

Modélisation et analyse des processus stochastiques

Laurent Lefèvre

18 mars 2022

Ces notes de cours ont été rédigées dans le cadre de l'organisation à distance de l'enseignement du cours MA332 *Modélisation et analyse des processus stochastiques*. Des visioconférences seront organisées pour vous présenter les différents chapitres et répondre à vos questions sur le cours ou les exercices. Il s'agira donc d'une expérience de *pédagogie inversée* pour laquelle il vous est demandé de lire les notes ou de faire les exercices, en amont des séances en visio qui seront organisées comme des sessions de questions/réponses et de discussion. Le programme de chaque séance vous sera communiqué au fur et à mesure de l'avancement du cours. La version en ligne de ces notes sur *chamilo.grenoble-inp.fr* sera également mise à jour au fur et à mesure (n'hésitez pas à me remonter les erreurs éventuelles). Pour toute question concernant l'organisation de cet enseignement, n'hésitez pas à me contacter directement (*laurent.lefevre@lcis.grenoble-inp.fr*).

Table des matières

1	Introduction générale	5
1.1	Objet et objectifs du cours	5
1.2	Organisation de l'enseignement	10
2	Processus de Poisson	13
2.1	Classification	13
2.2	Définition et caractérisation	18
2.2.1	propriétés élémentaires	20
2.2.2	première caractérisation	22
2.2.3	deuxième caractérisation	24
2.3	Âge et temps de vie résiduel	28
2.4	Décomposition probabiliste	32
2.5	Estimation	35
2.6	Cas non stationnaire	37
2.7	Exercices	40
3	Chaînes de Markov à temps discret	43
3.1	Introduction et définition	43
3.2	Quelques exemples classiques de CMTD	47
3.2.1	La chaîne à deux états	48
3.2.2	Modèle de diffusion	48
3.2.3	Promenade aléatoire	49
3.2.4	Le modèle de la ruine du joueur	51
3.2.5	Système dynamique déterministe	52
3.3	Probabilités d'états et probabilités de chemins	54
3.3.1	Probabilité d'état	54
3.3.2	Probabilité d'un chemin	55
3.4	Classes d'états dans une CMTD	58
3.4.1	Classes d'équivalence dans une CMTD	58
3.4.2	Périodicité dans une CMTD	59
3.4.3	Etats récurrents et transitoires	61
3.4.4	Exemple	66
3.5	Temps de séjour et lois d'entrée dans une sous-chaîne	70

3.5.1	Temps moyen d'entrée ou de séjour dans une sous-chaîne	70
3.5.2	Lois d'entrée dans une sous-chaîne	72
3.6	Régime permanent	75
3.7	Exercices	80
4	Chaînes de Markov à temps continu	87
4.1	Introduction et définition	87
4.2	Caractérisations des CMTC	94
4.2.1	Temps de séjours et probabilités de transition	94
4.2.2	Taux de transition	97
4.2.3	Le modèle de l'élève revisité	100
4.2.4	Le cas des rebouclages directs	102
4.3	Régimes transitoire et permanent	105
4.3.1	régime transitoire	105
4.3.2	régime permanent	107
4.3.3	méthode des coupes	113
4.4	Exercices	116
5	Files et réseaux de files d'attente	121
5.1	Introduction et définitions	121
5.1.1	La file simple	122
5.1.2	Les réseaux de files d'attente	128
5.2	Exemples	131
5.2.1	Machine repairman model	131
5.2.2	Systèmes informatiques	132
5.2.3	réseaux de communication	135
5.2.4	systèmes de production flexibles	138
5.3	Paramètres de performance	142
5.3.1	Définition des paramètres opérationnels	142
5.3.2	Stabilité et ergodicité	145
5.3.3	Loi de Little	150
5.4	Analyse des files simples	155
5.4.1	La file M/M/1	155
5.4.2	La file à capacité limitée $M/M/1/K$	161
5.4.3	La file à plusieurs serveurs $M/M/C$	165
5.4.4	La file <i>Infinite Servers</i> (IS) $M/M/\infty$	169
5.4.5	La file markovienne générale	172
5.5	Exercices	175
6	Corrections des exercices	181
6.1	Processus de Poisson	181
6.2	Chaînes de Markov à temps discret	198
6.3	Chaînes de Markov à temps continu	230
6.4	Files et réseaux de files d'attente	252

Chapitre 1

Introduction générale

1.1 Objet et objectifs du cours

Dans ce cours, nous allons apprendre à *modéliser de manière probabiliste* et *analyser les performances* de *systèmes à évènements discrets*. Afin de préciser les termes employés ci-dessus, nous allons considérer deux exemples.

Exemple 1 (lancer 1D6) Un exemple simple est celui du lancer d'un dé à six faces (1D6, en abrégé). Connaissant parfaitement la position (six coordonnées) et la vitesse initiales du dé par rapport à la surface de contact, ses caractéristiques élastiques (ainsi que celles de la surface de contact), il est théoriquement possible de prédire de manière exacte – en intégrant le modèle mécanique – la trajectoire de ce dé et le résultat final, c'est-à-dire la face sur laquelle il va s'arrêter. Cependant, la position et la vitesse initiales du dé ne peuvent être connues parfaitement. La modélisation parfaite du choc (rebond) est illusoire. La modélisation du contact air/dé et des éventuelles turbulences aériennes est elle aussi d'une précision limitée. Au final, un effort considérable de modélisation et de calcul ne fournira pas nécessairement des résultats plus précis qu'un modèle probabiliste statuant que le dé retombera une fois sur six sur n'importe laquelle des faces donnée !

Exemple 2 (réseau à commutation de paquets) dans un réseau de communication avec routage par paquets, la destination du paquet est déterminée à l'émission et son routage est déterminé à priori ou en temps réel, par exemple en fonction de l'occupation des différents noeuds et liaisons du réseau. L'état du réseau, défini par le nombre de paquets dans chaque noeud et chaque liaison, peut donc être prédit de manière déterministe, connaissant les performances des routeurs, des lignes de transmission, l'algorithme de routage, etc. Néanmoins, les réseaux (par exemple le réseau Internet) peuvent être de très grande taille, non homogène et soumis à des évènements "imprévisibles" (par exemple, des ruptures de lignes ou des pannes de routeurs). Un modèle probabiliste du réseau considérera que les paquets, lors-

qu'ils arrivent à un noeud du réseau, sont routés de manière probabiliste vers un des noeuds connectés à celui-ci (par exemple de manière équiprobable).

Modélisation probabiliste

Les systèmes réels sont le plus souvent déterministes (au moins du point de vue macroscopique), mais ils sont éventuellement *trop complexes* ou *insuffisamment connus* pour qu'un modèle déterministe puisse être développé ou s'avérer utile. Un modèle probabiliste permettra souvent, à partir d'une connaissance rudimentaire (ou d'une grande simplification) du système réel, des analyses et des prédictions partielles de son comportement qui peuvent être cependant utiles.

Systèmes à évènements discrets

Les systèmes considérés dans ce cours sont des processus stochastiques qui évoluent au cours du temps. Leur état à un instant donné est décrit par une variable d'état discrète (à valeur dans un espace dénombrable, fini ou non), notée X_i , dans le cas des systèmes à temps discret (où $i \in \mathbb{N}$ désigne l'instant t_i), ou notée $X(t)$, dans le cas des systèmes à temps continu (avec $t \in \mathbb{R}_+$). Les changements d'état (lorsque l'état du système passe d'une valeur discrète à une autre) se produisent lors des occurrences d'évènements considérées comme aléatoires. Le système est alors dit *à évènements discrets*. Par exemple, nous pouvons considérer, lors de lancers successifs d'1D6, que l'état du système est décrit par la somme des résultats obtenus. Chaque nouveau lancer est un "évènement" qui modifie la valeur de l'état courant. Dans le cas du réseau de communication par commutation de paquets, l'état est défini comme le nombre de paquets en attente ou en émission à chacun des noeuds du réseau, ainsi que l'état des canaux de transmission (libre ou saturé) et éventuellement des noeuds (serveurs). Un évènement est alors le début (ou la fin) de l'émission d'un paquet, ou encore son arrivée au noeud de réception, le début ou la fin de l'indisponibilité d'une ligne, etc.

Analyse de performances

La calcul ou l'analyse des performances d'un système doit parfois se faire à partir de son modèle. Les mesures directes peuvent être impossibles, par exemple dans la phase de conception (ou le système réel n'existe pas encore!) ou parce qu'il est impossible de modifier un système en cours d'exploitation (pour mettre en place les mesures nécessaires). Le modèle probabiliste est en général fondé sur une simplification extrême du comportement du système et ne fournit que des informations de nature probabiliste sur celui-ci. Cependant, le concepteur est souvent intéressé davantage à certains paramètres moyens (valeurs sur le long terme) qu'à la valeur de l'état du système à un instant particulier. Dans l'exemple des lancers successifs d'1D6,

il peut être intéressant de prédire la fréquence relative de chaque résultat ($1/6$, en l'occurrence) ou encore la valeur moyenne de l'état¹ ($n \cdot 3,5$ ou n désigne le nombre de lancers). Dans l'exemple du réseau à commutation de paquets, l'état du système à un instant précis présente peu d'intérêt. Par contre, le temps de réponse², le débit, le nombre de paquets dans les buffers ou les taux d'utilisation des routeurs sont des informations cruciales, aussi bien dans la phase de conception du réseau que pour sa surveillance (monitoring). Ces grandeurs d'intérêt sont des variables aléatoires dont les distributions peuvent être calculées, en principe, à partir du modèle probabiliste du système. En pratique cependant, le calcul des distributions complètes peut s'avérer trop compliqué. Par ailleurs, le plus souvent, seuls les premiers moments de ces distributions nous intéressent. Dans l'exemple du réseau à commutation, on s'intéressera essentiellement aux moments d'ordre un (espérances) : temps de séjour moyen, débit moyen dans le réseau, nombre moyen de clients dans les buffers, etc. Ces valeurs moyennes calculées sur des horizons de temps longs sont appelées *paramètres de performances* du processus stochastique. L'étude des réseaux comme processus stochastiques et l'analyse de leurs performances remontent au moins aux années 60's [Onaga, 1969]. Le lecteur intéressé se référera à [Park, 1976] pour une première synthèse ou [Bolch et al., 2006, Gelenbe and Mitrani, 2010] pour une présentation pédagogique plus récente.

De très nombreux systèmes peuvent être modélisés de manière probabiliste, comme des systèmes à événements discrets, et leurs performances analysées au sens développé ci-dessus. En plus des réseaux de communication, déjà évoqués, nous mentionnerons à titre d'exemples les files et réseaux de files d'attente, les systèmes de production et les systèmes informatiques. Historiquement, la théorie des processus stochastiques a d'abord été appliquée, de manière systématique, aux files d'attente simple (de type client/serveur, comme un guichet à la poste, par exemple) et aux systèmes de production dans des ateliers en général fortement mécanisés, voire automatisés [Buzacott, 1967, Forestier, 1980]. Un système de production est un réseau de lignes de production interconnectées où des pièces à usiner transitent entre différentes machines-outils. Les pièces sont prises en charge par des convoyeurs, parfois mises en attente dans des buffers en attendant que la machine suivante soit disponible. Selon les ateliers, les pièces peuvent être (ou non) routées pour suivre les opérations d'usinage dans des ordres différents, donnant parfois naissance à des produits finis différents. Développés pour des ateliers de production mécanique, les modèles stochastiques de ce type de processus ont rapidement été appliqués à une grande variété de processus

1. Dans les deux cas, la prédiction sera d'autant plus fiable que le nombre de lancers est grand, par application du théorème central-limite

2. délai entre l'émission et la réception d'un paquet, ou encore temps de séjour d'un paquet dans le réseau

de production et de processus logistiques (transports multi-modaux, stockage, reconditionnement, distribution) [Buzacott and Shanthikumar, 1993, Altioek, 2012].

Les systèmes informatiques ont été, eux aussi, modélisés comme des processus stochastiques. Dans ceux-ci, les processus clients circulent entre différentes ressources que sont les processeurs ou les unités entrée-sortie. Ils sont placés dans des files d'attente (buffers) en attendant que la ressource nécessaire soit disponible. L'état du système est caractérisé par le nombre de processus dans chacun des buffers, des processeurs et des unités entrée-sortie. Le traitement des processus dans chaque "station" est entièrement déterministe. Néanmoins, étant donné le nombre et la variétés de ces processus, il est plus facile de considérer que la durée de traitement d'un processus est une variable aléatoire. Les événements tels que le début ou la fin de traitement d'un processus dans un processeur, le début ou la fin d'écriture dans une unité de stockage, la demande d'accès ou la libération d'une ressource, etc. sont alors des événements discrets qui conditionnent le changement d'état du système. Cette approche permet d'analyser les performances d'un système informatique donné, notamment les taux d'utilisation des processeurs, le nombre moyen de processus exécuté par unité de temps (débit moyen), le temps moyen d'exécution des requêtes, le nombre moyen de processus en cours d'exécution ou en attente, le degré de parallélisme, etc. Elle a été appliquée avec succès à de nombreux exemples d'architecture avec serveur central ou multiprocesseurs [Menascé et al., 1994, Bolch et al., 2006].

Objectifs généraux du cours

- Donner une introduction détaillée aux différentes notions probabilistes nécessaires à la modélisation des processus stochastiques, plus particulièrement aux processus d'arrivées (ou de Poisson) et aux processus markovien (sans mémoire), en temps discret et en temps continu
- Développer la capacité à modéliser des systèmes réels à l'aide d'une approche probabiliste, puis à en analyser les performances
- Appliquer les notions vues à l'analyse des systèmes plus complexes, par exemple les files et réseaux de files d'attente
- Familiariser les étudiants avec le comportement des processus stochastiques étudiés, au travers de simulations numériques qui illustrent les différents résultats théoriques

Organisation générale du manuscrit

Ce polycopié est organisé comme suit. Un premier processus stochastique en temps continu est étudié dans le chapitre 2 : le processus de Poisson.

Il s'agit du modèle probabiliste d'un processus de comptage d'arrivées (ou d'événements, en général) indépendantes, très utilisé comme élément constitutif de systèmes plus complexes. Il s'agit également d'un premier exemple de processus markovien (en temps continu) et son étude détaillée prépare les généralisations abordées dans les chapitres suivants. Le chapitre 3 est dédié à l'étude des chaînes de Markov à temps discret (CMTD) et à leurs nombreuses applications. Les différentes classes de CMTD et leurs états sont présentés, ainsi que les méthodes de calcul des probabilités, transitoires ou stationnaires, les plus couramment rencontrées en pratique. Une analyse comparable des chaînes de Markov à temps continu (CMTC) fait l'objet du chapitre 4. Les deux approches, par probabilités de passage et temps de séjour dans les différents états, ou par taux de transition (générateur infinitésimal) sont présentées et comparées. Le calcul et l'analyse du régime permanent font l'objet d'une attention particulière. Dans le chapitre 5, les résultats relatifs aux CMTC sont appliqués à la modélisation et à l'analyse des performances des files simples et des réseaux de files d'attente, comme ceux rencontrés dans la modélisation des systèmes informatiques ou des réseaux de communication. Après un examen détaillé des files simples les plus courantes, des outils généraux sont introduits pour l'analyse de performance des réseaux de files d'attente, en particulier des réseaux à forme produit. Le chapitre 6 est consacré aux modèles de Markov cachés (MMC). Trois questions sont abordées : le calcul de la probabilité d'une séquence d'observations, celui du chemin le plus probable (estimation d'état) et celui de l'estimation des paramètres du MMC. Des algorithmes performants et utilisés en pratique sont présentés pour résoudre ces trois problèmes. Ces notions et algorithmes sont ensuite appliqués à des problèmes de détection de malware et de cryptanalyse. Le reste du polycopié est consacré à la correction des exercices proposés à la fin de chaque chapitre. Les exemples et exercices proposés dans ce cours sont – dans la mesure du possible – orientés vers des applications d'intérêt général ou susceptibles d'exciter la curiosité d'étudiants en informatique, réseaux et cybersécurité.

Pour les chapitres 1 à 5, le plan général suivi est celui proposé dans [Baynat, 2000] qui peut utilement servir de référence pour le cours. Pour une approche plus formelle et des démonstrations plus rigoureuses des principaux théorèmes, le lecteur pourra se référer à [Foata and Fuchs, 2004]. L'ouvrage [Bolch et al., 2006] est une présentation plus complète sur les files et réseaux de file d'attente, avec application à l'analyse de performance de nombreux systèmes, notamment de systèmes informatiques ou de réseaux de communication. La référence [Stamp, 2017] contient (entre autres) une introduction concise, mais opérationnelle, aux MMC et à leurs applications aux problèmes de cybersécurité.

1.2 Organisation de l'enseignement

L'enseignement est réparti en 9 séances de cours magistraux et 8 séances de travaux dirigés, dont deux séances pendant lesquelles les étudiants implémentent (sur leur propre machine, en utilisant le langage R³) des simulations de processus de Poisson, de chaînes de Markov en temps discret et de files simples. En 2021, suite aux mesures sanitaires décidées dans le cadre de la pandémie, seules les deux premières séances de cours (23 et 26 février) sont prévues en présentiel. Il est actuellement (22/02/2021) prévu que toutes les autres séances auront lieu à distance. Les chapitres à lire ou les exercices à faire en préparation de ces séances à distances seront précisés d'une fois sur l'autre. La documentation relative au cours sera régulièrement mise à jour et disponible sur chamilo.grenoble-inp.fr (cours MA332 Modélisation stochastique). La programmation des séances de cours est donnée à la table 1.1 ci-après, à titre indicatif. Elle pourra changer en fonction des circonstances et de l'avancement de l'enseignement.

La note finale (évaluation des étudiants) comprendra une note pour les comptes-rendus de TDM réalisés en binômes (note commune par binôme, comptant pour 25% de la note globale de chaque étudiant) et une note individuelle d'examen comptant pour 75% de la note globale. Si les conditions sanitaires le permettent, l'examen sera un examen écrit, d'une durée de 01h30, avec une feuille A4 recto-verso manuscrite (à rendre avec la copie d'examen) et calculatrice. L'examen comprendra une succession d'exercices similaires à ceux réalisés durant les séances d'exercices (TD). En cas d'évaluation à distance, l'examen prendra la forme d'un entretien oral (avec documents et calculatrices) qui s'appuiera sur les compte-rendu des TDM et sur les notions vues en cours. L'examen de rattrapage prendra la même forme que l'examen de première session. La note de l'examen de rattrapage remplace la note de l'examen de première session. La note du compte-rendu du TDM, commune à chaque binôme, n'est pas rattrapable.

3. GNU R est un logiciel libre distribué selon les termes de la licence GNU GPL. Il était à l'origine dédiés aux calculs statistiques et à la science des données (data science). Il est aujourd'hui très populaire, non seulement parmi les statisticiens (et data scientists) mais aussi parmi ceux qui étudient les systèmes complexes en réseaux (e.g. Internet, Facebook, les smart-grids, etc.) et ceux qui étudient l'intelligence artificielle en général et le machine learning en particulier. Ce langage sera également utilisé en quatrième année (IR&C) dans le cours de *méthodes statistiques appliquées à la sécurité* et en cinquième année dans le cours *complex network systems*.

Séance	Date	Type	Contenu	Préparation
N° 1	23/02	cours - en présentiel	Introduction générale + Processus de Poisson 1	aucune
N° 2	26/02	cours - en présentiel	Processus de Poisson 2	relire cours PP1
N° 3	03/03	cours - à distance	Chaînes de Markov à temps discret 1	relire cours PP2
N° 4	03/03	exercices - à distance	Processus de Poisson	préparer exercices PP : exercices 1 → 5
N° 5	09/03	cours - à distance	Chaînes de Markov à temps discret 2	lire cours CMTD : sections 3.1 → 3.3
N° 6	12/03	exercices - à distance	Processus de Poisson	préparer exercices PP : exercices 6 → 8
N° 7	16/03	cours - à distance	Chaînes de Markov à temps continu 1	lire cours CMTD : sections 3.4 → 3.6
N° 8	16/03	exercices - à distance	CMTD	préparer exercices 9 → 15
N° 9	23/03	cours - à distance	Chaînes de Markov à temps continu 2	lire cours CMTC sections 4.1 → 4.2
N° 10	23/03	exercices - à distance	CMTD	préparer exercices 16 → 22
N° 11	29/03	cours - à distance	Files et réseaux de files d'attente 1	lire cours CMTC section 4.3
N° 12	30/03	exercices - à distance	CMTC	préparer exercices 23 → 27
N° 13	01/04	cours - à distance	Files et réseaux de files d'attente 2	lire cours Files sections 5.1 → 5.3
N° 14	06/04	TDM - à distance	PP + CMTD	installer R apprendre à utiliser R
N° 15	13/04	exercices - à distance	Files	préparer exercices 28 → 35
N° 16	14/04	cours - à distance	Files et réseaux de files d'attente 3	lire cours Files sections 5.4 → 5.5
N° 17	14/04	TDM - à distance	CMTC + Files	programmation / compte-rendu

TABLE 1.1 – Déroulement chronologique des séances de cours et d'exercices
(prévisions faites au 22 février)

Chapitre 2

Processus de Poisson

Les processus de Poisson, font partie des processus stochastiques les plus utilisés en pratique, notamment dans la modélisation des files et réseaux de files d'attente. Ils modélisent et permettent de caractériser le nombre et la répartition d'évènement aléatoires indépendants et de même type (même probabilité d'occurrence). L'occurrence d'un de ces évènements est généralement appelée *arrivée*, par référence aux arrivées des clients dans les files d'attente. La présentation et la classification générale des processus stochastiques proposée dans la section 2.1 est inspirée de [Baynat, 2000] (section 2.4). Les démonstrations des propriétés des processus de Poisson (sections 2.2 et suivantes) sont détaillées dans [Baynat, 2000] (section 2.5) ou [Foata and Fuchs, 2004] (chapitre 3).

2.1 Classification des processus stochastiques

Un processus stochastique est un système dont l'état X varie au cours du temps et dont la valeur à chaque instant t , notée $X(t)$, dépend de l'issue d'une expérience aléatoire. De manière tout à fait générale, un processus stochastique peut donc être défini comme une famille de variables aléatoires

$$\{X(t)\}_{t \in T}$$

où l'indice des temps t appartient à un ensemble de valeurs $T \subset \mathbb{R}$ qui peut être *discret* (c'est-à-dire dénombrable, fini ou non) ou continu. On parlera de processus à *temps discret* lorsque T est discret (on considère souvent le cas $T = \mathbb{N}$), ou à *temps continu* lorsque T contient au moins un intervalle réel ouvert non vide dans \mathbb{R} (on considère souvent le cas $T = [0, \infty[$). Chacune des variables aléatoires de la famille est définie, de manière habituelle, comme une application

$$X(t) : \Omega \rightarrow E \subset \mathbb{R} : \omega \mapsto X(\omega)$$

où Ω désigne l'ensemble des issues ω possibles pour l'expérience aléatoire considérée et $X(\omega)$ la caractéristique réelle associée. L'espace E des

valeurs est appelé *espace d'état*. Il peut lui-même être continu ou discret (i.e. dénombrable). On parle dans ce dernier cas de processus à état discret ou de manière équivalence d'un *chaîne*. On supposera que Ω est un espace probabilisé, c'est-à-dire muni d'une tribu¹ d'évènements \mathcal{A} et d'une mesure de probabilité² $\mathbb{P}[\cdot]$ associée à ces évènements. Une même expérience aléatoire peut être associée à différentes caractéristiques réelles d'intérêt et donner naissance à différents processus stochastiques associées, comme le montre l'exemple suivant.

Exemple 3 (système informatique) On étudie pendant une heure le fonctionnement d'un système informatique auquel sont adressées des requêtes, sous la forme d'arrivées de processus (au sens informatique d'un processus). L'arrivée des requêtes est considérée comme aléatoire, ainsi que la durée nécessaire au traitement de chacune d'elles. Lorsque la ressource nécessaire au traitement d'un processus n'est pas disponible, celui-ci est mis en attente. Il est alors possible de définir différents processus stochastiques associés à l'arrivée des requêtes et à leur traitement par le systèmes informatique :

(3.1) Le nombre de requêtes

$$\{X(n)\}_{n \in \{1, \dots, 60\}}$$

adressées à chaque minute au système informatique est un exemple de chaîne à temps discret. Dans le cas des processus à temps discret, pour des valeurs $n \in \mathbb{N}$, on notera souvent X_n au lieu de $X(n)$

(3.2) Le temps moyen de séjour (attente et traitement)

$$\{T_n\}_{n \in \{1, \dots, 60\}}$$

des requêtes arrivées dans le systèmes pendant la minute $n \in \{1, \dots, 60\}$ est un exemple de processus stochastique à état continu et à temps discret

(3.3) Le nombre de processus

$$\{N(t)\}_{t \in [0, 60]}$$

présents dans le systèmes (en attente ou en cours de traitement) au temps t est un exemple de chaîne à temps continu

1. une tribu $\mathcal{A} \subset \mathcal{P}(\Omega)$ sur un ensemble Ω est un ensemble non vide de parties de Ω , stable par passage au complémentaire et par union dénombrable (donc aussi par intersection dénombrable)

2. une mesure de probabilité $\mathbb{P}[\cdot]$ sur \mathcal{A} est une mesure définie $\forall A \in \mathcal{A}$, à valeur dans $[0, 1]$, de masse unitaire ($\mathbb{P}[\Omega] = 1$) et telle que pour toute suite dénombrable d'évènements $A_1, A_2, \dots \in \mathcal{A}$ disjoints deux à deux :

$$\mathbb{P}\left[\bigcup_{i \geq 1} A_i\right] = \sum_{i \geq 1} \mathbb{P}[A_i]$$

(3.4) Le temps moyen de séjour

$$\{T(t)\}_{t \in [0,60]}$$

des processus arrivés dans le systèmes au temps t est un exemple de processus stochastique à état continu et à temps continu

Des réalisations particulières des quatres types de processus stochastiques, en temps discret ou continu, à états discrets ou continus, sont représentées à la figure 2.1. Dans la suite, par convention de notation, nous noterons $x(t)$ ($t > 0$) une réalisation particulière de la variable aléatoire $X(t)$. Nous noterons de même x_k ($k \in \mathbb{N}$) une réalisation particulière de la variable aléatoire $X(t_k)$.

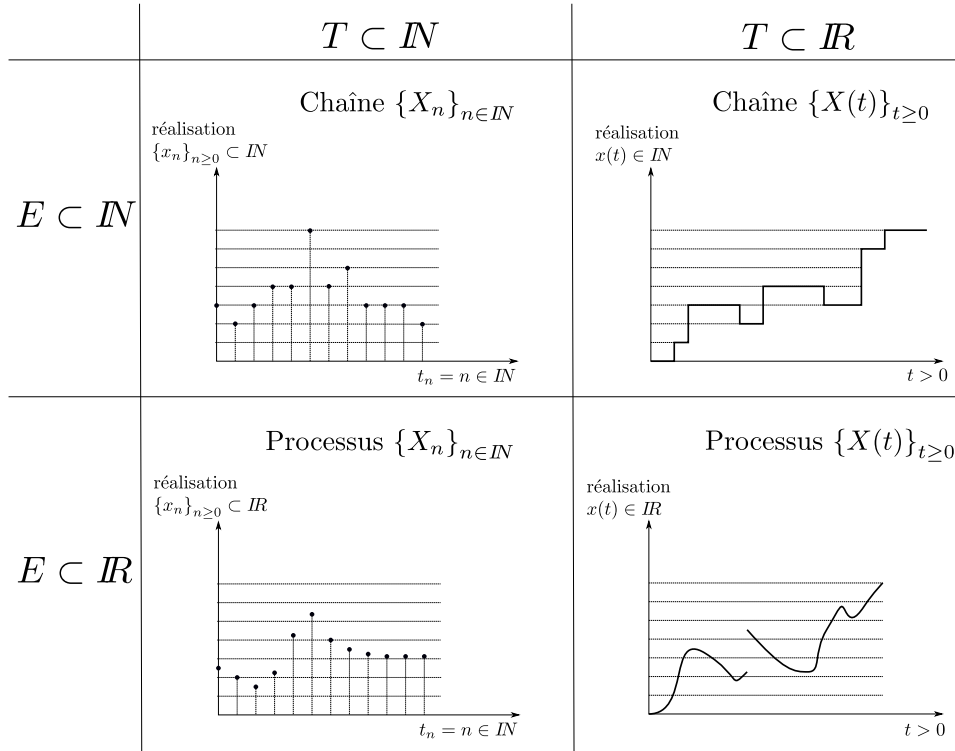


FIGURE 2.1 – Réalisations particulières pour quatre types de processus stochastiques : une chaîne à temps discret, une chaîne à temps continu, un processus à état continu et temps discret, un processus à état continu et temps continu

Dans ce cours, nous nous intéresserons exclusivement à des processus stochastiques à état discret, également appelés chaînes. Les processus de Poisson sont des exemples de chaînes à temps continu. Ce sont en particulier des *processus de comptage*

Définition 1 (processus de comptage) Un processus de comptage est une chaîne (i.e. un processus stochastique à état discret) dont la réalisation est nécessairement *nulle au temps initial* et *monotone non décroissante*.

- Une chaîne à temps discret est donc un processus de comptage si pour toute réalisation $\{x_n\}_{n \in \mathbb{N}}$, on a nécessairement $x_0 = 0$ et $x_n \leq x_{n+1}$, $\forall n \in \mathbb{N}$
- Une chaîne à temps continu est donc un processus de comptage si pour toute réalisation $x(t)$, on a nécessairement $x(0) = 0$ et $t \leq s \Rightarrow x(t) \leq x(s)$

La figure 2.2 montre un exemple de réalisation particulière d'un processus de comptage à temps continu. En principe, rien ne s'oppose à ce que la valeur de la chaîne augmente arbitrairement lors d'une transition dans un processus de comptage à valeurs dans $E = \mathbb{N}$. Dans le cas des processus de Poisson cependant, comme nous le verrons ultérieurement, la probabilité pour que la chaîne augment de plus d'une unité lors d'une transition est négligeable.

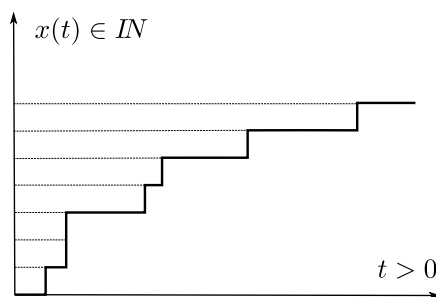


FIGURE 2.2 – réalisation d'un processus de comptage à temps continu

Pour définir complètement une chaîne, de manière probabiliste, il faut préciser la distribution de probabilité pour l'état initial (au temps 0) et la loi qui détermine les transitions entre états.

La distribution initiale de probabilité indique la probabilité de chacun des états de E au temps $t = 0$. En l'absence de toute information, dans le cas où E est fini, on considérera souvent une loi uniforme. Dans le cas d'un processus de comptage, on sait que cet état initial est nécessairement $x = 0$. Dans le cas où l'état initial x_0 est connu (avec certitude), la distribution initiale de probabilité est donnée par $\mathbb{P}[X(0) = x] = \delta_{x_0, x}$, $\forall x \in E$, où $\delta_{a,b}$ désigne le symbole de Kronecker³ sur E .

3. Le symbole de Kronecker sur E est défini par

$$\delta : E \times E \rightarrow \{0, 1\} : (x, y) \mapsto \delta_{x,y} = \delta(x, y) = \begin{cases} 1 & \text{si } x = y \\ 0 & \text{sinon} \end{cases}$$

Définir les probabilités de transition entre états est plus complexe. Nous nous limiterons aux processus causaux pour lesquels la valeur de l'état $x(\bar{t})$ à un instant particulier \bar{t} ne peut dépendre que des états passés de la chaîne⁴, c'est-à-dire du graphe $\{(t, x(t)) \in T \times E \mid t < \bar{t}\}$. Dans le cas général d'un processus causal, les probabilités de transition peuvent donc dépendre non seulement de la valeur de l'état à l'instant présent t , mais également de toute l'histoire passée du système, c'est-à-dire de tous les instants de transition entre états et de toute la séquence des états passés.

Considérons dans un premier temps le cas d'une chaîne causale à temps discret. Pour caractériser complètement l'évolution dans le temps de cette chaîne, de manière probabiliste, il est donc nécessaire de définir les probabilités

$$IP[X_n = x_n \mid X_{n-1} = x_{n-1}, \dots, X_1 = x_1, X_0 = x_0]$$

pour toutes les valeurs du temps $n \in \mathbb{N}$, pour tous les états courants x_n possible et pour toutes les séquences possibles d'états passés $x_0, x_1, \dots, x_{n-1} \in E$. La caractérisation de l'évolution des chaînes à temps continu est encore plus complexe. Pour celles-ci, la probabilité d'état au temps t peut dépendre de l'ensemble non dénombrable des valeurs des états passés. Une manière de définir une chaîne à temps continu est de donner l'ensemble des probabilités

$$IP[X(t_n) = x_n \mid X(t_{n-1}) = x_{n-1}, \dots, X(t_1) = x_1, X(t_0) = x_0]$$

pour tous les choix possibles d'instant $0 < t_0 < t_1 < \dots < t_n$, pour tous les états possibles $x_0, x_1, \dots, x_n \in E$ et pour tous les nombres d'échantillons $n \in \mathbb{N}$. Les caractérisations qui précèdent sont très générales mais ne permettent en général pas d'aboutir à des résultats analytiques concernant le comportement de la chaîne, ni de résoudre des problèmes simples de calcul de probabilités d'englober la très large classe des chaînes causales. Cependant, elles sont difficiles à spécifier et ne permettent pas – en général – d'obtenir l'évolution dans le temps des probabilités d'état, ni même de répondre à des questions simples sur le comportement de la chaîne par des calculs probabilistes explicites. Comme toujours, nous allons donc devoir restreindre la classe des systèmes considérés afin de pouvoir en dire quelque chose d'intéressant! En particulier, dans ce cours, nous allons considérer (sauf mention explicite) des chaînes stationnaires à incréments indépendants.

Définition 2 (Processus à incréments indépendants) Un processus stochastique à temps discret $\{X_n\}_{n \geq 0}$ est dit à *accroissements indépendants* si et seulement si les incréments

$$X_{n+1} - X_n, X_n - X_{n-1}, \dots, X_2 - X_1, X_1 - X_0$$

4. Dans des processus non causaux, la valeur de l'état $x(\bar{t})$ peut dépendre des valeurs $x(t)$ pour $t > \bar{t}$. Cela peut être le cas lorsque les valeurs futures l'état influencent le comportement actuel du système, mais aussi les cas où la variable t ne représente pas le temps, mais par exemple une variable d'espace.

sont des variables aléatoires indépendantes, quelque soit $n \in \mathbb{N}$. De même, un processus stochastique à *temps continu* $\{X(t)\}_{t \geq 0}$ est dit à *accroissements indépendants* si et seulement si les incréments

$$X(t_{n+1}) - X(t_n), X(t_n) - X(t_{n-1}), \dots, X(t_2) - X(t_1), X(t_1) - X(t_0)$$

sont des variables aléatoires indépendantes, pour toute suite d'instants d'échantillonnages $0 < t_0 < t_1 < \dots < t_n$ et quelque soit $n \in \mathbb{N}$.

Définition 3 (Processus stationnaire) Un processus stochastique à *temps discret* $\{X_n\}_{n \geq 0}$ est dit *stationnaire* si et seulement si les incréments

$$X_n - X_{n-1} \text{ et } X_1 - X_0$$

sont distribués selon la même loi, quelque soit $n \in \mathbb{N}$. De même, un processus stochastique à *temps continu* $\{X(t)\}_{t \geq 0}$ est dit *stationnaire* si et seulement si les incréments

$$X(s+t) - X(s) \text{ et } X(t) - X(0)$$

sont distribuées selon la même loi, quelque soit $s > 0$.

Pour définir un processus stationnaire et à incréments indépendants, il suffit donc de donner cette loi (ou famille de lois) unique pour tous les incréments, à savoir la *distribution commune* de probabilité des variables

$$\Delta_n := X_n - X_{n-1}$$

dans le cas discret, ou la famille de distributions de probabilité des variables

$$\Delta(t) := X(s+t) - X(s)$$

dans le cas continu⁵. Les processus stochastiques étudiés dans ce cours seront – sauf mention contraire – des chaînes stationnaires, à incréments indépendants, à temps discret ou à temps continu selon les cas. Le premier d'entre eux, que nous allons analyser en détail dans ce chapitre, est le processus de Poisson. Il s'agit d'un processus à état discret (i.e. une chaîne), à temps continu, de comptage, stationnaire et à incréments indépendants.

2.2 Définition et caractérisations des processus de Poisson

Le processus de Poisson est l'un des processus stochastiques les plus utilisés pour modéliser et analyser la répétition d'événements aléatoires du

5. Par définition, ces distributions sont indépendantes de l'instant particulier s considéré et ne dépendent que de la durée t de l'intervalle de temps pendant lequel l'accroissement de la variable X est calculé

même type. Les *arrivées* de *clients* dans les systèmes de files d'attente que nous étudierons plus loin, par exemple, seront modélisées par des processus de Poisson.

Le processus de Poisson est un processus de comptage à temps continu, stationnaire et à incréments indépendants. La famille de lois choisies comme distribution de ces incréments est une famille de lois de Poisson. Ce choix particulier de lois conduit à une propriété fondamentale du processus de Poisson : l'absence de mémoire. La durée d'attente avant la prochaine réalisation ne dépend pas du temps déjà attendu depuis la dernière réalisation. C'est une hypothèse très forte qui n'est pas toujours satisfaite dans les applications considérées. Cependant, elle facilite grandement les calculs et permet d'obtenir de nombreux résultats analytiques explicitement. Ces résultats, à leur tour, permettent de mieux comprendre qualitativement et d'analyser quantitativement le comportement de la chaîne. Ils fournissent une base intuitive solide à l'interprétation des résultats numériques qui peuvent être obtenus par simulation, dans le cas de processus de comptage avec d'autres lois plus générales ou plus complexes. Par ailleurs, en cas d'évolution des caractéristiques du processus d'arrivée lente par rapport à la constante de temps caractéristique de ces arrivées, il est toujours possible d'adapter régulièrement la valeur numérique paramètre du processus de Poisson pour "recoller" à la réalité mesurée.

Les exemples d'application du processus de Poisson à la modélisation de systèmes "réels" sont nombreuses :

- arrivées des clients dans une file d'attente : appels dans un central téléphonique, paquets dans un noeud de réseau à commutation, requêtes à un serveur, processus à un processeur, voitures à une barrière de péage, clients dans un magasin ou un service, etc.
- erreurs dans une ligne de transmission ou, plus généralement, dans le transfert d'information entre un émetteur et un récepteur
- désintégration des particules dans une masse donnée de matériau radio-actif
- mutations dans une séquence ADN, occurrence de certains motifs ou gènes dans une séquence ADN
- accidents ou pannes dans les installations industrielles, sinistres divers subis par les assurés, accidents de la route
- ...

Définition 4 (Processus de Poisson) Un processus stochastique $\{N(t)\}_{t \geq 0}$ à temps continu, à valeurs dans $E = \mathbb{N}$ (chaîne), est un processus de Poisson de paramètre λ ($\lambda \in \mathbb{R}_+^*$) si et seulement si :

- (i) $\{N(t)\}_{t \geq 0}$ est un processus de comptage
- (ii) $\{N(t)\}_{t \geq 0}$ est un processus stationnaire à incréments indépendants

$$(iii) \quad \mathbb{P}[N(t+s) - N(s) = k] = \frac{(\lambda t)^k}{k!} e^{-\lambda t}, \quad \forall k \in \mathbb{N}; \quad \forall t, s > 0$$

La propriété (iii) indique que la loi (discrète) du nombre d'arrivées pendant un intervalle de temps de longueur $t > 0$ est une loi de Poisson de paramètre λt . Cette loi est la même, quelque que soit l'instant s considéré (caractère stationnaire).

2.2.1 propriétés élémentaires

Beaucoup de problèmes pratiques peuvent être formulés de manière équivalente, soit en terme de nombre d'arrivées durant un intervalle de temps donné (donc par des calculs de variables aléatoires discrètes), soit en terme de durée entre les arrivées (donc avec des variables aléatoires continues). Nous pouvons, par exemple, nous interroger sur la probabilité qu'un client arrive dans les cinq prochaines minutes. Ce problème revient à calculer la probabilité que le temps avant la prochaine arrivée soit inférieur à cinq minutes. Nous verrons que ce temps est une variable aléatoire de distribution exponentielle. Mais ce même problème, de manière équivalente, revient à calculer la probabilité que le nombre de clients qui arrivent dans les cinq prochaines minutes soit supérieur ou égal à un. Nous allons voir que ce nombre est distribué selon une variable de Poisson. Parmi les propriétés du processus de Poisson, deux caractérisations jouent donc un rôle essentiel (notamment en pratique) : la première concerne la *probabilité d'arrivée d'un client pendant in intervalle de temps* (court) donné et la seconde concerne la *distribution du temps entre deux arrivées*.

Proposition 1 Soit $\{N(t)\}_{t>0}$, un processus de Poisson de paramètre λ . Alors, le nombre d'arrivées sur un intervalle de temps $[0, t]$ donnée ($t > 0$) suit une loi de Poisson de paramètre λt :

$$\mathbb{P}[N(t) = k] = \frac{(\lambda t)^k}{k!} e^{-\lambda t}, \quad \forall k \in \mathbb{N}$$

Démonstration : Il suffit d'appliquer le point (iii) de la définition 4 pour le cas particulier $s = 0$. On trouve :

$$\mathbb{P}[N(t) - N(0) = k] = \frac{(\lambda t)^k}{k!} e^{-\lambda t}, \quad \forall k \in \mathbb{N}$$

Comme $N(t)$ est un processus de comptage, $N(0) = 0$ et

$$\mathbb{P}[N(t) - N(0) = k] = \mathbb{P}[N(t) = k]$$

■

Proposition 2 Soit $\{N(t)\}_{t>0}$, un processus de Poisson de paramètre λ . Alors,

$$\mathbb{P}[N(t) = k | N(t_0) = k_0] = \frac{(\lambda(t - t_0))^{k-k_0}}{(k - k_0)!} e^{-\lambda(t-t_0)}, \quad \forall k \geq k_0, \quad \forall t > t_0$$

Démonstration : Par définition de probabilité conditionnelle, on obtient :

$$P := \mathbb{P}[N(t) = k | N(t_0) = k_0] = \frac{\mathbb{P}[N(t) = k; N(t_0) = k_0]}{\mathbb{P}[N(t_0) = k_0]}$$

Comme $\{N(t)\}_{t>0}$ est un processus de comptage, on a de manière équivalente

$$P = \frac{\mathbb{P}[N(t) - N(t_0) = k - k_0; N(t_0) - N(0) = k_0]}{\mathbb{P}[N(t_0) = k_0]}$$

Comme $\{N(t)\}_{t>0}$ est un processus à incréments indépendants, on a :

$$\begin{aligned} P &= \frac{\mathbb{P}[N(t) - N(t_0) = k - k_0] \cdot \mathbb{P}[N(t_0) - N(0) = k_0]}{\mathbb{P}[N(t_0) = k_0]} \\ &= \frac{\mathbb{P}[N(t) - N(t_0) = k - k_0] \cdot \mathbb{P}[N(t_0) = k_0]}{\mathbb{P}[N(t_0) = k_0]} \\ &= \mathbb{P}[N(t) - N(t_0) = k - k_0] \end{aligned}$$

En appliquant le point (iii) de la définition 4, on trouve finalement

$$P = \mathbb{P}[N(t) - N(t_0) = k - k_0] = \frac{(\lambda(t - t_0))^{k-k_0}}{(k - k_0)!} e^{-\lambda(t-t_0)}$$

■

Nous allons maintenant nous intéresser au nombre d'arrivées pendant un intervalle de temps de durée *infinitésimale* dt . Nous utiliserons la notation de Landau $o(dt)$ pour désigner une fonction arbitraire de dt dont la valeur est négligeable devant dt , au voisinage de 0 :

$$f(dt) = o(dt) \Leftrightarrow \lim_{dt \rightarrow 0} \frac{f(dt)}{dt} = 0$$

Proposition 3 Soit $\{N(t)\}_{t>0}$, un processus de Poisson de paramètre λ . Alors, quelque soit $t > 0$ et $k, j \in \mathbb{N}$:

$$\mathbb{P}[N(t + dt) = k + j | N(t) = k] = \begin{cases} \lambda dt + o(dt) & \text{si } j = 1 \\ o(dt) & \text{si } j > 1 \\ 1 - \lambda dt + o(dt) & \text{si } j = 0 \end{cases}$$

Démonstration : Ce résultat se déduit de la proposition 2. On a :

$$IP[N(t+dt) = k+j | N(t) = k] = \frac{(\lambda dt)^j}{j!} e^{-\lambda dt}$$

En utilisant le développement limité de l'exponentielle

$$e^{-\lambda dt} = 1 - \lambda dt + o(dt)$$

on obtient pour $j = 0$ (aucune arrivée sur $[t, t+dt]$) :

$$IP[N(t+dt) = k | N(t) = k] = 1 - \lambda dt + o(dt)$$

De même, on obtient pour $j = 1$:

$$IP[N(t+dt) = k+1 | N(t) = k] = \lambda dt (1 - \lambda dt + o(dt)) = \lambda dt + o(dt)$$

Enfin, pour $j > 1$, on a :

$$IP[N(t+dt) = k+j | N(t) = k] = \frac{(\lambda dt)^j}{j!} (1 - \lambda dt + o(dt)) = o(dt)$$

■

2.2.2 première caractérisation

La propriété 3 nous indique que la probabilité qu'il y ait plus d'une arrivée sur un intervalle de temps de longueur dt est négligeable devant dt et donc, en particulier, devant la probabilité qu'il y ait une ou aucune arrivée sur le même intervalle de temps. Par ailleurs, la probabilité d'une arrivée sur l'intervalle de temps $[t, t+dt]$ est *proportionnelle à la durée dt de cet intervalle* et indépendante⁶ de t . On peut montrer que la propriété 3 est un fait une caractérisation des processus de Poisson

Proposition 4 (première caractérisation) Soit $\{N(t)\}_{t \geq 0}$, un processus stochastique à espace discret et à temps continu qui vérifie les propriétés suivantes :

- (i) $\{N(t)\}_{t \geq 0}$ est un processus de comptage
- (ii) $\{N(t)\}_{t \geq 0}$ est un processus stationnaire à incréments indépendants
- (iii) pour tout $t \geq 0$, $k \in \mathbb{N}$ et $j = 2, 3, \dots$
 - $IP[N(t+dt) = k | N(t) = k] = 1 - \lambda dt + o(dt)$
 - $IP[N(t+dt) = k+1 | N(t) = k] = \lambda dt + o(dt)$
 - $IP[N(t+dt) = k+j | N(t) = k] = o(dt)$

Alors $\{N(t)\}_{t \geq 0}$ est un processus de Poisson de paramètre λ

6. la constante de proportionnalité λ ne dépend pas du temps t dans le cas d'un processus de Poisson stationnaire de paramètre λ

Démonstration : Il suffit de montrer la propriété (iii) de la définition 4. Le processus $\{N(t)\}_{t>0}$ étant un processus stationnaire, il suffit de montrer cette propriété pour $s = 0$. Comme $\{N(t)\}_{t>0}$ est un processus de comptage, cela revient à montrer :

$$P_k(t) := IP[N(t) = k] = \frac{(\lambda t)^k}{k!} e^{-\lambda t}, \forall k \in \mathbb{N}$$

Commençons par calculer $P_0(t)$. On obtient en appliquant la définition de probabilité conditionnelle et l'hypothèse $IP[N(t+dt) = k | N(t) = k] = 1 - \lambda dt + o(dt)$:

$$\begin{aligned} P_0(t+dt) &= IP[N(t+dt) = 0] \\ &= IP[N(t+dt) = 0 | N(t) = 0] \cdot IP[N(t) = 0] \\ &= (1 - \lambda dt + o(dt)) \cdot P_0(t) \end{aligned}$$

c'est-à-dire :

$$\frac{P_0(t+dt) - P_0(t)}{dt} = -\lambda P_0(t) + \frac{o(dt)}{dt}$$

Par passage à la limite $dt \rightarrow 0$, on obtient l'équation différentielle

$$P_0'(t) = -\lambda P_0(t)$$

avec la condition initiale $P_0(0) = 1$, dont l'unique solution est bien

$$P_0(t) = e^{-\lambda t}$$

Calculons maintenant de manière similaire, par récurrence, l'expression générale de $P_k(t) := IP[N(t) = k]$ pour $k \geq 1$. En appliquant la formule des probabilités totales, on trouve :

$$\begin{aligned} P_k(t+dt) &= IP[N(t+dt) = k] \\ &= \sum_{k_0=1}^k IP[N(t+dt) = k | N(t) = k_0] \cdot IP[N(t) = k_0] \\ &= (1 - \lambda dt + o(dt)) \cdot IP[N(t) = k] \\ &\quad + (\lambda dt + o(dt)) \cdot IP[N(t) = k-1] + \sum_{k_0=1}^{k-2} o(dt) IP[N(t) = k_0] \\ &= (1 - \lambda dt + o(dt)) P_k(t) \\ &\quad + (\lambda dt + o(dt)) P_{k-1}(t) + \sum_{k_0=1}^{k-2} o(dt) P_{k_0}(t) \end{aligned}$$

c'est-à-dire :

$$\frac{P_k(t+dt) - P_k(t)}{dt} = \left(-\lambda + \frac{o(dt)}{dt}\right) P_k(t) + \left(\lambda + \frac{o(dt)}{dt}\right) P_{k-1}(t) + \frac{o(dt)}{dt} \sum_{k_0=1}^{k-2} P_{k_0}(t)$$

Par passage à la limite $dt \rightarrow 0$, on obtient les équations différentielles

$$P'_k(t) = -\lambda P_k(t) + \lambda P_{k-1}(t)$$

avec les conditions initiales $P_k(0) = 0$. En posant $Q_k(t) := e^{\lambda t} P_k(t)$, ces équations différentielles s'écrivent

$$Q'_k(t) = \lambda Q_{k-1}(t)$$

avec les conditions initiales $Q_k(0) = P_k(0) = 0$. Il est alors trivial d'établir par récurrence

$$Q_k(t) = \frac{(\lambda t)^k}{k!}$$

et

$$P_k(t) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}$$

■

Remarque 1 (distribution conditionnelle des arrivées) Dans un processus de Poisson stationnaire (on dit aussi homogène), le paramètre λ est constant dans le temps. La probabilité d'une arrivée sur $[t, t + dt]$, égale à λdt , ne varie pas avec t . Si le nombre total n d'arrivées de clients sur $[0, t]$ est connu, les instants d'arrivées de chacune des $k \in \{1, \dots, n\}$ arrivées indépendantes, sont donc distribués de manière uniforme sur l'intervalle de temps $[0, t]$. Il ne peut en effet y avoir d'instant plus favorables que d'autres pour une arrivée puisque la densité λ (taux d'arrivée) est constante. Cette propriété de la distribution conditionnelle des instants d'arrivées⁷ permet de résoudre de manière très efficace des problèmes relatifs à la répartition dans le temps des arrivées (voir les exercices 1 et 4). Par exemple, si nous savons que 14 clients sont arrivés dans le système durant une heure donnée, nous savons que la probabilité que 10 d'entre eux soient arrivés pendant les vingt dernières minutes est

$$\binom{14}{10} \left(\frac{40}{60}\right)^4 \left(\frac{20}{60}\right)^{10} = \frac{14!}{10!4!} \frac{2^4}{3^{14}} \simeq 0,00334..$$

2.2.3 deuxième caractérisation

La deuxième caractérisation des Processus de Poisson, elle aussi très utilisée en pratique, concerne les intervalles de temps entre deux arrivées de clients. Nous noterons l'instant d'arrivée du client n dans le système

$$A_n := \inf \{t > 0 \mid N(t) = n\}$$

7. voir [Foata and Fuchs, 2004], pp. 39-41, pour une démonstration formelle

Par convention, on supposera $A_0 = 0$. Les intervalles inter-arrivées peuvent alors être définis par

$$T_n := A_n - A_{n-1}, \forall n \geq 1$$

Ces définitions des instants d'arrivées et des intervalles inter-arrivées sont illustrées schématiquement à la figure 2.3.

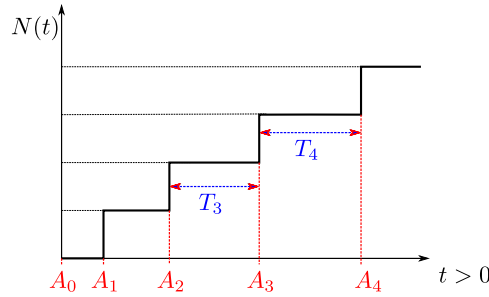


FIGURE 2.3 – Les instants d'arrivées A_n et les intervalles T_n inter-arrivées dans un processus de Poisson

Par définition, les intervalles de temps inter-arrivées sont disjoints. Le processus de Poisson étant un processus à incréments indépendants, les durées des intervalles inter-arrivées sont donc nécessairement des variables aléatoires continues et indépendantes.

Proposition 5 Soit $\{N(t)\}_{t>0}$, un processus de Poisson de paramètre λ . Alors, les intervalles inter-arrivées sont des variables aléatoires exponentielles indépendantes identiquement distribuées :

$$T_n \sim \text{Exp}(\lambda), \forall n \geq 1$$

Démonstration : Considérons $T_n = t - A_{n-1}$ qui mesure le temps entre l'arrivée client n et l'arrivée du client $n - 1$ (pour un choix de $n \geq 1$ particulier, mais quelconque). La probabilité que le client n arrive dans l'intervalle de temps $[\xi, \xi + d\xi]$ s'écrit :

$$\mathbb{P}[T_n \leq \xi + d\xi | T_n > \xi] = \lambda d\xi + o(d\xi)$$

Par définition de la probabilité conditionnelle, on a :

$$\begin{aligned} \mathbb{P}[T_n \leq \xi + d\xi | T_n > \xi] &= \frac{\mathbb{P}[\xi < T_n \leq \xi + d\xi]}{\mathbb{P}[T_n > \xi]} \\ &= \frac{F_{T_n}(\xi + d\xi) - F_{T_n}(\xi)}{1 - F_{T_n}(\xi)} \end{aligned}$$

où $F_{T_n}(\xi) = \mathbb{P}[T_n \leq \xi]$ désigne la fonction de répartition de la variable T_n . En rapprochant les deux derniers résultats, on obtient :

$$\frac{F_{T_n}(\xi + d\xi) - F_{T_n}(\xi)}{d\xi} = \left(\lambda + \frac{o(d\xi)}{d\xi} \right) (1 - F_{T_n}(\xi))$$

Par passage à la limite $d\xi \rightarrow 0$, on obtient ensuite

$$F'_{T_n}(\xi) = \lambda (1 - F_{T_n}(\xi))$$

avec la condition $F_{T_n}(\xi) = 0, \forall \xi \leq 0$ et en particulier la condition initiale $F_{T_n}(0) = 0$. L'unique solution de cette équation différentielle linéaire à coefficient constant (mais non homogène) qui satisfasse la condition initiale $F_{T_n}(0) = 0$ est

$$F_{T_n}(\xi) = \begin{cases} 1 - e^{-\lambda\xi} & \text{pour } x > 0 \\ 0 & \text{pour } x \leq 0 \end{cases}$$

qui n'est autre que la fonction de répartition de la variable $\text{Exp}(\lambda)$. ■

Remarque 2 (deuxième caractérisation) Le résultat 5 est lui aussi une caractérisation des processus de Poisson. Tout processus de comptage, stationnaire, à incréments indépendants, dont les intervalles inter-arrivées sont de variables aléatoires exponentielles de paramètre λ , est nécessairement un processus de Poisson (da paramètre λ).

Si les intervalles inter-arrivées sont des exponentielles indépendantes, les instants d'arrivées des clients successifs, eux, ne sont pas des variables aléatoires indépendantes. En effet, l'instant d'arrivée du client $n \geq 1$, peut s'écrire $A_n = \sum_{k=1}^n T_k$. La somme de $n \geq 1$ variables aléatoires exponentielles indépendantes de paramètre λ est une variable aléatoire de loi Gamma (voir lemme 1, ci-après).

Définition 5 (Loi $\Gamma(n, \lambda)$) Soit $\lambda \in \mathbb{R}, \lambda > 0$ et $n \in \mathbb{N}, n \geq 1$. Une variable aléatoire de loi $X \sim \Gamma(n, \lambda)$ est une variable aléatoire continue, à valeurs réelles, dont la densité de probabilité est donnée par

$$f_X(\xi) := \begin{cases} \frac{\lambda(\lambda\xi)^{n-1}}{(n-1)!} e^{-\lambda\xi} & \text{si } \xi \geq 0 \\ 0 & \text{sinon} \end{cases}$$

Lemme 1 (Somme d'exponentielles indépendantes) Soient X_i ($i \in \{1, \dots, n\}$) des variables aléatoires exponentielles indépendantes de même paramètre $\lambda > 0$. Alors

$$Z_n := \sum_{i=1}^n X_i \sim \Gamma(n, \lambda)$$

Démonstration : On démontre cette proposition par récurrence sur n . Commençons par démontrer le cas “trivial” $n = 2$. On a alors, par définition de la fonction de répartition de la somme de deux variables aléatoires :

$$F_{Z_2}(\xi) := IP[Z_2 \leq \xi] = IP[X_1 + X_2 \leq \xi] = \iint_{x_1 + x_2 \leq \xi} f_{(X_1, X_2)}(x_1, x_2) dx_1 dx_2$$

où $f_{(X_1, X_2)}$ désigne la densité conjointe du couple de variables aléatoires (X_1, X_2) . Par indépendances des variables X_1 et X_2 , on obtient :

$$\begin{aligned} \iint_{x_1 + x_2 \leq \xi} f_{(X_1, X_2)}(x_1, x_2) dx_1 dx_2 &= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\xi - x_2} f_{(X_1, X_2)}(x_1, x_2) dx_1 \right) dx_2 \\ &= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\xi - x_2} f_{X_1}(x_1) dx_1 \right) f_{X_2}(x_2) dx_2 \end{aligned}$$

La densité de la variable Z_2 (somme) s’écrit donc

$$f_{Z_2}(\xi) = \frac{dF_{Z_2}}{d\xi}(\xi) = \int_{-\infty}^{\infty} f_{X_1}(\xi - x_2) f_{X_2}(x_2) dx_2 = (f_1 * f_2)(\xi)$$

Dans le cas de deux variables exponentielles de paramètres λ , on obtient $f_{Z_2}(\xi) = 0$ pour $\xi < 0$ et

$$f_{Z_2}(\xi) = \int_0^{\xi} \lambda e^{-\lambda(\xi - x_2)} \lambda e^{-\lambda x_2} dx_2 = \lambda^2 \xi e^{-\lambda \xi}$$

dans le cas $\xi \geq 0$. On a donc bien $Z_2 \sim \Gamma(2, \lambda)$. Démontrons maintenant la récurrence. Pour ce faire, nous supposons que la propriété est vraie au rang n , c’est-à-dire :

$$Z_n := \sum_{i=1}^n X_i \sim \Gamma(n, \lambda)$$

On a alors

$$Z_{n+1} := Z_n + X_{n+1}$$

avec $X_{n+1} \sim \text{Exp}(\lambda)$. Par hypothèse, les variables Z_n et X_{n+1} sont indépendantes.

On a donc :

$$f_{Z_{n+1}}(\xi) = (f_n * f_2)(\xi) = (f_2 * f_n)(\xi)$$

où $f_2(\xi)$ est la densité de la loi exponentielle. Par hypothèse de récurrence, $f_n(\xi)$ est la densité d’une loi $\Gamma(n, \lambda)$ et la convolution précédente s’écrit (pour $\xi > 0$)

$$\begin{aligned} f_{Z_{n+1}}(\xi) &= \int_0^{\xi} \lambda e^{-\lambda(\xi - x_2)} \frac{\lambda (\lambda x_2)^{n-1}}{(n-1)!} e^{-\lambda x_2} dx_2 \\ &= \frac{\lambda^2 e^{-\lambda \xi}}{(n-1)!} \int_0^{\xi} (\lambda x_2)^{n-1} dx_2 \\ &= \frac{\lambda (\lambda \xi)^n}{n!} e^{-\lambda \xi} \end{aligned}$$

qui n'est autre que la fonction de répartition de la variable $\Gamma(n+1, \lambda)$. Nous avons donc montré que la propriété est vraie au rang $n+1$ (en supposant qu'elle est vraie au rang n). ■

Proposition 6 (Distributions des instants d'arrivées) Soit $\{N(t)\}_{t>0}$, un processus de Poisson de paramètre λ . Alors A_n , l'instant d'arrivée du n -ième client, est distribué selon une loi Gamma :

$$A_n \sim \Gamma(n, \lambda), \forall n \geq 1$$

Démonstration : C'est une conséquence immédiate du lemme 1, l'instant d'arrivée $A_n = \sum_{k=1}^n T_k$ étant la somme des variables variables exponentielles T_k , indépendantes et de paramètres λ . ■

Remarque 3 La variable $\Gamma(n, \lambda)$ étant la somme de n variables aléatoires indépendantes, il est aisé de calculer l'espérance et la variance des instants d'arrivées A_n . On a

$$E(A_n) = E\left(\sum_{k=1}^n T_k\right) = \frac{n}{\lambda}$$

et

$$\text{var}(A_n) = \text{var}\left(\sum_{k=1}^n T_k\right) = \frac{n}{\lambda^2}$$

car les variables T_k sont indépendantes. La distribution de l'instant d'arrivée du client n est donc décalée vers la droite et de plus en plus dispersée lorsque n augmente. La figure 2.4 représente les distributions de lois $\Gamma(k, \theta)$ pour différentes valeurs de k et θ

2.3 Âge et temps de vie résiduel

D'après les définitions de la section précédente, $A_{N(t)}$ désigne l'instant de la dernière arrivée avant le temps t , tandis que $A_{N(t)+1}$ désigne l'instant de la prochaine arrivée après le temps t . Prenons l'exemple d'un processus de Poisson où les arrivées sont les pannes d'une machine et où la machine en fonctionnement est remplacée par une neuve à chaque panne. L'âge de la machine au temps t est donc la durée $T_a := t - A_{N(t)}$ depuis la dernière panne. Le temps de vie résiduel est le temps restant avant la prochaine panne, c'est-à-dire $T_r := A_{N(t)+1} - t$. Ces deux variables aléatoires sont représentées schématiquement à la figure 2.5.

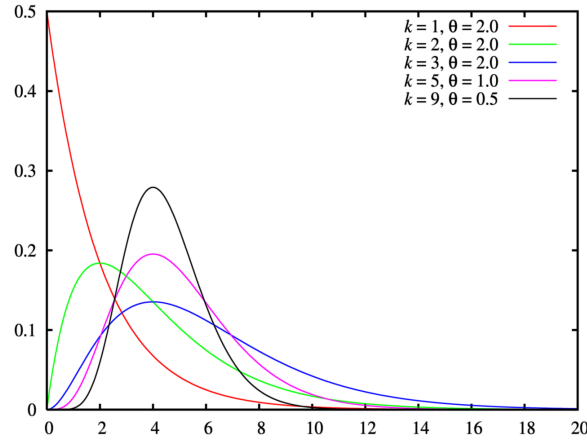


FIGURE 2.4 – Distributions de probabilités de lois $\Gamma(k, \theta)$ pour différentes valeurs des paramètres k et θ

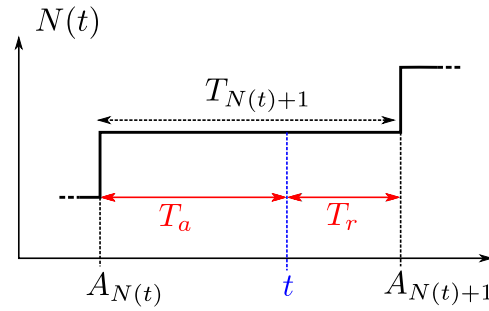


FIGURE 2.5 – Représentation schématique des variables aléatoires *âge* et *temps de vie résiduel*, évaluées au temps t , pour une réalisation particulière d'un processus de Poisson

Proposition 7 (âge et temps de vie résiduel) Soit $\{N(t)\}_{t>0}$, un processus de Poisson de paramètre λ . Alors, l'âge au temps t défini par

$$T_a := t - A_{N(t)}$$

et la *durée de vie résiduelle* définie par

$$T_r := A_{N(t)+1} - t$$

sont deux variables aléatoires indépendantes. De plus, la durée de vie résiduelle est distribuée selon une loi exponentielle :

$$T_r \sim \text{Exp}(\lambda)$$

La fonction de répartition de l'âge au temps t est donnée par

$$F_{T_a}(\xi) := \mathbb{P}[T_a \leq \xi] = \begin{cases} 0 & \text{pour } \xi \leq 0 \\ 1 - e^{-\lambda\xi} & \text{pour } 0 \leq \xi \leq t \\ 1 & \text{pour } t \leq \xi \end{cases}$$

Démonstration : Par construction, les variables aléatoires $T_a = t - A_{N(t)}$ et $T_r = A_{N(t)+1} - t$ sont liées aux arrivées sur les intervalles de temps $]0, t[$ et $]t, \infty[$ disjoints. Par définition du processus de Poisson, ces arrivées correspondent donc à des incréments indépendants. Par suite, les variables aléatoires $T_a = t - A_{N(t)}$ et $T_r = A_{N(t)+1} - t$ sont donc indépendantes.

La distribution exponentielle de $T_r = A_{N(t)+1} - t$ résulte de l'absence de mémoire de la loi exponentielle ou, de manière équivalente, du caractère stationnaire du processus de Poisson. Le fait qu'il s'est déjà écoulé un certain temps $T_a = t - A_{N(t)}$ depuis la dernière arrivée n'a pas d'influence sur la distribution du temps qui reste jusqu'à la prochaine arrivée. On peut formellement démontrer la distribution exponentielle de $T_r = A_{N(t)+1} - t$ en passant par sa fonction de répartition :

$$\begin{aligned} F_{T_r}(\xi) &:= \mathbb{P}[T_r \leq \xi] \\ &= 1 - \mathbb{P}[T_r > \xi] \\ &= 1 - \mathbb{P}[0 \text{ arrivée sur } [t, t + \xi]] \\ &= 1 - e^{-\lambda\xi} \end{aligned}$$

qui est bien la fonction de répartition d'une variable aléatoire exponentielle de paramètre λ .

Par construction, la variable $T_a := t - A_{N(t)}$ prend ses valeurs dans l'intervalle $[0, t]$. On a, pour tout $\xi \in [0, t]$

$$\begin{aligned} F_{T_a}(\xi) &:= \mathbb{P}[T_a \leq \xi] \\ &= 1 - \mathbb{P}[T_a > \xi] \\ &= 1 - \mathbb{P}[t - A_{N(t)} > \xi] \\ &= 1 - \mathbb{P}[A_{N(t)} < t - \xi] \\ &= 1 - \mathbb{P}[0 \text{ arrivée sur } [t - \xi, t]] \\ &= 1 - \mathbb{P}[0 \text{ arrivée sur } [0, \xi]] \\ &= 1 - e^{-\lambda\xi}, \end{aligned}$$

ce qui achève la démonstration. ■

Le paradoxe de l'inspection

Illustrons ce paradoxe sur un exemple. On suppose disposer d'un stock d'ampoules de durées de vie indépendantes et distribuées selon des lois

$T_i \sim \text{Exp}(\lambda)$ (≥ 1) identiques et indépendantes. Lorsqu'une ampoule s'éteint, on la remplace par une nouvelle. Les arrivées considérées sont les instants où les ampoules s'éteignent. Le processus $\{N(t)\}_{t \geq 0}$ qui compte les ampoules consommées est donc un processus de Poisson de paramètre λ (car les intervalles entre deux arrivées sont distribués selon des variables exponentielles). L'espérance de la durée de vie d'une ampoule est

$$E(T_i) = \frac{1}{\lambda}$$

Le paramètre λ (exprimé par exemple en heure⁻¹) est donc le nombre moyen d'ampoules qui tombent en panne par unité de temps (dans l'exemple, par heure). On se propose d'estimer ce taux λ par une inspection à l'instant t . A l'instant t , on examine le système et on note la durée de vie de l'ampoule qui brûle à ce moment là. Ceci revient, du fait de l'absence de mémoire de la variable aléatoire exponentielle, à étudier la variable aléatoire $T_{N(t)+1}$. On peut écrire :

$$T_{N(t)+1} = A_{N(t)+1} - A_{N(t)} = t - A_{N(t)} + A_{N(t)+1} - t = T_a(t) + T_r(t)$$

où les deux variables aléatoires $T_a(t)$ et $T_r(t)$ sont indépendantes. On a alors

$$E(T_{N(t)+1}) = E(T_a(t)) + E(T_r(t))$$

avec $E(T_r(t)) = \frac{1}{\lambda}$ et

$$E(T_a(t)) = \int_0^t e^{-\lambda\xi} d\xi = \frac{1 - e^{-\lambda t}}{\lambda}$$

Au final, l'espérance de vie de l'ampoule inspectée au temps t vaut donc

$$E(T_{N(t)+1}) = \frac{1}{\lambda} + \frac{1 - e^{-\lambda t}}{\lambda} = \frac{2}{\lambda} - \frac{e^{-\lambda t}}{\lambda} > \frac{1}{\lambda}$$

La durée de vie moyenne de l'ampoule inspectée à l'instant t est strictement supérieure à la durée de vie moyenne des ampoules non inspectées. Tout se passe comme si l'observation de l'ampoule à l'instant t produit une perturbation qui dilate la durée de vie moyenne de l'ampoule inspectée. Une explication à posteriori de ce paradoxe est la suivante. Parmi les ampoules, il y en a qui brûlent plus longtemps que les autres. Lors de l'inspection au temps t , on a donc plus de chance de tomber sur une ampoule qui brûle longtemps que sur une ampoule qui brûle moins longtemps.

Remarque 4 Lorsque l'instant d'observation $t \rightarrow \infty$

$$E(T_{N(t)+1}) \rightarrow \frac{2}{\lambda}$$

La durée de vie moyenne de l'ampoule observée est donc pratiquement le double de la durée de vie moyenne des autres ampoules. La loi de probabilité de la durée de vie de l'ampoule inspectée est donnée par

$$f_{T_{N(t)+1}}(\xi) = (f_{T_a(t)} * f_{T_r(t)})(\xi)$$

Lorsque $t \rightarrow \infty$, la loi de $T_{N(t)+1}$ tend vers la convolution de deux exponentielles indépendantes, c'est-à-dire vers une loi $\Gamma(2, \lambda)$.

2.4 Décomposition probabiliste d'un processus de Poisson

Considérons l'exemple de l'arrivées de véhicules à une barrière de péage d'autoroute. Supposons que ces véhicules arrivent selon un processus de Poisson $\{N(t)\}_{t>0}$ de paramètre λ . Le gestionnaire souhaite pouvoir caractériser les arrivées des différents types de véhicules : motos, autos et camions. Dans un premier temps, on suppose que les probabilités des différentes catégories ne varient pas au cours du temps. Nous noterons les probabilités associées aux différents types de véhicules de la manière suivante :

$$\begin{aligned} p_a &:= \mathbb{P}[\text{auto} \mid \text{un véhicule est arrivé}] \\ p_m &:= \mathbb{P}[\text{moto} \mid \text{un véhicule est arrivé}] \\ p_c &:= \mathbb{P}[\text{camion} \mid \text{un véhicule est arrivé}] \end{aligned}$$

Nous supposons que les arrivées des différents types de véhicules sont des processus indépendants⁸ que nous noterons $\{N_i(t)\}_{t>0}$ avec $i \in \{a, m, c\}$. La question que se pose le gestionnaire est donc : “comment caractériser les processus $\{N_i(t)\}_{t>0}$?” afin, ensuite, de pouvoir effectuer les calculs utiles basés sur cette description décomposée, par exemple le calcul des tarifs pour chaque type de véhicule qui garantit (avec une certaine précision et un taux de confiance donnés) un revenu global déterminé.

Proposition 8 (décomposition probabiliste) Soit $\{N(t)\}_{t>0}$, un processus de Poisson de paramètre λ qui se décompose de façon probabiliste en K processus, chaque nouvelle arrivée participant au processus $k \in \{1, \dots, K\}$ avec une probabilité p_k . Alors, chacun des processus $\{N_k(t)\}_{t>0}$ d'arrivées de type k est un processus de Poisson de paramètre λp_k .

8. On pourrait imaginer à contrario que moins de camions circulent lorsque la densité de voitures est importante, par exemple le dimanche ou les jours fériés. Mais nous nous concentrons ici sur le cas simple où les différents processus d'arrivées sont indépendants les uns des autres.

Démonstration : Il suffit d'utiliser la première caractérisation des processus de Poisson. Par définition de probabilité conditionnelle, on a :

$$\begin{aligned}
 & IP [N_k(t + dt) = n + 1 | N_k(t) = n] \\
 &= IP [\text{arrivée de type } k | \text{arrivée sur } [t, t + dt]] \cdot IP [\text{arrivée sur } [t, t + dt]] \\
 &= p_k \cdot IP [\text{arrivée sur } [t, t + dt]] \\
 &= p_k \cdot (\lambda dt + o(dt)) \\
 &= (\lambda p_k) dt + o(dt)
 \end{aligned}$$

Par la première caractérisation des processus de Poisson, nous pouvons en déduire que $\{N_k(t)\}_{t \geq 0}$ est un processus de Poisson de paramètre λp_k ■

Remarque 5 (décomposition non stationnaire) Dans de nombreux exemples, la répartition des différents types de clients varie avec le temps. On a alors

$$IP [\text{arrivée de type } k | \text{arrivée sur } [t, t + dt]] =: p_k(t)$$

Par exemple, si les arrivées du processus de Poisson $\{N(t)\}_{t \geq 0}$ correspondent à la production de pièces par une machine outil sur une ligne de production, et si $p_1(t)$ désigne la proportion de pièces défectueuses parmi celles produites au temps t et $p_2(t) := 1 - p_1(t)$ le proportion de pièces conformes au cahier des charges, il est naturel (dans certains cas) de considérer une probabilité $p_1(t)$ qui augmente avec t (usure de la machine, échauffement des pièces, etc.). Dans un tel cas, les processus d'arrivées de pièces défectueuses et conformes ne sont plus des processus de Poisson stationnaires : les taux $\lambda p_1(t)$ et $\lambda p_2(t)$ varient au cours du temps. Il est possible d'étendre la définition de processus de Poisson au cas non stationnaire (on dit aussi non homogène, voir section 2.6). Dans l'exemple, les processus d'arrivées $N_1(t)$ et $N_2(t)$, respectivement de pièces défectueuses et conformes, sont ainsi deux processus de Poisson non homogènes. En particulier, pour tout $0 < s < t$, les nombres de pièces respectivement défectueuses et conformes produites sur l'intervalle de temps $[s, t]$ sont deux variables aléatoires de Poisson de paramètres respectifs

$$\mu_1^{(s,t)} := \int_s^t \lambda p_1(\xi) d\xi$$

et

$$\mu_2^{(s,t)} := \int_s^t \lambda p_2(\xi) d\xi$$

La démonstration de ce résultat est la même que celle de la proposition 8 pour le cas stationnaire, mais elle utilise le résultat de caractérisation des processus de Poisson non homogène de la section 2.6 (équations (2.2) et (2.3)). Par ailleurs, la loi conditionnelle du nombre $N_1^{\{N(t)=n\}}$ de pièces défectueuses sur $[0, t]$, sachant que le nombre total de pièces produites est

$N(t) = n$, est une loi binomiale $B(n, \bar{p}_1)$ où \bar{p}_1 est la probabilité moyenne sur $[0, t]$ définie par

$$\bar{p}_1 := \frac{1}{t} \int_0^t p_1(\xi) d\xi$$

La démonstration est une conséquence immédiate de la remarque 7 dans la section section 2.6 consacrée aux processus de Poisson non homogènes.

Inversément à ce qui précède, dans de nombreuses applications, on est intéressé au processus global qui résulte de la superposition de processus de Poisson. Pour continuer avec un exemple dans le domaine du trafic routier (évidemment transposable à d'autres problèmes de trafic sur des réseaux), notre gestionnaire de réseau autoroutier peut s'intéresser aux caractéristiques de trafic sur un tronçon d'autoroute, sachant que les processus d'arrivées des véhicules à la barrière de péage d'entrée et aux différentes bretelles d'accès du tronçon ont été préalablement caractérisés comme des processus de Poisson de paramètres λ_i , $i \in \{1, \dots, N\}$.

Proposition 9 (superposition) Soient $\{N_i(t)\}_{t>0}$, $i \in \{1, \dots, N\}$, des processus de Poisson indépendants de paramètres respectifs $\lambda_i > 0$ qui se superpose en un seul processus $\{N(t)\}_{t>0}$ total défini par

$$N(t) = \sum_{i=1}^N N_i(t)$$

Alors $\{N(t)\}_{t>0}$ est un processus de Poisson de paramètre $\lambda := \sum_{i=1}^N \lambda_i$.

Démonstration : Par construction $\{N(t)\}_{t>0}$ est un processus de comptage stationnaire à incréments indépendants. Il reste donc à montrer que le nombre total d'arrivée sur $[0, t]$ suit une loi de Poisson de paramètre $\lambda t = \sum_{i=1}^N \lambda_i t$. Ce résultat découle immédiatement du lemme 2 présenté ci-après

■

Lemme 2 (somme de lois de Poisson) Soient X_1 et X_2 deux variables de Poisson indépendantes, de paramètres respectifs λ_1 et λ_2 . Alors, la somme $X := X_1 + X_2$ est une variable de Poisson de paramètre $\lambda := \lambda_1 + \lambda_2$.

Démonstration : En appliquant la formule des probabilités totales, on obtient, pour tout $k \geq 0$:

$$\begin{aligned} \mathbb{P}[X = k] &= \sum_{j=0}^k \mathbb{P}[X = k | X_1 = j] \cdot \mathbb{P}[X_1 = j] \\ &= \sum_{j=0}^k \mathbb{P}[X_2 = k - j | X_1 = j] \cdot \mathbb{P}[X_1 = j] \end{aligned}$$

Par indépendance de X_1 et X_2 , on peut écrire :

$$\mathbb{P}[X_2 = k - j | X_1 = j] \cdot \mathbb{P}[X_1 = j] = \mathbb{P}[X_2 = k - j] \cdot \mathbb{P}[X_1 = j]$$

Comme X_1 et X_2 sont des variables de Poisson, on obtient finalement :

$$\begin{aligned} \mathbb{P}[X = k] &= \sum_{j=0}^k \frac{(\lambda_2)^{k-j}}{(k-j)!} e^{-\lambda_2} \cdot \frac{(\lambda_1)^j}{j!} e^{-\lambda_1} \\ &= \frac{e^{-(\lambda_1+\lambda_2)}}{k!} \sum_{j=0}^k \frac{k!}{(k-j)!j!} (\lambda_1)^j (\lambda_2)^{k-j} \\ &= \frac{(\lambda_1 + \lambda_2)^k e^{-(\lambda_1+\lambda_2)}}{k!} \end{aligned}$$

qui n'est autre qu'une loi de Poisson de paramètre $(\lambda_1 + \lambda_2)$

■

Le lemme 2 s'étend facilement, par récurrence, au cas de la somme d'un nombre arbitraire de variables de Poisson indépendantes.

2.5 Estimation de la densité d'un processus de Poisson

Une question d'un grand intérêt pratique est la suivante : disposant d'une suite d'observations sous la forme du nombre n d'arrivées sur $[0, T]$ et des instants $0 \leq a_1, a_2, \dots, a_n \leq T$ de ces arrivées, ayant décidé de modéliser le processus d'arrivée par un processus de Poisson, comment choisir au mieux le paramètre λ . Il s'agit donc de proposer un estimateur $\hat{\lambda}$ dont la réalisation empirique produise une estimation "intéressante" pour le paramètre λ . Nous allons construire cet estimateur en appliquant la méthode du *maximum de vraisemblance* : il s'agit de choisir la valeur de λ qui rende maximale la probabilité d'observer les valeurs n et a_1, a_2, \dots, a_n , supposant que le processus d'arrivée est un processus de Poisson de paramètre λ . Cette probabilité est proportionnelle à la fonction de vraisemblance

$$L(n, a_1, a_2, \dots, a_n; \lambda) := \mathbb{P}[N(T) = n] \cdot g(a_1, a_2, \dots, a_n | N(T) = n)$$

où $g(a_1, \dots, a_n | N(T) = n)$ est la densité conditionnelle de probabilité du n -uplet de variables aléatoires continues (A_1, \dots, A_n) , sachant que l'évènement $N(T) = n$ est réalisé. Or, si le nombre total d'arrivées sur $[0, T]$ est connu, alors les instants d'arrivées sont des variables aléatoires indépendantes de distribution uniforme. On a donc :

$$g(a_1, a_2, \dots, a_n | N(T) = n) = \begin{cases} \frac{n!}{T^n} & \text{si } a_1, \dots, a_n \in [0, T] \\ 0 & \text{sinon} \end{cases}$$

La factorielle $n!$ prend en compte toutes les permutations possibles des clients qui correspondent à la même séquence (a_1, a_2, \dots, a_n) des instants d'arrivées. La fonction de vraisemblance à maximiser s'écrit donc :

$$L(n, a_1, a_2, \dots, a_n; \lambda) = \frac{e^{-\lambda T} (\lambda T)^n}{n!} \cdot \frac{n!}{T^n} = \lambda^n e^{-\lambda T}$$

Cette fonction est maximale lorsque son logarithme

$$\ln L = n \ln(\lambda) - \lambda T$$

est, lui aussi, maximal, c'est-à-dire lorsque

$$\frac{\partial}{\partial \lambda} \ln L = \frac{n}{\lambda} - T = 0$$

On obtient ainsi pour l'estimateur par maximum de vraisemblance

$$\hat{\lambda} = \frac{N(T)}{T}$$

Il est remarquable que l'estimateur $\hat{\lambda}$ ne dépende pas des instants d'arrivées observés (a_1, a_2, \dots, a_n) , mais uniquement du nombre total d'arrivées sur $[0, T]$. Fort simple, cet estimateur possède néanmoins des propriétés remarquables : il s'agit d'un estimateur non biaisé (i.e. $E(\hat{\lambda}) = \lambda$) et parmi tous les estimateurs non biaisés, il est celui de variance minimale (voir [Foata and Fuchs, 2004], section 4.7). Autrement dit, la distribution de cet estimateur présente une dispersion minimale autour de la "vraie valeur" λ .

Remarque 6 (estimation à partir des intervalles inter-arrivées X_n)

D'autres expériences sont possibles en vue de l'estimation du paramètre λ . On peut par exemple observer le processus jusqu'à l'arrivée du $n^{\text{ième}}$ client (n étant fixé à priori) et noter les instants d'arrivées ou, de manière équivalente, les intervalles inter-arrivées (x_1, x_2, \dots, x_n) observés. La fonction de vraisemblance s'écrit alors

$$L(x_1, x_2, \dots, x_n; \lambda) = \prod_{k=1}^n \lambda e^{-\lambda x_k} = \lambda^n e^{-\lambda(x_1 + \dots + x_n)}$$

Cette fonction est maximale lorsque

$$\frac{\partial}{\partial \lambda} \ln L = \frac{n}{\lambda} - (x_1 + \dots + x_n) = 0$$

c'est-à-dire pour

$$\hat{\lambda} = \frac{n}{x_1 + \dots + x_n}$$

Comme $x_1 + \dots + x_n$ n'est rien d'autre que l'instant d'arrivée a_n du client n , l'estimateur par maximum de vraisemblance donne dans ce cas

$$\hat{\lambda} = \frac{n}{A_n}$$

Il n'est pas nécessaire de noter tous les intervalles inter-arrivées. L'instant d'arrivée a_n du $n^{\text{ième}}$ client suffit.

2.6 Processus de Poisson non homogène

Les hypothèses qui définissent un processus de Poisson conduisent aux deux caractérisations qui rendent ce processus simple à analyser, que ce soit en terme de répartition dans le temps des arrivées ou en nombre d'arrivées sur un intervalle donné. Dans la plupart des applications réelles, cependant, ces hypothèses ne sont pas satisfaites. Prenons l'exemple d'un atelier de production dans lequel des machines peuvent tomber en panne (ce sont les événements ou "arrivées" qui nous intéressent). Supposer que les pannes surviennent selon un processus de Poisson stationnaire revient à supposer, entre autres, que la probabilité d'une panne est indépendante de la durée de fonctionnement, c'est-à-dire de l'âge ou de l'usure de la machine. Ceci n'est bien sûr vrai qu'en première approximation, pour des durées de simulation relativement courtes par rapport à la durée de vie moyenne d'une machine. Lorsque ces pannes surviennent plus fréquemment, une première approche consiste à subdiviser le processus d'arrivée en une suite de processus de Poisson stationnaires consécutifs, avec un taux d'arrivée $\lambda(t)$ constant sur chacun de ces processus individuels (modélisation adaptative). Une deuxième approche, plus directe, consiste à analyser un processus de Poisson à accroissements indépendants, mais non stationnaire. C'est cette seconde approche que nous développons ici. Nous allons montrer comment généraliser les caractérisations obtenues pour les processus de Poisson stationnaires au cas où le taux d'arrivée $\lambda(t)$ dépend explicitement du temps.

Définition 6 (Processus de Poisson non homogène) Soit

$$\lambda : \mathbb{R}_+ \rightarrow \mathbb{R}_+$$

une fonction continue (taux d'arrivée instantané). Alors, le processus stochastique $\{N(t)\}_{t \geq 0}$ à temps continu, à valeurs dans $E = \mathbb{N}$ est un *processus de Poisson non homogène* de densité $\lambda(t)$ si et seulement si :

- (i) $\{N(t)\}_{t \geq 0}$ est un *processus de comptage*
- (ii) $\{N(t)\}_{t \geq 0}$ est un *processus à incréments indépendants*
- (iii) pour tout $0 < s < t$, le nombre $N(t) - N(s)$ d'arrivées sur l'intervalle de temps $[s, t]$ est une variable aléatoire de Poisson de paramètre

$$\mu^{(s,t)} := \int_s^t \lambda(\xi) d\xi$$

Beaucoup de propriétés des Processus de Poisson stationnaires se généralisent au cas non homogène en remplaçant λt , le nombre moyen d'arrivées sur $[0, t]$, par

$$\mu(t) := \int_0^t \lambda(\xi) d\xi$$

Comme $N(t)$ est un processus de comptage, on a $N(0) = 0$. Dès lors, $N(t)$ est une variable de Poisson de paramètre $\mu(t)$. On a donc

$$E(N(t)) = \mu(t) ; \text{ var } (N(t)) = \mu(t) \quad (2.1)$$

La fonction $\mu(t)$ est continûment différentiable, à valeurs positives, monotone non décroissante et telle que $\mu(0) = 0$. Elle est appelée *fonction moyenne* du processus, ou encore *fonction de renouvellement*. Par définition du processus de Poisson, on a pour tout $h > 0$:

$$\begin{aligned} IP[N(t+h) - N(t) = 1] &= \frac{(\mu(t+h) - \mu(t))^1}{1!} e^{-(\mu(t+h) - \mu(t))} \\ &= (\mu(t) + h\mu'(t) + o(h)) - \mu(t) e^{-(\mu(t+h) - \mu(t))} \\ &= h\lambda(t) + o(h) \end{aligned} \quad (2.2)$$

et

$$\begin{aligned} IP[N(t+h) - N(t) \geq 2] &= \sum_{k \geq 2} \frac{(\mu(t+h) - \mu(t))^k}{k!} e^{-(\mu(t+h) - \mu(t))} \\ &= \left(e^{(\mu(t+h) - \mu(t))} - 1 - (\mu(t+h) - \mu(t)) \right) e^{-(\mu(t+h) - \mu(t))} \\ &= 1 - e^{-(\mu(t+h) - \mu(t))} (1 + (\mu(t+h) - \mu(t))) \\ &= 1 - e^{-(h\lambda(t) + o(h))} (1 + (h\lambda(t) + o(h))) \\ &= 1 - (1 - (h\lambda(t) + o(h))) (1 + (h\lambda(t) + o(h))) \\ &= h^2 \lambda^2(t) + o(h) \\ &= o(h) \end{aligned} \quad (2.3)$$

Les propriétés (2.1) (2.2) et (2.3) sont des généralisations immédiates de propriétés analogues obtenues dans le cas stationnaire. En fait, de manière générale, l'analyse d'un processus de Poisson non homogène peut toujours se ramener à celle d'un processus de Poisson stationnaire.

Proposition 10 (changement d'horloge) Tout processus de Poisson non homogène $\{N(t)\}_{t \geq 0}$, de densité $\lambda(t) > 0$ et de fonction de renouvellement $\mu(t)$, peut se ramener à un processus de Poisson stationnaire $\{N^*(t)\}_{t \geq 0}$ défini par le changement d'horloge

$$N^*(t) := N(\mu^{-1}(t))$$

Démonstration : En effet, dans le cas où $\lambda(t)$ est une fonction à valeur strictement positives, $\mu(t)$ est une fonction continue (et même continûment différentiable) strictement croissante sur \mathbb{R}_+ avec $\mu(0) = 0$. Elle est donc bijective et $\mu^{-1}(t)$ est elle aussi continûment différentiable, strictement croissante et telle que $\mu^{-1}(0) = 0$. Le processus $\{N^*(t)\}_{t \geq 0}$ est donc un processus

de comptage, à incréments indépendants. De plus, pour tout $0 < s < t$, le nombre $N^*(t) - N^*(s)$ d'arrivées sur l'intervalle de temps $[s, t]$ est une variable aléatoire de Poisson de paramètre

$$\mu^{(s,t)} := \int_{\mu^{-1}(s)}^{\mu^{-1}(t)} \lambda(\xi) d\xi$$

En posant $\eta := \mu(\xi)$, on trouve

$$\mu^{(s,t)} := \int_s^t \lambda(\mu^{-1}(\eta)) d(\mu^{-1}(\eta))$$

Or $\mu(\xi)$ est la primitive de $\lambda(\xi)$. On a donc

$$\mu^{(s,t)} := \int_s^t \frac{d}{d\eta} (\mu(\mu^{-1}(\eta))) d\eta = \int_s^t \frac{d\eta}{d\eta} d\eta = (t - s)$$

Ainsi, $\{N^*(t)\}_{t \geq 0}$ est bien un processus de Poisson de paramètre constant $\lambda = 1$. ■

Remarque 7 (distribution conditionnelle des instants A_k) Dans le cas non homogène, aussi bien que dans le cas stationnaire, si le nombre total n d'arrivées sur l'intervalle $[s, t]$ (avec $0 < s < t$) est connu, nous pouvons appeler A_k l'instant d'arrivée⁹ du client k et conclure que les variables aléatoires A_k sont n variables aléatoires indépendantes distribuées selon la même loi. En appliquant le changement d'horloge de la proposition précédente, nous arrivons à la conclusion que la fonction de répartition (commune) de ces instants d'arrivée est la fonction :

$$IP[A_k \leq u | N(t) - N(s) = n] = F(\xi) \begin{cases} 0 & \text{pour } \xi \leq s \\ \frac{\mu(\xi) - \mu(s)}{\mu(t) - \mu(s)} & \text{pour } s \leq \xi \leq t \\ 0 & \text{pour } t \leq \xi \end{cases}$$

quel que soit $k \in \{1, \dots, n\}$. On en déduit :

$$IP[N(t') - N(s') = k | N(t) - N(s) = n] = \binom{n}{k} \left[\frac{\mu(t') - \mu(s')}{\mu(t) - \mu(s)} \right]^k \left[1 - \frac{\mu(t') - \mu(s')}{\mu(t) - \mu(s)} \right]^{n-k}$$

quels que soient (s, s', t, t') avec $0 < s < s' < t' < t$ et $0 \leq k \leq n$.

9. L'indice $k \in \{1, \dots, n\}$ est juste un label qui désigne le client k et non pas son ordre d'arrivée

2.7 Exercices

Exercice 1

Robert est un grand romantique. Pour conquérir le coeur de Stéphanie il a décidé de sortir le grand jeu et de l'emmener contempler les étoiles filantes. Pour préparer sa sortie, et par commodité, il décide de modéliser le processus d'apparition des étoiles filantes par un processus de Poisson, $N(t)$ (comptant le nombre d'étoiles filantes apparues entre le début de l'observation et le temps t , exprimé en minutes), de paramètre λ . Lors d'une observation préparatoire, la veille du jour j , il a compté 30 étoiles filantes entre 22h00 et 23h00. Il décide donc de tenter sa chance le lendemain, à la même heure.

1. Calculez le temps moyen que Stéphanie et Robert attendront entre 2 étoiles filantes. En déduire λ
2. Ce qui est capital se dit Robert c'est qu'il y ait tout de suite des étoiles, sinon elle va s'impatiser. Calculez la probabilité qu'il y ait une étoile dans les 5 premières minutes.
3. Calculer la probabilité que la première étoile apparaisse avant 22h06
4. Calculez la probabilité pour que la dixième étoile arrive après 23 heures
5. Stéphanie est jolie mais pas très romantique ; elle voudra partir si au bout de 10 minutes il n'y a pas d'étoiles. calculez la probabilité de ce (sinistre) évènement
6. Sachant que 20 étoiles sont apparues dans la première heure, calculez la probabilité qu'exactly 10 étoiles sont apparues dans le deuxième quart d'heure.

Exercice 2

On suppose que des foies arrivent à l'hôpital suivant un processus de Poisson de paramètre λ . Deux malades attendent d'être greffés. Leurs durées de vie suivent des lois exponentielles indépendantes entre elles (et indépendantes du processus d'arrivée des foies), respectivement de paramètres λ_1 et λ_2 . Le premier foie va au premier patient, si celui est encore en vie ; sinon au second (s'il est encore en vie...).

1. Calculer la probabilité que le premier patient soit greffé
2. Calculer la probabilité que le second patient soit greffé

Exercice 3

Les élèves de 1ere année IR&C (ils sont 50 en tout) arrivent en retard à un cours en suivant un processus de Poisson de paramètre $\lambda = 4$. On compte le temps en minutes.

1. Ce jour là le professeur est de bonne humeur, il se contente de faire une remarque chaque fois que 3 élèves sont passés. On note Z_n le temps qui sépare la $(n - 1)^{ieme}$ remarque de la n^{ieme} . Calculer la loi de Z_n

2. Ce jour là le professeur est de mauvaise humeur, il décide de ne plus accepter les élèves se présentant avec plus de 5 min de retard. On note Z le nombre d'élèves refusés au cours. Calculer la loi de Z

Exercice 4

Robert a trouvé un job de standardiste, on suppose que les appels téléphoniques se succèdent à ce standard selon un processus de Poisson de paramètre λ . De 8h à 12h, il y a eu 10 appels cette matinée

1. Calculer la probabilité que 5 appels arrivent dans la première heure, par deux méthodes.
2. Robert s'est absenté de 10h à 10h 15 pour draguer sa collègue. Quelle est la probabilité pour que Robert aie raté un appel ?

Exercice 5

Un signal $X(t)$ ne peut prendre que les valeurs 0 ou 1. Les instants auxquels il change de valeur définissent un processus de Poisson de paramètre λ . On suppose qu'à l'instant $t = 0$ le signal vaut 1. Calculer $P(X(t) = 1)$

Exercice 6

Le nombre de personnes franchissant une barrière automatique de métro à l'heure de pointe est modélisé par un processus de Poisson. Il arrive une personne toutes les 2 secondes en moyenne. Parmi les personnes certaines fraudent en n'ayant pas de ticket. La probabilité pour qu'une personne soit un fraudeur est $p = 0.1$.

1. Modéliser cette situation
2. Quelle est la probabilité qu'au moins une personne fraude dans un intervalle d'une minute ?
3. Sachant que 200 personnes sont passées en 12 minutes, quel est le nombre moyen de fraudeurs ?
4. 60 personnes sont passées en 8 minutes. Quelle est la probabilité que 5 fraudeurs soient passés dans les 5 premières minutes
5. Chaque fraudeur fait perdre 1 euro à la compagnie de transports. Combien la compagnie perd-elle en moyenne à cette barrière en une heure ?

Exercice 7

On suppose que les clients d'un magasin arrivent dans un magasin selon un processus de Poisson au rythme de 3 par minute. Le temps de séjour d'un client dans le magasin suit une loi exponentielle de paramètre $\mu = 0.1$. Le temps que passe chaque client dans le magasin est indépendant du temps qu'y passent les autres clients. On note $N(T)$ le nombre de clients encore dans le magasin à l'instant $T = 30$ min. On note $N_p(T)$ le nombre de clients partis à l'instant T . Donner la loi de ces deux variables aléatoires.

Exercice 8

On modélise le téléchargement d'un fichier par l'envoi de segments. Lorsqu'on télécharge un fichier, la transmission de ces segments peut être erronée.

La liaison téléphonique entre Robert et son fournisseur d'accès internet est mauvaise : on suppose que des erreurs de transmission se produisent selon un processus de Poisson. Le temps moyen entre deux erreurs de transmission est de 1 minute. Robert télécharge un fichier sur un site internet ; Entre le FAI de Robert et ce site les erreurs de transmission sont également modélisées par un processus de Poisson. Elles sont en moyenne espacées de 5 minutes.

1. Le téléchargement dure une demi heure. En moyenne, combien de segments sont erronés parmi ceux qu'a reçus Robert ?
2. Quelle est la probabilité que le téléchargement se déroule sans accroc ?

Exercice 9

On suppose que les arrivées d'un bus à un arrêt suivent un processus de Poisson de paramètre λ .

Soit $t > 0$. Montrer que le temps moyen séparant le dernier autobus arrivé avant t et le premier après t est strictement plus grand que le temps moyen séparant 2 autobus ; et que si t est grand, il vaut approximativement le double.

Exercice 10

Les employés d'une entreprise arrivent au travail selon un processus de Poisson $N(t)$ d'intensité λ . On note A_k l'instant d'arrivée du $k^{ième}$ employé.

1. Exprimer le nombre d'heures totales $X(t)$ de travail effectuées dans l'entreprise jusqu'à la date t en fonction de $N(t)$, et des A_k
2. On cherche à calculer $E(X(t))$

(a) Montrer que
$$E(X(t)) = \sum_{k=0}^{+\infty} E(X(t) | N(t) = n) \mathbb{P}[N(t) = n]$$

(b) En déduire que
$$E(X(t)) = \frac{\lambda t^2}{2}$$

Chapitre 3

Chaînes de Markov à temps discret

3.1 Introduction et définition

Nous allons dans ce chapitre considérer des processus stochastiques à espace d'état discret E et à *temps discret* d'un type très particulier : nous supposons que l'état futur du système $X(t_{n+1})$ peut dépendre de l'état présent $X(t_n)$, mais pas des valeurs passées $X(t_k)$ pour $k < n$. Le processus sera dit “sans mémoire” ou encore “non héréditaire”. On parlera de manière équivalente de *chaîne de Markov à temps discret*. La loi de transition qui établit les probabilités de passage d'un état au temps t_n à un autre état au temps t_{n+1} de variera pas au cours du temps. On parlera dans ce contexte de chaînes de Markov *homogènes*¹.

L'utilisation de chaînes de Markov est très fréquente dans la modélisation des processus stochastiques. Cette popularité s'explique en dernier ressort par les raisons suivantes :

- (i) une partie des phénomènes complexes naturels ou humains sont intrinsèquement markovien (sans mémoire) et homogènes (stationnaires)
- (ii) l'hypothèse markovienne permet des calculs explicites des probabilités d'état au cours du temps ou des paramètres de performances des systèmes étudiés
- (iii) l'état présent d'une chaîne de Markov peut être augmenté (avec celui des quelques instants précédents) pour représenter une influence du passé récent dans le cas où le système considéré est héréditaire mais que seuls les états récents ont une influence importante sur le futur
- (iv) de nombreux processus varient peu au cours du temps et peuvent être modélisés comme des processus markoviens homogènes avec une

1. nous avons dans le chapitre précédent appelés *stationnaires* de tels processus stochastiques dont les lois d'évolution ne dépendent pas explicitement du temps

stratégie adaptative de modélisation, c'est-à-dire une remise à jour périodique des paramètres du modèle pour le “recaler” sur la réalité observée

Définition 7 (chaîne de Markov à temps discret) Soit $(X(t_n))_{n \in \mathbb{N}}$ une suite de variables aléatoires à valeurs dans E , un espace d'état discret. Alors, par définition, $(X(t_n))_{n \in \mathbb{N}}$ est une *chaîne de Markov* si et seulement si

$$\begin{aligned} IP[X(t_{n+1}) = x_{n+1} | X(t_n) = x_n; \dots; X(t_0) = x_0] = \\ IP[X(t_{n+1}) = x_{n+1} | X(t_n) = x_n] \\ \forall n \in \mathbb{N}, \forall x_{n+1} \in E, \forall (x_0, x_1, \dots, x_n) \in E^n \end{aligned} \quad (3.1)$$

La probabilité de transition de x_n à x_{n+1} ne dépend donc pas des états passé $(x_0, x_1, \dots, x_{n-1})$, mais seulement de l'état présent x_n , du temps t_n et de l'état futur x_{n+1} .

Définition 8 (homogénéité) Soit $(X(t_n))_{n \in \mathbb{N}}$ une chaîne de Markov à temps discret à valeurs dans E . Elle est dite *homogène* si et seulement si

$$IP[X(t_{n+1}) = x_j | X(t_n) = x_i] = p_{x_i, x_j}, \forall x_i, x_j \in E \quad (3.2)$$

La *probabilité de transition* p_{x_i, x_j} ne dépend donc pas dans ce cas du temps, mais seulement de l'état d'origine x_i et de l'état de destination x_j .

Afin de simplifier les notations, nous adopterons dans la suite de ce chapitre - et ce sans perte de généralité - les conventions suivantes :

- les instants discrets t_n correspondent aux instants où un évènement se produit qui entraîne le changement de la valeur de l'état (éventuellement, la nouvelle valeur peut être identique à la valeur précédente). Ces instants peuvent être équidistants (e.g. dans le cas d'un échantillonnage) ou non (e.g. lorsqu'il s'agit des instants d'occurrence d'un évènement aléatoire). Dans tous les cas, ils peuvent être rangés par ordre croissant et dénombrés :

$$t_0 = 0 < t_1 < \dots < t_n < t_{n+1} < \dots$$

Dans la suite, **on désignera donc par n ($n \in \mathbb{N}$) l'instant t_n , par X_n la variable aléatoire $X(t_n)$ qui désigne l'état au temps t_n et par x_n la valeur de sa réalisation $x(t_n)$.**

- l'espace d'état E est par hypothèse dénombrable, fini ou non, selon les exemples étudiés. Il est donc possible de ranger les éléments de E dans un certain ordre “naturel”, c'est-à-dire de construire une bijection $b : E \rightarrow \mathbb{N} : e \mapsto b(e)$ entre E et \mathbb{N} (ou une partie finie $\{0, 1, \dots, N\} \subset \mathbb{N}$ dans le cas d'un ensemble d'état E fini avec N éléments). **Dans la**

suite, on désignera ainsi par $b(e) \in \mathbb{N}$ l'élément $e \in E$. On écrira donc l'espace d'état E sous la forme :

$$E = \{1, 2, \dots, i-1, i, i+1, \dots\}$$

au lieu de

$$E = \{e_1, e_2, \dots, e_{i-1}, e_i, e_{i+1}, \dots\}$$

Les définitions 7 et 8 peuvent être ré-écrites avec ces conventions de notations. On obtient alors la définition qui suit.

Définition 9 (CMTD homogène) Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires à valeurs dans E , un espace d'état discret. Alors, $(X_n)_{n \in \mathbb{N}}$ est une *chaîne de Markov* si et seulement si

$$\begin{aligned} IP[X_{n+1} = j | X_n = i; X_{n-1} = i_{n-1}; \dots; X_0 = i_0] &= IP[X_{n+1} = j | X_n = i] \\ \forall n \in \mathbb{N}, \forall i, j \in E, \forall i_0, i_1, \dots, i_{n-1} \in E \end{aligned} \quad (3.3)$$

Cette chaîne de Markov est *homogène* si les probabilités $IP[X_{n+1} = j | X_n = i]$ ne dépendent pas du temps n . Dans la suite de ce cours, on abrégiera *chaîne de Markov en temps discret* par l'acronyme *CMTD*. Dans le cas d'une CMTD homogène, on notera :

$$p_{i,j} := IP[X_{n+1} = j | X_n = i] \quad \forall n \in \mathbb{N}, \forall i, j \in E \quad (3.4)$$

les probabilités de transition de l'état i vers l'état j . La *matrice*² *de transition* du processus, \mathbf{P} , sera définie par :

$$(\mathbf{P})_{i,j} := p_{i,j} \quad (3.5)$$

La *matrice de transition* \mathbf{P} possède quelques propriétés élémentaires liées au fait qu'il s'agit d'une *matrice de probabilités* :

- (i) $0 \leq p_{i,j} \leq 1$, $\forall i, j \in E$. Dans le cas d'une CMTD avec *rebouclage direct*, il est possible d'avoir $p_{i,i} > 0$. La transition peut donc se faire de l'état i à l'état i avec une probabilité non nulle.
- (ii) $\sum_{j \in E} p_{i,j} = 1$, $\forall i \in E$
- (iii) le vecteur colonne $\mathbf{v} = \mathbf{1} = (1, 1, \dots, 1)^T$ ³ est un vecteur propre à droite de la matrice \mathbf{P} , associé à la valeur propre $\lambda = 1$, c'est-à-dire :

$$\mathbf{P}\mathbf{1} = \lambda\mathbf{1}$$

En effet, $(\mathbf{P}\mathbf{1})_i = \mathbf{P}_{(i)}\mathbf{1} = \sum_{j \in E} p_{i,j} = 1 = (\lambda \cdot \mathbf{1})_i$, $\forall i \in E$, pour la valeur particulière $\lambda = 1$.

2. Dans ce chapitre, les variables notées en gras désigneront des vecteurs ou des matrices. Ceux-ci pourront être "infinis" dans le cas où l'espace d'état E n'est pas fini. La notation $\mathbf{P}^{(j)}$ désignera la j -ième colonne de la matrice \mathbf{P} considérée comme un vecteur colonne. Le vecteur ligne correspondant à la i -ième ligne de \mathbf{P} sera noté $\mathbf{P}_{(i)}$.

3. On notera \mathbf{A}^T la transposée d'une matrice \mathbf{A} , c'est-à-dire la matrice définie par $(\mathbf{A}^T)_{i,j} := \mathbf{A}_{j,i}$

Exemple 4 (labyrinthe) A titre d'illustration, considérons une expérience dans laquelle des rats doivent trouver la sortie d'un labyrinthe. Le labyrinthe est représenté schématiquement à la figure 3.1 (partie gauche). Dans ce labyrinthe, les rats doivent emprunter les corridors dans le sens indiqué par les double-flèches. Le labyrinthe comporte 5 pièces, numérotées 1 à 5, ainsi qu'une sortie, numérotée 6. Quand un rat est dans une pièce, il emprunte aléatoirement un des corridors qui s'ouvrent devant lui avec une distribution de probabilité uniforme. Par exemple, un rat dans la pièce 3 ira au hasard vers les pièces 1, 2, 5 ou 6 avec une probabilité $\frac{1}{4}$ pour chacune de ces pièces. L'hypothèse markovienne se traduit ici par le fait que le rat choisit parmi les corridors qui s'ouvrent devant lui toujours avec la même distribution de probabilités, quel que soit son parcours antérieur dans le labyrinthe : il n'a littéralement aucune mémoire ! Nous allons représenter la marche aléatoire du rat dans le labyrinthe par une chaîne de Markov. L'état au temps n sera le numéro de la pièce dans laquelle se trouve le rat au temps n . L'espace d'état est donc

$$E := \{1, \dots, 6\}$$

Il est fréquent de représenter une CMTD par un *graphe associé*. Les noeuds sont les états alors que les arrêtes représentent les transitions entre états dont la probabilité de transition est non nulle. Cette représentation, dans le cas de la marche aléatoire du rat dans le labyrinthe, est donnée à la figure 3.1 (partie droite). Chaque arrête est pondérée par la probabilité de la transition qu'elle représente. Il est courant d'omettre cette pondération lorsque la probabilité de transition correspondante vaut 1 (il n'y a alors qu'une seule manière de sortir de l'état d'origine), comme par exemple quand le rat se trouve dans les pièces 4, 5 ou 6.

Alternativement - et de manière équivalente - cette même CMTD peut être représentée par sa *matrice de transition*. On obtient ici :

$$\mathbf{P} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} & 0 & 0 \\ \frac{1}{4} & \frac{1}{4} & 0 & 0 & \frac{1}{4} & \frac{1}{4} \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \quad (3.6)$$

La CMTD qui modélise la marche aléatoire du rat dans le labyrinthe est complètement définie, soit par la matrice de transition (3.6), soit par le graphe associé décrit à la figure 3.1. Toutes les probabilités relatives à ce problème peuvent alors être calculées à l'aide de ce modèle et de la *distribution initiale de probabilité*⁴. Voici des exemples de questions (parmi

4. La distribution initiale de probabilité indique notre connaissance à priori sur le système au moment où nous commençons à regarder son évolution (au temps 0). Par exemple, si nous savons que le rat démarre dans la pièce 2, nous choisirons une distribution initiale telle que $\mathbb{P}[X_0 = 2] = 1$ (voir la section 3.3, ci-après)

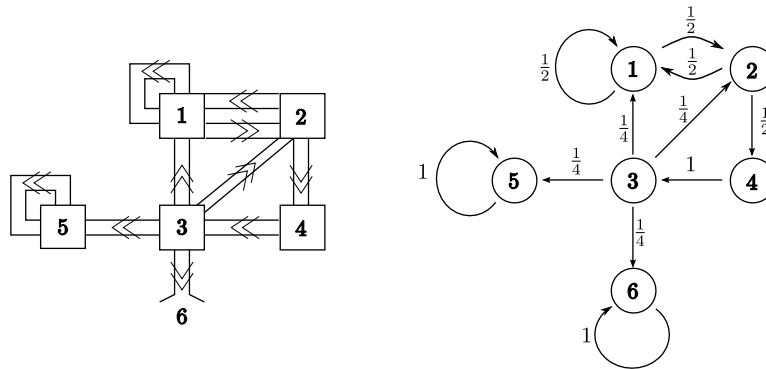


FIGURE 3.1 – Représentation schématique du labyrinthe de l'exemple 4 (à gauche) et graphe de la chaîne de Markov associée (à droite). Les noeuds représentent les états et les poids sur les arrêtes représentent les probabilités de transition correspondantes.

beaucoup d'autres) que nous allons pouvoir résoudre grâce aux méthodes développées dans ce chapitre :

- Quelle est la probabilité pour un rat dans la pièce 2 de revenir à la pièce 2 après 4 déplacements ?
- Combien de fois en moyenne un rat passera-t-il par la pièce 3 ?
- Quel est la probabilité pour un rat de trouver la sortie ? Quel est le temps de séjour moyen dans le labyrinthe des rats qui réussissent à trouver la sortie ?

3.2 Quelques exemples classiques de CMTD

Dans cette section, nous allons passer en revue d'autres exemples de CMTD finies ou non. Nous donnerons leurs représentations sous forme de graphes et leurs matrices de transition. Il s'agira pour la plupart d'exemples simples, à vocation pédagogique, même si certains sont utilisés dans des applications réelles et parfois complexes.

D'une manière générale, les systèmes markoviens sont partout ... ou presque ! Ils sont entre autres très utilisés en physique (physique statistique), en théorie de l'information et de la communication, dans les outils de compression de signaux et images, dans les outils de reconnaissance des formes (images, langue naturelle) ou en intelligence artificielle (e.g. algorithme de Viterbi), dans la modélisation et l'analyse des réseaux (e.g. PageRank, modèles de votes, réseaux sociaux, etc.), dans la modélisation des files d'attente ou encore pour analyser la sûreté de fonctionnement (fiabilité, disponibilité) des systèmes techniques ou la sécurité des systèmes informatiques et des réseaux (détection d'attaques, propagation de virus, etc.). Les chaînes de

Markov fondent les systèmes de Bonus/Malus mis au point par les actuaires des sociétés d'assurances automobiles. Elles sont également utilisées en bio-informatique (génétique, épidémiologie, écologie, etc.).

3.2.1 La chaîne à deux états

La CMTD à deux états est l'exemple le plus simple de chaîne de Markov, après le cas trivial (chaîne à un état). Son graphe associé est représenté à la figure 3.2. Les probabilités de transition satisfont :

$$0 \leq \alpha, \beta \leq 1$$

La matrice de transition de cette chaîne de Markov est donnée par :

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

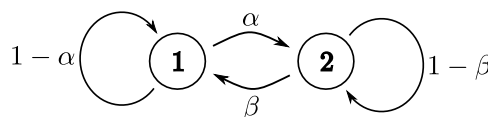


FIGURE 3.2 – Graphe associé à une chaîne de Markov à deux états

3.2.2 Modèle de diffusion

Le modèle des urnes d'Ehrenfest (1907) est un modèle élémentaire de diffusion des particules qui a connu une certaine popularité, notamment dans les controverses autour des fondements de la mécanique statistique, à l'époque de la naissance de cette discipline (la théorie cinétique des gaz est formulée par Boltzmann autour de 1872). Ce modèle des urnes décrit notamment l'évolution vers l'équilibre thermodynamique d'un gaz. On procède par l'esprit à l'expérience virtuelle décrite à la figure 3.3. Deux urnes A et B sont séparées à l'instant initial par une cloison étanche. L'urne A contient N boules, numérotées de 1 à N (représentant les particules du gaz). Ensuite la cloison est enlevée. Les particules de gaz vont donc pouvoir diffuser d'une urne dans l'autre. Le modèle stochastique associé à cette diffusion est le suivant : à chaque pas de temps, on tire au hasard un numéro j compris entre 1 et N et on transfère la boule numéro j dans l'autre urne (celle où elle n'était pas). L'état du système X_n au temps n est le nombre de boule dans l'urne A au temps n . L'espace d'état de la chaîne de Markov associée à cette expérience est donc :

$$E := \{0, 1, \dots, N\}$$

La graphe associée à cette chaîne de Markov est représenté à la figure 3.4. La matrice de transition de la CMTD est donnée par

$$\mathbf{P} = \begin{pmatrix} 0 & 1 & 0 & \cdots & \cdots & \cdots & 0 \\ \frac{1}{N} & 0 & \frac{N-1}{N} & \ddots & & & \vdots \\ 0 & \frac{2}{N} & 0 & \ddots & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 & \frac{2}{N} & 0 \\ \vdots & & & \ddots & \frac{N-1}{N} & 0 & \frac{1}{N} \\ 0 & \cdots & \cdots & \cdots & 0 & 1 & 0 \end{pmatrix}$$

La figure 3.5 représente les résultats obtenus lors d'une réalisation de l'expérience des urnes d'Ehrenfest avec $N = 1000$ boules dans les urnes pendant 10 000 pas de temps. On notera en particulier l'évolution vers l'équilibre thermodynamique où le nombre de boules dans les deux compartiments est approximativement le même et fluctue assez peu. Il s'agit de l'état d'entropie maximum qui correspond au nombre maximum de configurations possibles (au sens combinatoire).

3.2.3 Promenade aléatoire

Une *marche aléatoire* (ou promenade aléatoire, ou encore *random walk*, en anglais) est un modèle mathématique fréquemment utilisé en physique ou en économie, notamment. Elle est composée d'une succession de pas (ou choix) aléatoires, parmi un nombre fini de pas (ou choix) possibles, et décorrélés les uns des autres. C'est cette indépendance qui justifie le caractère markovien : le pas suivant ne dépend que de la position à l'instant présent et pas du chemin parcouru auparavant pour atteindre cet état présent. Une promenade aléatoire à une dimension peut se représenter

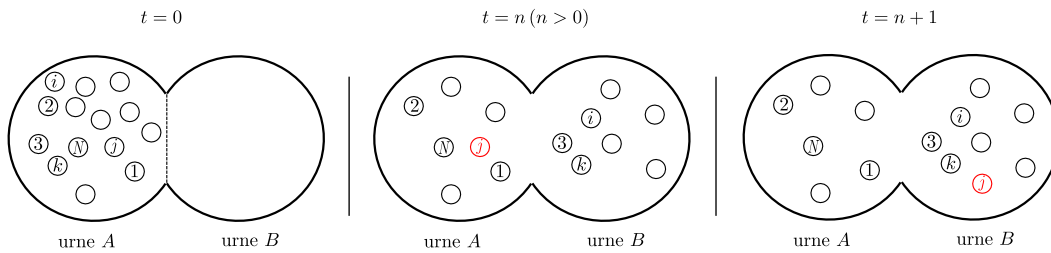


FIGURE 3.3 – L'expérience des urnes d'Ehrenfest. Au temps 0, les particules (boules) sont toutes dans le compartiment (urne) de gauche. A chaque pas de temps n , on choisit une particule au hasard (ici, la particule j) et on la transfère dans l'autre compartiment (ici, B).

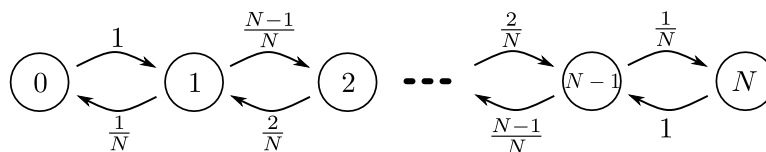


FIGURE 3.4 – Le graphe de la CMTD associée à l'expérience aléatoire virtuelle des urnes d'Ehrenfest

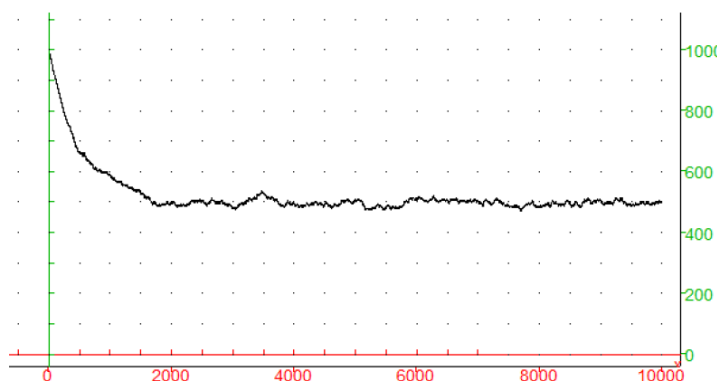


FIGURE 3.5 – Nombres de boules X_n dans l'urne A (ordonnées), en fonction du temps n (abscisses) lors de la réalisation d'une expérience aléatoire avec 1000 boules et 10 000 tirages.

comme une marche aléatoire sur \mathbb{Z} . La chaîne de Markov associée possède un espace d'état (infini mais dénombrable) :

$$E := \mathbb{Z}$$

Les éléments de la matrice de transition \mathbf{P} (elle aussi infinie) s'écrivent :

$$p_{i,j} := \begin{cases} p & \text{si } j = i + 1 \\ 1 - p & \text{si } j = i - 1 \\ 0 & \text{sinon} \end{cases}$$

avec $p \in [0, 1]$ et $q := 1 - p$. Le graphe associé à cette chaîne de Markov est représenté à la figure 3.6.

La figure 3.7 montre trois exemples de simulation (itérations) d'une marche aléatoire *isotrope* ($p = q = \frac{1}{2}$) sur \mathbb{Z} , à partir de l'état initial $X_0 = 0$.

Le même principe de marche aléatoire peut être utilisé avec un réseau (*lattice*) 2D ou 3D régulier (par exemple sur \mathbb{Z}^2 ou \mathbb{Z}^3) pour rendre compte

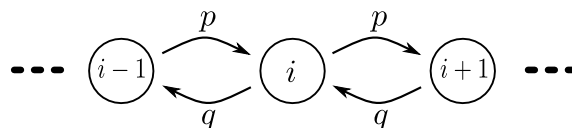


FIGURE 3.6 – Graphe associé à la chaîne de Markov modélisant un marche aléatoire sur \mathbb{Z} .

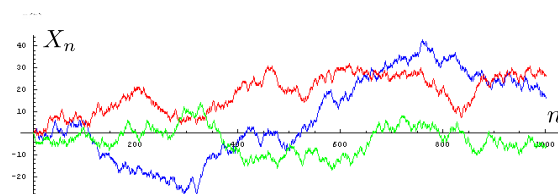


FIGURE 3.7 – Marche aléatoire isotrope sur \mathbb{Z} (1000 itérations)

de certains phénomènes naturels comme le mouvement brownien, correspondant aux mouvements en apparence aléatoires de particules libres. Les marches aléatoires sur des graphes plus complexes permettent de modéliser d'autres phénomènes comme la navigation d'un internaute sur le web. C'est la méthode utilisée par le moteur de recherche Google pour parcourir, identifier et classer les pages du réseau internet.

3.2.4 Le modèle de la ruine du joueur

Le modèle de la ruine du joueur représente la situation suivante. Un joueur A joue contre un joueur B une suite de parties indépendantes de pile ou face (ou tout autre jeu binaire) avec une pièce éventuellement pipée. Au départ, A possède a euros et B en possède b . La somme totale de leurs fortunes est donc $a + b$ euros. A chaque partie on convient que le joueur A gagne 1 euro si le résultat est pile, sinon il perd un euro. La probabilité d'obtenir pile est notée p avec $0 < p < 1$. Le jeu s'arrête dès que l'un des joueurs est ruiné. Pour chaque $n \geq 1$, on désigne par X_n la fortune du joueur A à l'issue de la n -ième partie. La suite $(X_n)_{n \geq 0}$ est une chaîne de Markov, dont l'ensemble des états est

$$E = \{0, 1, \dots, a + b\}$$

Sa matrice de transition est donnée par

$$\mathbf{P} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ q & 0 & p & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & q & 0 & p \\ 0 & \cdots & 0 & 0 & 1 \end{pmatrix}$$

Le graphe associé est illustré à la figure 3.8. Les états 0 (ruine de A) et $N = a + b$ (ruine de B) sont dits *absorbants* : une fois atteints, on y reste indéfiniment. Le graphe a 3 composantes fortement connexes qui seront appelées en classes dans la suite. Seulement deux de ces classes sont fermées, les classes $\{0\}$ et $\{N\}$. Elles seront appelées *récurrentes* (composées d'états récurrents) alors que la classe $\{1, \dots, N-1\}$ sera dite transitoire (composée d'états transitoires).

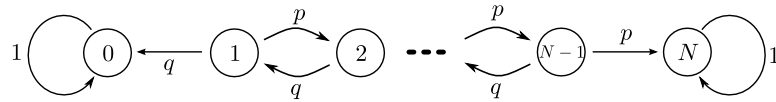


FIGURE 3.8 – Graphe de la CMTD associée au problème de la ruine du joueur, pour une dotation initiale de $N := a + b$ euros, avec $p \in [0, 1]$ et $q := 1 - p$

3.2.5 Système dynamique déterministe

On peut voir à l'examen du modèle de la ruine du joueur qu'un certain nombre de propriétés des modèles markoviens sont déterminées par leurs propriétés topologiques, c'est-à-dire par la topologie du graphe associé, de manière indépendante des valeurs numériques particulières des probabilités de transition. En poussant l'analyse, on peut aller jusqu'au cas "extrême" où toutes les probabilités de transition valent 1 (ou 0). On retombe alors sur un modèle d'évolution déterministe de la forme :

$$X_{n+1} = f(X_n) \tag{3.7}$$

où $X_n \in E$ désigne l'état au temps n , E est l'espace d'état (fini ou dénombrable) et $f : E \rightarrow E$ l'application qui permet de déterminer de manière univoque l'état suivant (image) en fonction du précédent (antécédent). Le modèle déterministe (3.7) peut être défini comme une chaîne de Markov sur E avec les probabilités de transition :

$$p_{i,j} := \begin{cases} 1 & \text{si } f(i) = j \\ 0 & \text{sinon} \end{cases}, \forall i, j \in E$$

Le graphe associé à cette chaîne de Markov est le graphe au sens usuel de l'application f . Un exemple de système dynamique déterministe est proposé à la figure 3.9. Sa matrice de transition est donnée par

$$\mathbf{P} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

Sur cet exemple, l'application $f : E \rightarrow E$ (avec $E = \{1, \dots, 7\}$) n'est ni

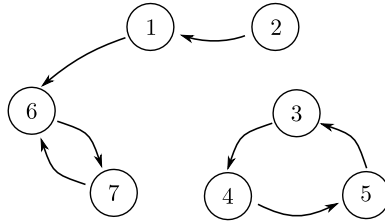


FIGURE 3.9 – Exemple de système dynamique déterministe à sept états ($E = \{1, \dots, 7\}$)

surjective, ni injective. L'image 6 a deux antécédents car $f(7) = f(1) = 6$ alors que l'état 2 n'a pas d'antécédent. Cela correspond aux faits que la sixième colonne de \mathbf{P} a deux éléments égaux à 1 (en première et septième lignes), alors que la première colonne de \mathbf{P} n'a pas d'élément non nul.

Si l'ensemble E est fini et si l'application f est une permutation⁵, alors la matrice de transition \mathbf{P} est une matrice de permutation (composée exclusivement de 0 et de 1, avec un seul élément 1 par ligne et par colonne). Dans ce cas, l'image de E satisfait $f(E) = E$ et l'ensemble des états de E sont des états récurrents. Dans le cas général (lorsque f n'est pas une permutation), on a $E \supset f(E)$. Les états successifs appartiennent donc à la suite décroissante des espaces

$$E \supset f(E) \supset f^2(E) \supset \dots$$

où $f^2(E) := (f \circ f)(E)$. Lorsque l'espace d'état E est fini, il existe nécessairement un plus petit sous-espace $E_{\min} \subset E$ tel que :

$$\exists N \geq 0 : f^n(E) = E_{\min}, \forall n \geq N$$

5. une application $f : E \rightarrow E$ avec E fini est une permutation si c'est une bijection sur E . Dans ce cas particulier, elle est bijective si et seulement si elle est surjective (ou si et seulement si elle est injective)

La restriction de f à E_{\min} est donc nécessairement une permutation. Comme toute permutation, celle-ci se décompose en un nombre fini r de cycles disjoints :

$$\begin{aligned} E_{\min} &= C_1 \oplus C_2 \oplus \dots \oplus C_r \\ C_i &\cap C_j = \emptyset, \forall i, j \end{aligned}$$

On définit alors les orbites associées à chacun de ces cycles disjoints comme l'ensemble des états de E à partir desquels on arrive, après un certain nombre d'itérations, dans le cycle correspondant :

$$O_i := \{X_0 \in E \mid \exists N \geq 0 : n \geq N \Rightarrow X_n \in C_i\}$$

où X_n est la trajectoire solution de $X_{n+1} = f(X_n)$ calculée à partir de la condition initiale X_0 . On peut donc décomposer l'espace d'état E en r orbites disjointes :

$$\begin{aligned} E &= O_1 \oplus O_2 \oplus \dots \oplus O_r \\ O_i &\cap O_j = \emptyset, \forall i, j \\ C_i &\subset O_i, \forall i \end{aligned} \tag{3.8}$$

Ainsi, d'une manière générale, toutes les trajectoires aboutiront, après un nombre fini d'étapes transitoires, dans un cycle qu'elle parcoureront ensuite indéfiniment. Dans l'exemple de la figure 3.9, les cycles sont $C_1 = \{6, 7\}$, $C_2 = \{2\}$ et $C_3 = \{3, 4, 5\}$. Les orbites correspondantes sont $O_1 = C_1 \cup \{1\} = \{6, 7, 1\}$, $O_2 = C_2 = \{2\}$ et $O_3 = C_3 \cup \{3, 4, 5\}$.

3.3 Probabilités d'états et probabilités de chemins

3.3.1 Probabilité d'état

Soit $\{X_n\}_{n \in \mathbb{N}}$ une CMTD homogène sur un espace d'état E discret. Le vecteur ligne $\pi^{(n)}$ des *probabilités d'état* au temps n ($n \geq 0$) est défini composante par composante par :

$$\pi_j^{(n)} := \mathbb{P}[X_n = j], \forall j \in E$$

On appelle *distribution initiale* de probabilités le vecteur des probabilités d'état à l'instant $n = 0$:

$$\pi_j^{(0)} := \mathbb{P}[X_0 = j], \forall j \in E$$

Celui-ci doit être connu pour pouvoir déterminer l'évolution ultérieure du vecteur des probabilités d'état $\pi^{(n)}$ pour $n > 0$. Il représente notre connaissance de la configuration initiale du système. Si nous n'avons aucune information sur l'état initial de la chaîne, on pourra choisir une distribution

uniforme de probabilités initiales, par exemple :

$$\pi_j^{(0)} = \frac{1}{N}, \forall j \in E := \{1, \dots, N\}$$

Si par contre nous savons qu'au temps $n = 0$ la chaîne est dans l'état i , on choisira plutôt la distribution initiale de probabilités :

$$\pi_j^{(0)} = \delta_{i,j} := \begin{cases} 1 & \text{si } j = i \\ 0 & \text{sinon} \end{cases}, \forall j \in E$$

Une fois connue la distribution initiale de probabilités $\pi^{(0)}$ et les probabilités de transition $p_{i,j}$, toutes les valeurs ultérieures du vecteur des probabilités d'état $\pi^{(n)}$ peuvent être calculées grâce à l'hypothèse markovienne et à la formule des probabilités totales. On a en effet :

$$\begin{aligned} \pi_j^{(n)} &= IP[X_n = j] \\ &= \sum_{i \in E} IP[X_n = j | X_{n-1} = i] IP[X_{n-1} = i] \\ &= \sum_{i \in E} p_{i,j} \pi_i^{(n-1)} \end{aligned}$$

Sous forme matricielle, on obtient pour le vecteur des probabilités d'état :

$$\pi^{(n)} = \pi^{(n-1)} \mathbf{P} \quad (3.9)$$

ou encore, par récurrence sur n :

$$\pi^{(n)} = \pi^{(0)} \mathbf{P}^n \quad (3.10)$$

L'équation (3.10) est parfois appelée relation de Chapman-Kolmogorov.

3.3.2 Probabilité d'un chemin

Soit $\{X_n\}_{n \in \mathbb{N}}$ une CMTD homogène sur un espace d'état E discret. Soit $(x_0, x_1, \dots, x_n) \in E^n$ un chemin quelconque de longueur n dans l'espace des états. La probabilité que l'état de la chaîne parcoure dans le même ordre les valeurs de ce chemin lors d'une suite de n transitions aléatoires successives est appelée *probabilité du chemin* (x_0, x_1, \dots, x_n) et s'écrit :

$$\begin{aligned} p_{x_0, x_1, \dots, x_n} &:= IP[X_0 = x_0, X_1 = x_1, \dots, X_n = x_n] \\ &= IP[X_n = x_n, \dots, X_1 = x_1, X_0 = x_0] \\ &= IP[X_n = x_n, \dots, X_1 = x_1 | X_0 = x_0] IP[X_0 = x_0] \\ &= IP[X_n = x_n, \dots, X_1 = x_1 | X_0 = x_0] \pi_{x_0}^{(0)} \end{aligned}$$

en utilisant la définition de la probabilité conditionnelle. La probabilité conditionnelle $P^{\{X_0=x_0\}}$ définie par

$$\mathbb{P}^{\{X_0=x_0\}}[\cdot] := \mathbb{P}[\cdot | X_0 = x_0]$$

étant elle même une mesure de probabilité, on peut répéter le raisonnement ci-dessus et obtenir en particulier :

$$\begin{aligned} p_{x_0, x_1, \dots, x_n} &:= \mathbb{P}[X_n = x_n, \dots, X_1 = x_1 | X_0 = x_0] \pi_{x_0}^{(0)} \\ &= \mathbb{P}^{\{X_0=x_0\}}[X_n = x_n, \dots, X_1 = x_1] \pi_{x_0}^{(0)} \\ &= \mathbb{P}^{\{X_0=x_0\}}[X_n = x_n, \dots, X_2 = x_2 | X_1 = x_1] \mathbb{P}^{\{X_0=x_0\}}[X_1 = x_1] \pi_{x_0}^{(0)} \\ &= \mathbb{P}^{\{X_0=x_0\}}[X_n = x_n, \dots, X_2 = x_2 | X_1 = x_1] p_{x_0, x_1} \pi_{x_0}^{(0)} \\ &= \mathbb{P}[X_n = x_n, \dots, X_2 = x_2 | X_1 = x_1] p_{x_0, x_1} \pi_{x_0}^{(0)} \end{aligned}$$

grâce à l'hypothèse markovienne⁶. En appliquant le raisonnement ci-dessus de manière itérative, on obtient pour la probabilité du chemin (x_0, x_1, \dots, x_n) la valeur :

$$p_{x_0, x_1, \dots, x_n} = p_{x_{n-1}, x_n} p_{x_{n-2}, x_{n-1}} \cdots p_{x_1, x_2} p_{x_0, x_1} \pi_{x_0}^{(0)} \quad (3.11)$$

La probabilité de suivre un chemin quelconque du graphe associé à une CMTD homogène est donc le produit des probabilités de transition entre les noeuds du graphe (pris deux à deux) qui composent ce chemin. Nous allons illustrer cette méthode sur l'exemple du labyrinthe

Exemple 5 (labyrinthe, suite) Reprenons l'exemple 4 du labyrinthe (voir section 3.1). Nous allons en particulier calculer la probabilité qu'un rat parte de la pièce numéro 2 et y revienne après quatre déplacements, c'est à dire $P(X_4 = 2 | X_0 = 2)$. Nous allons calculer cette probabilité de deux manières différentes : avec la relation de Chapman-Kolmogorov (3.10), d'une part, et avec la formule de la probabilité d'un chemin (3.11), d'autre part. Commençons avec la méthode des probabilités d'état. On a par hypothèse $\pi_i^{(0)} = \delta_{i,2}$ et d'après la formule (3.10) :

$$\mathbb{P}[X_4 = 2 | X_0 = 2] = \pi_2^{(4)} = \sum_{i=1}^6 \pi_i^{(0)} (\mathbf{P}^4)_{i,2} = (\mathbf{P}^4)_{2,2}$$

En utilisant la définition (3.6) pour la matrice de transition associée à

6. D'après cette hypothèse, aussi appelée propriété d'oubli, la probabilité de transition vers l'état x_2 et les états suivant ne peut dépendre que de l'état x_1 , mais pas de l'état x_0 au temps $n = 0$

l'exemple du labyrinthe, on obtient :

$$\mathbf{P}^4 = \frac{1}{16} \begin{pmatrix} 6 & 4 & 2 & 2 & 1 & 1 \\ 5 & 3 & 2 & 2 & 2 & 2 \\ 3 & 2 & 1 & 1 & \frac{9}{2} & \frac{9}{2} \\ 3 & 2 & 2 & 1 & 4 & 4 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \quad (3.12)$$

On a donc :

$$\mathbb{P}[X_4 = 2 | X_0 = 2] = (\mathbf{P}^4)_{2,2} = \frac{3}{16}$$

Cette première approche algébrique est utile pour des problèmes de faible complexité (lorsque l'espace d'état E est de faible dimension et/ou qu'il y a très peu de transition possibles) et pour calculer les probabilités d'état sur des horizons de temps assez courts. Elle nécessite par ailleurs l'usage d'un ordinateur ou d'une calculatrice (ou à défaut des calculs assez laborieux!). Pour des chaînes de Markov de grande taille et pour des simulations sur des horizons de temps importants, cette méthode conduit cependant à des problèmes de complexité (grande taille des matrices manipulées) et de stabilité numérique (pour les valeurs importantes de n , les modes dominants ont tendance à écraser tous les autres dans le calcul de \mathbf{P}^n). Une alternative consiste à utiliser le graphe associé à la CMTD et la formule de probabilité d'un chemin. L'examen du graphe associé à la CMTD, dans l'exemple du labyrinthe, montre qu'il n'y a que trois chemins possibles pour partir de la pièce 2 et y revenir en quatre étapes (voir figure 3.1, partie droite). Il s'agit des chemins :

$$\begin{aligned} (x_0, x_1, x_2, x_3, x_4) &= (2, 1, 2, 1, 2) \\ (x_0, x_1, x_2, x_3, x_4) &= (2, 1, 1, 1, 2) \\ (x_0, x_1, x_2, x_3, x_4) &= (2, 4, 3, 1, 2) \end{aligned}$$

Les probabilités associées à ces chemins sont :

$$\begin{aligned} \mathbb{P}[(x_0, x_1, x_2, x_3, x_4) = (2, 1, 2, 1, 2)] &= \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{16} \\ \mathbb{P}[(x_0, x_1, x_2, x_3, x_4) = (2, 1, 1, 1, 2)] &= \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{16} \\ \mathbb{P}[(x_0, x_1, x_2, x_3, x_4) = (2, 4, 3, 1, 2)] &= \frac{1}{2} \cdot 1 \cdot \frac{1}{4} \cdot \frac{1}{2} = \frac{1}{16} \end{aligned}$$

La probabilité de partir de la pièce 2 et de revenir à la pièce 2 après 4 déplacements est donc la probabilité de parcourir l'un ou l'autre de ces chemins (il s'agit d'événements disjoints) :

$$\mathbb{P}[X_4 = 2 | X_0 = 2] = \frac{1}{16} + \frac{1}{16} + \frac{1}{16} = \frac{3}{16}$$

Comme on le voit dans l'exemple ci-dessus, la méthode des chemins permet de tenir compte de la topologie particulière du graphe associé à la CMTD pour simplifier les calculs (en l'occurrence ne calculer que les probabilités associées aux chemins qui répondent au cahier des charges du problème). Par ailleurs, cette approche permet également de traiter des problèmes de très grandes tailles en utilisant des méthodes de représentation et des algorithmes de parcours spécifiques aux graphes.

3.4 Classes d'états dans une CMTD

Nous allons dans cette section examiner certaines propriétés structurelles des états et des sous-chaînes dans une CMTD. Ces propriétés structurelles sont celles qui ne dépendent pas des valeurs numériques spécifiques des probabilités de transition (hormis du fait qu'elles soient nulles ou non), mais plutôt de la topologie du graphe associé à la CMTD. Afin de caractériser ces propriétés, nous serons amenés à calculer les probabilités de passages d'un état à un autre, en un nombre donné d'étapes ou en un nombre quelconque d'étapes. Ces probabilités permettront à leur tour de résoudre de nouveaux problèmes associés aux CMTD comme nous l'illustreront à nouveau avec l'exemple du labyrinthe.

3.4.1 Classes d'équivalence dans une CMTD

Nous allons décomposer les états en classes d'équivalence qui seront les composantes fortement connexes⁷ du graphe orienté associé à la CMTD. Ensuite, nous distinguerons parmi ces classes d'équivalence celles qui sont récurrentes et celles qui sont transitoires.

Définition 10 (classes d'équivalence) Soit $\{X_n\}_{n \in \mathbb{N}}$ une CMTD homogène sur un espace d'état E discret. L'état $j \in E$ est dit *atteignable* depuis l'état $i \in E$ (on notera $i \rightsquigarrow j$) si et seulement si

$$\exists n \in \mathbb{N} : p_{i,j}^{(n)} > 0$$

où

$$p_{i,j}^{(n)} := (\mathbf{P}^n)_{i,j} = P(X_n = j | X_0 = i)$$

désigne la probabilité d'aller de i à j en n étapes. Les états $i, j \in E$ *communiquent* (on notera $i \longleftrightarrow j$) si et seulement si

$$i \rightsquigarrow j \text{ et } j \rightsquigarrow i$$

7. une composante fortement connexe d'un graphe orienté est un sous-graphe maximal pour la propriété suivante : pour tout couple de noeuds (i, j) du sous-graphe, il existe un chemin de i à j et de j à i

La relation de communication est une relation d'équivalence entre les états de E , c'est-à-dire une relation binaire sur E réflexive, symétrique et transitive. On définit la *classe d'équivalence* de $i \in E$ par

$$[i] := \{j \in E \mid i \rightsquigarrow j\}$$

On appelle alors i un *représentant* de la classe $[i]$. L'espace d'état E se *partitionne* en l'ensemble des classes d'équivalence de E qui sont toutes non vides et disjointes deux à deux. La chaîne de Markov $\{X_n\}_{n \in \mathbb{N}}$ est dite *irréductible* si et seulement si tous les états de E communiquent entre eux, c'est-à-dire si il n'y a qu'une seule classe d'équivalence dans E .

Exemple 6 Appliquons ces définitions à la chaîne de Markov sur $E := \{1, \dots, 7\}$ dont le graphe est donné à la figure 3.10. Les états 6 et 7, par exemple, communiquent. Ils sont donc dans la même classe d'équivalence. Par contre, 6 et 1 ne communiquent pas car 1 n'est pas atteignable depuis 6. Il en va d'ailleurs de même pour les états $\{1, 2, 3, 4, 5\}$. Nous obtenons ainsi une première classe d'équivalence

$$[6] = [7] = \{6, 7\}$$

L'espace d'état de la chaîne de Markov correspondant à la figure 3.10 se partitionne donc en trois classes d'équivalence disjointes

$$E = \{1, 2\} \cup \{3, 4, 5\} \cup \{6, 7\}$$

Parmi ces trois classes, $[6]$ et $[3]$ sont deux sous-chaînes absorbantes : dès qu'on y est entré, on n'en sortira plus. La classe $[1]$ par contre est une sous-chaîne transitoire. On est sûr de la quitter à un moment ou un autre (pour peu qu'on démarre dans cette classe) et on n'y renviendra ensuite plus. Les états des classes $[6]$ et $[3]$ sont tous des états récurrents : lorsqu'on les quitte, on est certain d'y revenir, à un moment ou un autre. Les états de la classe $[1]$ sont tous transitoires.

On montre facilement, par l'absurde, la propriété suivante.

Proposition 11 Toute chaîne de Markov *non irréductible* contient au moins une sous-chaîne *absorbante*

3.4.2 Périodicité dans une CMTD

La notion d'état périodique dans une CMTD diffère quelque peu de la notion de périodicité d'un état dans un système dynamique discret déterministe. En effet, le caractère aléatoire des transitions empêche d'observer une réelle périodicité dans la suite des états successifs x_n d'une réalisation particulière de $\{X_n\}_{n \in \mathbb{N}}$.

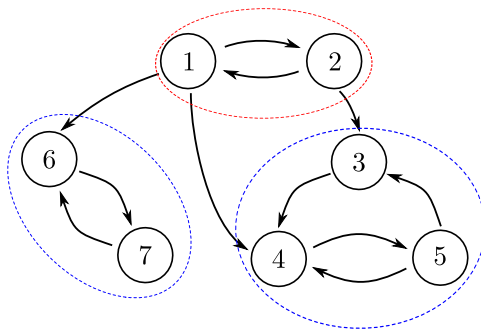


FIGURE 3.10 – Graphe de la CMTD non irréductible étudiée dans l'exemple 6

Définition 11 (Périodicité) Soit $\{X_n\}_{n \in \mathbb{N}}$ une CMTD homogène sur un espace d'état E discret. L'état $j \in E$ est dit *périodique* si et seulement si, partant de cet état j , il n'est possible d'y revenir qu'après un nombre d'étapes multiple de $k > 1$, c'est-à-dire si et seulement si

$$\exists k > 1 : p_{j,j}^{(m)} = 0, \forall m \notin \{kp \mid p \in \mathbb{N}\}$$

Le nombre k est appelé la *période* de l'état j . La période d'une CMTD est le plus grand commun diviseur (PGCD) des périodes de chacun de ses états. Une CMTD est dite périodique si sa période est strictement supérieure à 1.

Une définition alternative de la période d'un état j est le PGCD de l'ensemble

$$M_{j,j} := \left\{ m \mid p_{j,j}^{(m)} > 0 \right\}$$

c'est-à-dire le PGCD de l'ensemble des longueurs de chacun des circuits du graphe associé à la CMTD qui partent de j et reviennent à j . La période de la CMTD peut ainsi être vue comme le PGCD de tous les circuits du graphe associé. Si deux états i et j communiquent, ils ont nécessairement la même période puisqu'il est possible de construire un cycle qui part de i , pour aller à j , puis revenir à i , la longueur de ce cycle étant donc à la fois dans $M_{i,i}$ et $M_{j,j}$. Ainsi, tous les états d'une même classe d'équivalence ont nécessairement la même période.

Reprenons la CMTD de l'exemple précédent, représentée à la figure 3.10. Les états de la classe $[6] = \{6, 7\}$ sont périodiques de période 2. Il en va de même pour les états de la classe $[1] = \{1, 2\}$. Par contre, les états de la classe $[3] = \{3, 4, 5\}$ ne sont pas périodiques. En effet, il est possible de partir de l'état 3 (par exemple) et d'y revenir en 3 ou 5 étapes. Comme 3 et 5 n'ont que 1 comme diviseur commun, on en déduit que l'état 3 n'est pas périodique. Les états 4 et 5 ne peuvent donc pas l'être non plus. La chaîne dans son ensemble ne peut donc pas non plus être périodique.

Examinons maintenant les exemples de chaînes de Markov irréductibles illustrés à la figure 3.11. Dans la chaîne de gauche, l'état 3 ne peut être périodique puisqu'il y a un terme de rebouclage direct (cycle de longueur 1). Donc aucun des états de cette chaîne de gauche ne peut être périodique puisqu'il n'y a qu'une classe d'équivalence. Les états de la chaîne de droite ne peuvent pas non plus être périodiques. En effet, il existe des cycles de longueurs 2, 3, 5, 7, 8, etc. permettant de revenir à l'état 3. Aucun des états de la chaîne de droite ne peut donc être périodique puisqu'il n'y a qu'une classe d'équivalence. Dans la chaîne du milieu, l'état 1 (par exemple) est périodique de période 3 (on peut y revenir en 3, 6, 9, etc. étapes). Donc tous les états de cette chaîne sont périodiques de période 3 (puisque'il n'y a qu'une classe d'équivalence). La chaîne elle-même est donc irréductible et périodique de période 3. Lorsqu'on observe une trajectoire (réalisation) particulière de la chaîne 3, par exemple

$$1, 3, 4, 1, 3, 4, 2, 3, 4, 1, 3, 4, 2, 3, 4, 2, 3, 4, \dots$$

on ne reviendra à un état donné qu'après un nombre de pas de temps multiple de 3. Par contre, la suite des réalisations n'est pas une suite périodique.

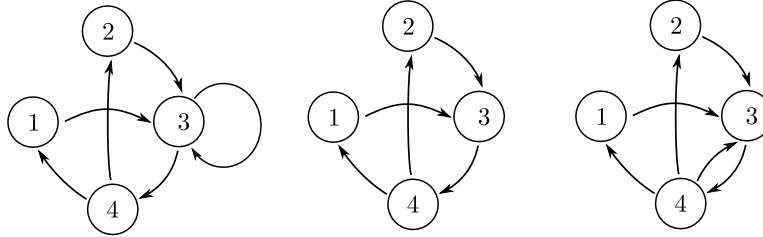


FIGURE 3.11 – Trois exemples de CMTD pour illustrer la notion de périodicité

3.4.3 Etats récurrents et transitoires

Nous allons distinguer les états *récurrents* des états *transitoires*. Les états récurrents sont ceux où on est certain (au sens probabiliste) de revenir. La probabilité de retour à un état $i \in E$ récurrent est donc égale à 1. Toutefois, dans le cas des chaînes de Markov où l'espace d'état est dénombrable mais infini, le temps moyen de retour peut être lui-même infini. Les états transitoires sont ceux où on ne revient en moyenne qu'un nombre fini de fois. La probabilité de retour pour un tel état transitoire est donc strictement inférieure à 1. Ces notions d'états transitoires ou récurrents nécessitent des définitions - et le calcul - des probabilités de retour, des probabilités de premier retour et de temps moyen de retour.

Définition 12 (Probabilités de transition en n étapes) Soit $\{X_n\}_{n \in \mathbb{N}}$, une CMTD homogène sur un espace d'état E discret. On appelle probabilité de transition de l'état $i \in E$ à l'état $j \in E$ en n ($n \in \mathbb{N}$) étapes *exactement* :

$$f_{i,j}^{(n)} := P(X_n = j; X_{n-1} \neq j; \dots; X_1 \neq j | X_0 = i) \quad (3.13)$$

La probabilité que le premier retour en $i \in E$ ait lieu en n étapes exactement est donc $f_{i,i}^{(n)}$. Le temps moyen de premier retour en $i \in E$ s'écrit :

$$M_i := \sum_{n \in \mathbb{N}} n f_{i,i}^{(n)} L a \quad (3.14)$$

La probabilité de transition de l'état $i \in E$ à l'état $j \in E$ en un nombre *quelconque* d'étapes vaut :

$$f_{i,j} := \sum_{n \in \mathbb{N}} f_{i,j}^{(n)} \quad (3.15)$$

La probabilité revenir en $i \in E$ s'écrit alors $f_{i,i}$.

Attention : la probabilité de transition en n étapes exactement, notée $f_{i,j}^{(n)}$ est différente de la probabilité d'aller de i à j en n étapes, en y étant éventuellement déjà passé (une ou plusieurs fois) entre les temps 0 et n . Cette dernière probabilité vaut $p_{i,j}^{(n)} := (\mathbf{P}^n)_{i,j} = P(X_n = j | X_0 = i)$.

Le calcul direct des probabilités $f_{i,j}^{(n)}$ et $f_{i,i}$ peut s'avérer compliqué (voir les exemples dans la section 3.4.4 ci-après) et dans certains cas inextricable. Heureusement, une approche récursive de ce problème permet d'établir une récurrence pour les $f_{i,j}^{(n)}$ et un système d'équations linéaires pour obtenir les $f_{i,j}$. L'idée générale est de décomposer le trajet de i à j en n étapes comme un trajet de i à k en une étape, puis un trajet de k à j en $(n-1)$ étapes. On obtient alors, grâce au caractère markovien :

$$\begin{aligned} f_{i,j}^{(1)} &= p_{i,j} \\ f_{i,j}^{(n)} &:= \sum_{k \neq j} P(X_1 = k | X_0 = i) f_{k,j}^{(n-1)}, \text{ pour } n \geq 1 \\ &= \sum_{k \neq j} p_{i,k} f_{k,j}^{(n-1)}, \text{ pour } n \geq 1 \end{aligned} \quad (3.16)$$

pour tout $i, j \in E$. L'utilisation de la formule de récurrence (3.16) est illustrée à la section 3.4.4, sur l'exemple du labyrinthe. En appliquant la même idée

aux probabilités de transition en un nombre quelconque d'étapes, on obtient :

$$\begin{aligned}
f_{i,j} &:= \sum_{n \in \mathbb{N}} f_{i,j}^{(n)} \\
&= f_{i,j}^{(1)} + \sum_{n \geq 2} f_{i,j}^{(n)} \\
&= f_{i,j}^{(1)} + \sum_{n \geq 2} \sum_{k \neq j} p_{i,k} f_{k,j}^{(n-1)} \\
&= p_{i,j} + \sum_{k \neq j} p_{i,k} \sum_{n \geq 2} f_{k,j}^{(n-1)} \\
&= p_{i,j} + \sum_{k \neq j} p_{i,k} \sum_{m \geq 1} f_{k,j}^{(m)} \\
&= p_{i,j} + \sum_{k \neq j} p_{i,k} f_{k,j} \tag{3.17}
\end{aligned}$$

pour tout $i, j \in E$. Les équations (3.17) constituent un système linéaire dont les inconnues sont les probabilités $f_{i,j}$, au nombre de N^2 si E est de dimension finie N . Nous pouvons maintenant définir les notions d'états récurrents et transitoires, pour des chaînes finies ou non.

Définition 13 (états récurrents et transitoires) Soit $\{X_n\}_{n \in \mathbb{N}}$, une CMTD homogène sur un espace d'état E discret. Un état $j \in E$ est *transitoire* si et seulement si

$$f_{j,j} < 1$$

Il est dit *récurrent* dans le cas contraire, c'est-à-dire lorsque

$$f_{j,j} = 1$$

Un état récurrent j est dit *récurrent non nul* si

$$M_j < \infty$$

Dans le cas contraire, il est dit *récurrent nul*

Un état récurrent est donc tel que la chaîne y revient presque sûrement (i.e. avec une probabilité 1). Dans le cas d'une chaîne finie, le temps moyen de retour est alors nécessairement fini (états récurrents non nuls) et le nombre moyen de retours est infini. Par contre, dans le cas d'une chaîne infinie, on peut imaginer des situations où la probabilité de retour est égale à 1, mais où l'espérance du temps de retour est infinie. C'est le cas, par exemple, de la marche aléatoire symétrique (isotrope, $p = \frac{1}{2}$) sur \mathbb{Z} pour laquelle tous les états sont récurrents nuls. Dans le cas d'une marche aléatoire anisotrope ($p \neq \frac{1}{2}$), les états sont par contre tous transitoires. En effet, dans ce dernier cas, à partir d'un certain moment, la chaîne ne reviendra plus sur sa position initiale (lente dérive moyenne, vers la gauche ou vers la droite).

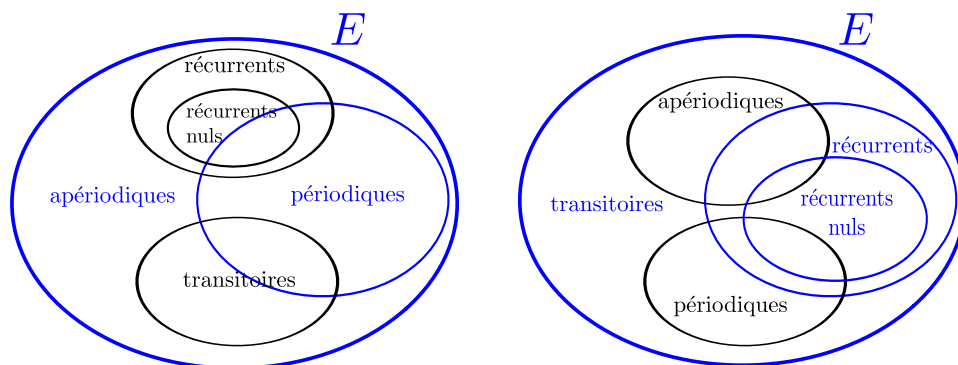


FIGURE 3.12 – Deux classements possibles (en bleu) des états d’une CMTD : états périodiques ou non (à gauche), états transitoires et récurrents (à droite). Les états récurrents peuvent à leur tour être subdivisés en états récurrents nuls ou non nuls. Les états transitoires peuvent être également subdivisés en état périodiques et apériodiques

En résumé, les états d’une CMTD peuvent être classés selon deux aspects (voir figure 3.12) : selon leur périodicité ou non, d’une part, et selon leur caractère récurrent ou transitoire, d’autre part. Les deux classements ne se recouvrent pas. Par exemple, un état périodique peut être transitoire, récurrent non nul ou même récurrent nul, comme dans le cas de la promenade aléatoire symétrique sur \mathbb{Z} évoqué précédemment.

Examinons à nouveau l’exemple 6 de la figure 3.10. Il est repris à la figure 3.13 avec un choix particulier de valeurs pour les probabilités de transition $p_{i,j}$. Les états 1 et 2 de la classe [1] sont transitoires. Après un certain nombre d’étapes, la chaîne va quitter la classe [1] et n’y reviendra jamais. On a par ailleurs :

$$f_{1,1} = f_{1,1}^{(2)} = \frac{1}{4} < 1$$

Les états 6 et 7 de la classe [6] sont récurrents. Une fois la chaîne dans un de ces deux états, on est sûr d’y revenir. On a par exemple :

$$f_{6,6} = f_{6,6}^{(2)} = 1$$

Les états 6 et 7 sont en fait récurrents non nuls car la classe [6] est finie. On a par ailleurs :

$$M_6 := \sum_{n \geq 1} n f_{6,6}^{(n)} = 2 f_{6,6}^{(2)} = 2 < \infty$$

Les états 3, 4 et 5 de la classe [3] sont également récurrents. Une fois la chaîne dans un de ces trois états, on est sûr d’y revenir. On a par exemple :

$$\begin{aligned} f_{3,3} &= f_{4,3} = f_{5,3} \\ f_{5,3} &= \frac{1}{2} + \frac{1}{2} f_{5,3} \end{aligned}$$

d'où l'on conclut $f_{3,3} = 1$. Comme on vient de le voir à travers ces exemples, la connaissance de la topologie du graphe associé à la CMTD permet parfois de simplifier considérablement le calcul des $f_{i,j}^{(n)}$, $f_{i,j}$ et M_i par rapport à la résolution explicite des équations (3.16) et (3.17).

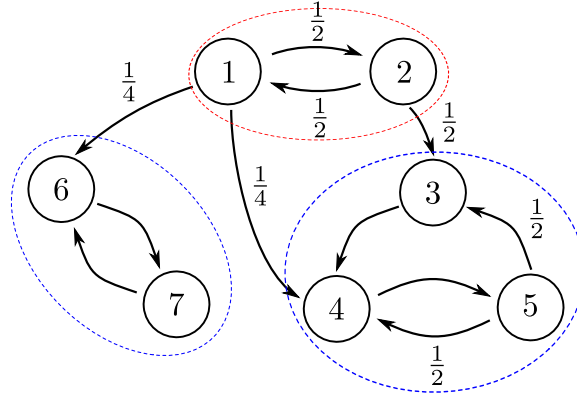


FIGURE 3.13 – Graphe de la CMTD de l'exemple 6 avec un choix particulier des probabilités de transition $p_{i,j}$

On montre facilement les propriétés suivantes :

- (i) Dans une CMTD irréductible (ou dans une classe d'équivalence), tous les états sont de même nature : ils sont tous, simultanément, transitoires, récurrents nuls ou récurrents non nuls
- (ii) Dans une CMTD finie, si j est un état récurrent, alors (partant de j) la chaîne repassera nécessairement une infinité de fois en j . A contrario, si j est transitoire, elle n'y repassera qu'un nombre fini de fois.
- (iii) Tous les états d'une CMTD irréductible finie sont récurrents non nuls
- (iv) Dans une CMTD finie, il y a au moins un état récurrent (donc au moins une classe dont tous les états sont récurrents)
- (v) Dans une CMTD finie, la chaîne se retrouve nécessairement, après un nombre fini d'étapes, dans l'ensemble des états récurrents

Remarque 8 Soit $\{X_n\}_{n \in \mathbb{N}}$ une CMTD finie. Le *nombre de passage* de la chaîne en i est la variable aléatoire définie par :

$$N_i := \sum_{n \in \mathbb{N}} 1_{\{X_n = i\}} \quad (3.18)$$

où $1_{\{X_n = i\}}$ est la variable aléatoire qui vaut 1 si $X_n = i$ et 0 sinon. L'*espérance*

du nombre de retours en i vaut :

$$\begin{aligned}
 E(N_i | X_0 = i) &:= \sum_{n \in \mathbb{N}} E(1_{\{X_n=i\}} | X_0 = i) \\
 &= \sum_{n \in \mathbb{N}} 1 \cdot P^{\{X_0=i\}}(X_n = i) \\
 &= \sum_{n \in \mathbb{N}} P(X_n = i | X_0 = i) \\
 &= \sum_{n \in \mathbb{N}} p_{i,i}^{(n)} \tag{3.19}
 \end{aligned}$$

On a donc la caractérisation

$$i \text{ récurrent} \Leftrightarrow f_{i,i} := \sum_{n \in \mathbb{N}} f_{i,i}^{(n)} = 1 \Leftrightarrow \sum_{n \in \mathbb{N}} p_{i,i}^{(n)} \text{ diverge} \tag{3.20}$$

ou encore

$$i \text{ transitoire} \Leftrightarrow f_{i,i} < 1 \Leftrightarrow \sum_{n \in \mathbb{N}} p_{i,i}^{(n)} < \infty \tag{3.21}$$

3.4.4 Exemple

Considérons à nouveau l'exemple 4 du labyrinthe présenté dans l'introduction de ce chapitre et illustré à la figure 3.1. Nous sommes maintenant en mesure de répondre à de nouvelles questions relatives à cet exemple.

Q1 *Supposons que la souris est dans la pièce 2. Combien de déplacement fera-t-elle en moyenne pour revenir à la pièce 2 ?*

Il s'agit du temps moyen de retour en 2, noté précédemment (voir (3.14)) :

$$M_2 = \sum_{n \geq 1} n f_{2,2}^{(n)} = \sum_{n \geq 2} n f_{2,2}^{(n)}$$

Le calcul des $f_{2,2}^{(n)}$ s'obtient par exemple par la formule de récurrence (3.16). On obtient :

$$\begin{aligned}
 f_{2,2}^{(n)} &= \sum_{k \neq 2} p_{2,k} f_{k,2}^{(n-1)} = p_{2,1} f_{1,2}^{(n-1)} + p_{2,4} f_{4,2}^{(n-1)} \\
 &= \frac{1}{2} f_{1,2}^{(n-1)} + \frac{1}{2} f_{4,2}^{(n-1)}, \text{ pour } n \geq 2
 \end{aligned}$$

avec $f_{2,2}^{(1)} = p_{2,2} = 0$. En appliquant la même formule (3.16), on obtient d'autre part :

$$f_{1,2}^{(n-1)} = p_{1,1} f_{1,2}^{(n-2)} = \frac{1}{2} f_{1,2}^{(n-2)}, \text{ pour } n \geq 3$$

avec $f_{1,2}^{(1)} = p_{1,2} = \frac{1}{2}$, ainsi que :

$$f_{4,2}^{(n-1)} = p_{4,3} f_{3,2}^{(n-2)} = f_{3,2}^{(n-2)}, \text{ pour } n \geq 3$$

avec $f_{4,2}^{(1)} = 0$. Enfin, toujours en appliquant (3.16), on obtient encore :

$$f_{3,2}^{(n-2)} = p_{3,1} f_{1,2}^{(n-3)} = \frac{1}{4} f_{1,2}^{(n-3)}, \text{ pour } n \geq 4$$

avec $f_{3,2}^{(1)} = \frac{1}{4}$. En combinant ces quatre derniers résultats, on obtient :

$$\begin{aligned} f_{1,2}^{(n)} &= \frac{1}{2^n}, \text{ pour } n \geq 1 \\ f_{3,2}^{(n)} &= \frac{1}{2^{n+1}}, \text{ pour } n \geq 1 \\ f_{4,2}^{(n)} &= \frac{1}{2^n}, \text{ pour } n \geq 2 \\ f_{2,2}^{(n)} &= \frac{1}{2^{n-1}}, \text{ pour } n \geq 3 \end{aligned} \quad (3.22)$$

avec $f_{2,2}^{(2)} = \frac{1}{4}$. On obtient finalement pour le nombre moyen de déplacements pour revenir en 2 :

$$\begin{aligned} M_2 &= 2 \cdot \frac{1}{4} + \sum_{n \geq 3} n \cdot \frac{1}{2^{n-1}} \\ &= \frac{1}{2} + \sum_{n \geq 3} \left[\frac{d}{dx} x^n \right] \Big|_{x=\frac{1}{2}} \\ &= \frac{1}{2} + \left[\frac{d}{dx} \sum_{n \geq 3} x^n \right] \Big|_{x=\frac{1}{2}} \\ &= \frac{1}{2} + \left[\frac{d}{dx} \left(\sum_{n \geq 0} x^n - 1 - x - x^2 \right) \right] \Big|_{x=\frac{1}{2}} \\ &= \frac{1}{2} + \left[\frac{d}{dx} \left(\frac{1}{1-x} - 1 - x - x^2 \right) \right] \Big|_{x=\frac{1}{2}} \\ &= \frac{5}{2} \end{aligned}$$

Il est intéressant de noter que, dans cet exemple, certaines probabilités de transition en n étapes exactement peuvent être trouvées de manière plus directe (et plus intuitive) en raisonnant directement à partir de la topologie du le graphe associé à la CMTD du labyrinthe. Ainsi, en raisonnant directement sur le graphe de la figure 3.1, il est facile de se convaincre que :

— le seul chemin pour aller de l'état 2 à l'état 2 en 2 étapes est le chemin $2 \rightarrow 1 \rightarrow 2$ et que, en conséquence :

$$f_{2,2}^{(2)} = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$$

- il y a deux chemins pour aller de l'état 2 à l'état 2 en 3 étapes : le chemin $2 \rightarrow 4 \rightarrow 3 \rightarrow 2$ et le chemin $2 \rightarrow 1 \rightarrow 1 \rightarrow 2$. En conséquence :

$$f_{2,2}^{(3)} = \frac{1}{2} \cdot 1 \cdot \frac{1}{4} + \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$$

- il y a deux chemins pour aller de l'état 2 à l'état 2 en n étapes lorsque $n \geq 4$: le chemin $2 \rightarrow 4 \rightarrow 3 \rightarrow 1 \rightarrow \dots \rightarrow 1 \rightarrow 2$ et le chemin $2 \rightarrow 1 \rightarrow \dots \rightarrow 1 \rightarrow 2$. En conséquence :

$$f_{2,2}^{(n)} = \frac{1}{2} \cdot 1 \cdot \frac{1}{4} \cdot \frac{1}{2^{n-4}} \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{2^{n-2}} \cdot \frac{1}{2} = \frac{1}{2^{n-1}}$$

Nous retrouvons bien sûr le résultat (3.22), mais par le biais d'un raisonnement basé sur la topologie du graphe et non plus par la voie "algébrique".

Q2 *Supposons que la souris est au départ dans la pièce 2. Quelle est la probabilité qu'elle trouve la sortie du labyrinthe ?*

Il s'agit ici de calculer la probabilité $f_{2,6}$ d'aller de l'état 2 à l'état 6 en un nombre quelconque d'étape. D'après la formule (3.17), on obtient :

$$\begin{aligned} f_{2,6} &= p_{2,6} + \sum_{k \neq 6} p_{2,k} f_{k,6} = p_{2,1} f_{1,6} + p_{2,4} f_{4,6} \\ &= \frac{1}{2} f_{1,6} + \frac{1}{2} f_{4,6} \end{aligned}$$

D'autre part, en appliquant toujours cette formule (3.17), on obtient également :

$$\begin{aligned} f_{1,6} &= p_{1,6} + \sum_{k \neq 6} p_{1,k} f_{k,6} = p_{1,1} f_{1,6} + p_{1,2} f_{2,6} \\ &= \frac{1}{2} f_{1,6} + \frac{1}{2} f_{2,6} \end{aligned}$$

ainsi que :

$$\begin{aligned} f_{4,6} &= p_{4,6} + \sum_{k \neq 6} p_{4,k} f_{k,6} = p_{4,3} f_{3,6} \\ &= f_{3,6} \end{aligned}$$

et

$$\begin{aligned} f_{3,6} &= p_{3,6} + \sum_{k \neq 6} p_{3,k} f_{k,6} = p_{3,6} + p_{3,1} f_{1,6} + p_{3,2} f_{2,6} \\ &= \frac{1}{4} + \frac{1}{4} f_{1,6} + \frac{1}{4} f_{2,6} \end{aligned}$$

En combinant ces quatre équations, on obtient :

$$f_{1,6} = f_{2,6} = f_{3,6} = f_{4,6} = \frac{1}{2}$$

Ce résultat était prévisible. En effet, quelle que soit la pièce de départ parmi $\{1, 2, 3, 4\}$ la souris terminera dans un des deux états absorbants, l'état 5 ou l'état 6. Comme elle ne peut entrer dans l'état 5 ou dans l'état 6 qu'en passant par l'état 3 et comme les deux probabilités de transition $p_{3,6}$ et $p_{3,5}$ sont égales, il y a nécessairement une chance sur deux qu'elle termine dans la pièce 6 (et une chance sur deux qu'elle termine dans la pièce 5).

Q3 *Sachant que la souris est initialement dans la pièce 2, quel est le nombre moyen de passages qu'elle fera dans la pièce 3, avant de sortir en 6 ou d'être piégée dans la pièce 5 ?*

Pour répondre à cette question, nous allons établir un résultat général donnant le nombre moyen $R_{i,j}$ de passages à un état j , venant d'un autre état i . Afin d'obtenir une réponse, il faut supposer que l'état j est transitoire (i.e. $f_{j,j} < 1$). Ce nombre moyen de passages s'écrit

$$R_{i,j} := \sum_{n \geq 1} n \cdot P(n \text{ passages par } j | X_0 = i)$$

Nous décomposons l'évènement $\{n \text{ passages par } j\}$ en trois évènements succesifs :

$$\{\text{aller de } i \text{ à } j\} \cap \{\text{revenir } n \text{ fois en } j\} \cap \{\text{ne plus revenir en } j\}$$

On obtient alors

$$P(n \text{ passages par } j | X_0 = i) = f_{i,j} \cdot (f_{j,j})^{n-1} \cdot (1 - f_{j,j})$$

D'où

$$\begin{aligned} R_{i,j} &= \sum_{n \geq 1} n \cdot f_{i,j} \cdot (f_{j,j})^{n-1} \cdot (1 - f_{j,j}) \\ &= f_{i,j} (1 - f_{j,j}) \sum_{n \geq 1} n (f_{j,j})^{n-1} \\ &= f_{i,j} (1 - f_{j,j}) \sum_{n \geq 1} \left[\frac{d}{dx} x^n \right] \Big|_{x=f_{j,j}} \\ &= \frac{f_{i,j}}{1 - f_{j,j}} \end{aligned} \tag{3.23}$$

Le nombre moyen de passage par la pièce 3 en partant de la pièce 2 vaut donc :

$$R_{2,3} = \frac{f_{2,3}}{1 - f_{3,3}}$$

La probabilité $f_{3,3}$ peut être obtenue à l'aide de la formule (3.17), de manière similaire à ce que nous avons fait au paragraphe précédent :

$$f_{3,3} = \frac{1}{4}f_{1,3} + \frac{1}{4}f_{2,3}$$

On voit par ailleurs sur le graphe de la CMTD que, partant de l'état 1 ou de l'état 2, on finira certainement par passer par l'état 3, quoiqu'il arrive. On a donc $f_{2,3} = f_{1,3} = 1$ et, en conséquence $f_{3,3} = \frac{1}{2}$, donc $R_{2,3} = 2$.

A l'examen des réponses aux questions ci-dessus, il est facile d'imaginer que les calculs de probabilités sur des CMTD peuvent être très complexes et l'application des formules (3.16) ou (3.17) très laborieuse. A l'instar de ce que nous avons fait par le nombre moyen de passages par j (formule (3.23)), il peut être utile de proposer des résultats pour une famille de problèmes. En particulier, dans certains cas, on peut gagner à considérer la formulation de tous les problèmes d'une même famille sous forme matricielle (ou vectorielle), plutôt que de résoudre séparément toutes les équations scalaires qui résultent de l'application des formules (3.16) ou (3.17). C'est le cas, notamment, pour le calcul des temps de séjour ou des lois d'entrées dans des sous-classes qui sont abordés dans la section suivante.

3.5 Temps de séjour et lois d'entrée dans une sous-chaîne

3.5.1 Temps moyen d'entrée ou de séjour dans une sous-chaîne

Définition 14 (Temps de séjour dans une sous-chaîne) Soit $\{X_n\}_{n \in \mathbb{N}}$, une CMTD homogène sur un espace d'état E discret. Soit $F \subset E$ une sous-chaîne (un ensemble d'état). On appelle *temps d'entrée* dans F (ou encore *temps de séjour* dans $\bar{F} := E \setminus F$, la grandeur

$$T_F := \inf \{n \geq 0 \mid X_n \in F\}$$

Par convention, $T_F = +\infty$ lorsque l'état n'entre jamais dans F . On notera :

$$\mathbf{T}_i := E \left(T_F^{\{X_0=i\}} \right) = E(T_F \mid X_0 = i) , \quad i \in \bar{F} := E \setminus F$$

l'espérance du temps d'entrée dans F (ou du temps de séjour dans \bar{F}), partant de l'état initial $X_0 = i \in \bar{F}$

Examinons à nouveau l'exemple du labyrinthe illustré à la figure 3.1. Nous allons nous intéresser au temps de séjour dans le labyrinthe, c'est-à-dire au temps d'entrée dans la sous-chaîne absorbante $F := \{5, 6\}$ ou encore au temps de séjour dans la sous-chaîne $\bar{F} := \{1, 2, 3, 4\}$. L'examen du graphe

3.5. TEMPS DE SÉJOUR ET LOIS D'ENTRÉE DANS UNE SOUS-CHAÎNE 71

associé à la CMTD montre⁸

$$\begin{aligned}\mathbf{T}_1 &= p_{1,1}(1 + \mathbf{T}_1) + p_{1,2}(1 + \mathbf{T}_2) \\ \mathbf{T}_2 &= p_{2,1}(1 + \mathbf{T}_1) + p_{2,4}(1 + \mathbf{T}_4) \\ \mathbf{T}_3 &= (p_{3,5} + p_{3,6}) \cdot 1 + p_{3,1}(1 + \mathbf{T}_1) + p_{3,2}(1 + \mathbf{T}_2) \\ \mathbf{T}_4 &= (1 + \mathbf{T}_3)\end{aligned}$$

Ces équations peuvent être ré-écrites sous la forme :

$$\begin{aligned}\mathbf{T}_1 &= (p_{1,1} + p_{1,2}) + p_{1,1}\mathbf{T}_1 + p_{1,2}\mathbf{T}_2 \\ \mathbf{T}_2 &= (p_{2,1} + p_{2,4}) + p_{2,1}\mathbf{T}_1 + p_{2,4}\mathbf{T}_4 \\ \mathbf{T}_3 &= (p_{3,1} + p_{3,2} + p_{3,5} + p_{3,6}) + p_{3,1}\mathbf{T}_1 + p_{3,2}\mathbf{T}_2 \\ \mathbf{T}_4 &= (p_{4,3}) + p_{4,3}\mathbf{T}_3\end{aligned}$$

qui fait apparaître les sommes des lignes de la matrice de transition

$$p_{1,1} + p_{1,2} = p_{2,1} + p_{2,4} = p_{3,1} + p_{3,2} + p_{3,5} + p_{3,6} = p_{4,3} = 1$$

On a donc sous forme matricielle :

$$\begin{pmatrix} \mathbf{T}_1 \\ \mathbf{T}_2 \\ \mathbf{T}_3 \\ \mathbf{T}_4 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} + \begin{pmatrix} p_{1,1} & p_{1,2} & p_{1,3} & p_{1,4} \\ p_{2,1} & p_{2,2} & p_{2,3} & p_{2,4} \\ p_{3,1} & p_{3,2} & p_{3,3} & p_{3,4} \\ p_{4,1} & p_{4,2} & p_{4,3} & p_{4,4} \end{pmatrix} \begin{pmatrix} \mathbf{T}_1 \\ \mathbf{T}_2 \\ \mathbf{T}_3 \\ \mathbf{T}_4 \end{pmatrix}$$

On trouve pour la solution dans le cas du labyrinthe de la figure 3.1 :

$$\mathbf{T} := \begin{pmatrix} \mathbf{T}_1 \\ \mathbf{T}_2 \\ \mathbf{T}_3 \\ \mathbf{T}_4 \end{pmatrix} = \begin{pmatrix} 15 \\ 13 \\ 8 \\ 9 \end{pmatrix}$$

L'espérance du temps de séjour dans le labyrinthe

$$E(T_F) = E(E(T_F | X_0 = i)) = \sum_{i \in \bar{F}} \pi_i^{(0)} \cdot E(T_F | X_0 = i) = \boldsymbol{\pi}^{(0)} \mathbf{T}$$

dépend bien sûr de la distribution initiale de probabilité $\boldsymbol{\pi}^{(0)}$. Le raisonnement conduit dans l'exemple du labyrinthe restant valable pour toute CMTD homogène, nous venons d'établir le résultat suivant.

8. Par exemple, partant de la pièce 1, on va dans la pièce 1 avec une probabilité $\frac{1}{2}$ ou dans la pièce 2 avec une probabilité $\frac{1}{2}$ également. Dans les deux cas, le déplacement aura pris un pas de temps. Une fois arrivé dans la pièce 1, il nous faudra encore un temps moyen \mathbf{T}_1 pour sortir du labyrinthe, alors que dans la pièce 2, il nous faudra un temps moyen \mathbf{T}_2 .

Proposition 12 Soit $\{X_n\}_{n \in \mathbb{N}}$, une CMTD homogène sur un espace d'état E discret. Soit $F \subset E$ une sous-chaîne (un ensemble d'état). Soit \mathbf{T} le vecteur colonne des espérances des temps d'entrée dans F , de composantes $E(T_F | X_0 = i)$, avec $i \in \bar{F} := E \setminus F$. Notons

$$\mathbf{P}|_{\bar{F}}$$

la matrice obtenue en supprimant dans la matrice de transition les lignes et les colonnes qui correspondent aux états de F . On a alors :

$$\mathbf{T} = \mathbf{1} + \mathbf{P}|_{\bar{F}} \mathbf{T}$$

ou encore

$$\mathbf{T} = (\mathbf{I} - \mathbf{P}|_{\bar{F}})^{-1} \mathbf{1}$$

où \mathbf{I} désigne la matrice identité de même dimension que $\mathbf{P}|_{\bar{F}}$, tandis que $\mathbf{1}$ désigne le vecteur colonne de même dimension que \mathbf{T} dont toutes les composantes valent 1. Par conséquent l'espérance du temps d'entrée dans F vaut :

$$E(T_F) = \boldsymbol{\pi}^{(0)} (\mathbf{I} - \mathbf{P}|_{\bar{F}})^{-1} \mathbf{1}$$

Maintenant que nous avons caractérisé les temps moyen d'entrée dans une sous-chaîne, nous allons aborder la question des lois d'entrée. Dans l'exemple du labyrinthe, il s'agira par exemple de déterminer les probabilités d'entrer dans la sous-chaîne (absorbante) $\{5, 6\}$, soit par la pièce 5 (cul de sac), soit par la pièce 6 (sortie du labyrinthe), ce qui peut dans certaines circonstances avoir un intérêt pratique ...

3.5.2 Lois d'entrée dans une sous-chaîne

Définition 15 (Lois d'entrée) Soit $\{X_n\}_{n \in \mathbb{N}}$, une CMTD homogène sur un espace d'état E discret. Soit $F \subset E$ une sous-chaîne. La *probabilité d'entrer dans F en partant de i* est

$$\mathbf{V}^{(i)} := P(T_F < \infty | X_0 = i), \forall i \in E$$

Les *probabilités d'entrer dans F par l'état j* , en partant de l'état i sont

$$\mathbf{W}_j^{(i)} := P(X_{T_F} = j | X_0 = i), \forall j \in F, \forall i \in \bar{F} \text{ tel que } \mathbf{T}_i < \infty$$

Examinons l'exemple du labyrinthe illustré à la figure 3.1. Nous savons que la sous-chaîne $F := \{5, 6\}$ est absorbante. Nous nous attendons donc à trouver pour cette sous-chaîne les valeurs $\mathbf{V}^{(i)} = 1$ quel que soit $i \in \bar{F}$. L'examen du graphe associé à la CMTD montre que les $\mathbf{V}^{(i)}$ sont solutions du système :

$$\begin{aligned} \mathbf{V}^{(1)} &= p_{1,1} \mathbf{V}^{(1)} + p_{1,2} \mathbf{V}^{(2)} \\ \mathbf{V}^{(2)} &= p_{2,1} \mathbf{V}^{(1)} + p_{2,4} \mathbf{V}^{(4)} \\ \mathbf{V}^{(3)} &= p_{3,1} \mathbf{V}^{(1)} + p_{3,2} \mathbf{V}^{(2)} + p_{3,5} \mathbf{V}^{(5)} + p_{3,6} \mathbf{V}^{(6)} \\ \mathbf{V}^{(4)} &= p_{4,3} \mathbf{V}^{(3)} \end{aligned}$$

3.5. TEMPS DE SÉJOUR ET LOIS D'ENTRÉE DANS UNE SOUS-CHAÎNE73

avec $\mathbf{V}^{(5)} = \mathbf{V}^{(6)} = 1$ puisque les états 5 et 6 sont dans la classe F . Ces équations correspondent aux quatre premières lignes (i.e. les lignes qui correspondent aux états de \bar{F}) de la forme matricielle :

$$\mathbf{V} = \mathbf{P}\mathbf{V} \quad (3.24)$$

avec

$$\mathbf{V} := \begin{pmatrix} \mathbf{V}^{(1)} \\ \mathbf{V}^{(2)} \\ \mathbf{V}^{(3)} \\ \mathbf{V}^{(4)} \\ \mathbf{V}^{(5)} \\ \mathbf{V}^{(6)} \end{pmatrix}$$

Il faut de plus tenir compte des contraintes

$$\begin{aligned} \mathbf{V}^{(j)} &= 1, \forall j \in F := \{5, 6\} \\ 0 &\leq \mathbf{V}^{(i)} \leq 1, \forall i \in E \end{aligned} \quad (3.25)$$

On montre facilement que la seule solution de (3.24) avec les contraintes (3.25) est :

$$\mathbf{V}^{(1)} = \mathbf{V}^{(2)} = \mathbf{V}^{(3)} = \mathbf{V}^{(4)} = 1,$$

ce qui est bien le résultat attendu dans cet exemple. Examinons, maintenant les probabilités $\mathbf{W}_5^{(i)}$. Etant donné la symétrie du problème, on sait à l'avance que $\mathbf{W}_5^{(i)} = \mathbf{W}_6^{(i)} = \frac{1}{2}$, $\forall i \in \bar{F}$. Par ailleurs, on a comme précédemment

$$\begin{aligned} \mathbf{W}_5^{(1)} &= p_{1,1} \mathbf{W}_5^{(1)} + p_{1,2} \mathbf{W}_5^{(2)} \\ \mathbf{W}_5^{(2)} &= p_{2,1} \mathbf{W}_5^{(1)} + p_{2,4} \mathbf{W}_5^{(4)} \\ \mathbf{W}_5^{(3)} &= p_{3,1} \mathbf{W}_5^{(1)} + p_{3,2} \mathbf{W}_5^{(2)} + p_{3,5} \mathbf{W}_5^{(5)} + p_{3,6} \mathbf{W}_5^{(6)} \\ \mathbf{W}_5^{(4)} &= p_{4,3} \mathbf{W}_5^{(3)} \end{aligned}$$

avec, par définition, $\mathbf{W}_5^{(5)} = 1$ et $\mathbf{W}_5^{(6)} = 0$. Ces équations peuvent être ré-écrites sous la forme matricielle :

$$\mathbf{W}_5 = \mathbf{P}\mathbf{W}_5 \quad (3.26)$$

avec

$$\mathbf{W}_5 := \begin{pmatrix} \mathbf{W}_5^{(1)} \\ \mathbf{W}_5^{(2)} \\ \mathbf{W}_5^{(3)} \\ \mathbf{W}_5^{(4)} \\ \mathbf{W}_5^{(5)} \\ \mathbf{W}_5^{(6)} \end{pmatrix}$$

et les contraintes

$$\begin{aligned} \mathbf{W}_j^{(i)} &= \delta_{i,j}, \forall i, j \in F := \{5, 6\} \\ 0 &\leq \mathbf{W}_j^{(i)} \leq 1, \forall i \in \bar{F}, \forall j \in F \end{aligned} \quad (3.27)$$

On montre facilement que la seule solution de (3.26) avec les contraintes (3.27) est :

$$\mathbf{W}_5^{(1)} = \mathbf{W}_5^{(2)} = \mathbf{W}_5^{(3)} = \mathbf{W}_5^{(4)} = \frac{1}{2},$$

ce qui est bien le résultat attendu dans cet exemple. Le même raisonnement s'applique de manière symétrique au calcul des $\mathbf{W}_6^{(i)}$ et conduit au même résultat

$$\mathbf{W}_6^{(1)} = \mathbf{W}_6^{(2)} = \mathbf{W}_6^{(3)} = \mathbf{W}_6^{(4)} = \frac{1}{2}$$

Il est intéressant de noter que les vecteurs \mathbf{V} et \mathbf{W}_j ($j \in F$) sont solutions de la même équation (3.24) ou (3.26). Il s'agit de l'équation des vecteurs propres associés à la valeur propre $\lambda = 1$ de la matrice \mathbf{P} des probabilités de transition. Cependant, parmi l'infinité de solutions possibles pour cette équation, il faut choisir celle satisfaisant les conditions (3.25) pour obtenir la solution \mathbf{V} et celle satisfaisant les conditions (3.27) pour obtenir la solution \mathbf{W}_j . Le raisonnement conduit dans l'exemple du labyrinthe restant valable pour toute CMTD homogène, nous venons d'établir le résultat suivant.

Proposition 13 Soit $\{X_n\}_{n \in \mathbb{N}}$, une CMTD homogène sur un espace d'état E discret. Soit $F \subset E$ une sous-chaîne (un ensemble d'état). Soit \mathbf{V} le vecteur des probabilités d'entrée dans la classe F . Alors \mathbf{V} est solution des lignes de

$$\mathbf{V} = \mathbf{P}\mathbf{V}$$

qui correspondent aux états de \bar{F} , dans la matrice de Passage \mathbf{P} , et doit également satisfaire les contraintes :

$$\mathbf{V}^{(j)} = 1, \forall j \in F$$

Soit \mathbf{W}_j le vecteur des probabilités d'entrer dans la classe F par l'état j ($j \in F$). Alors \mathbf{W}_j est solution des lignes de

$$\mathbf{W}_j = \mathbf{P}\mathbf{W}_j$$

qui correspondent aux états de \bar{F} , dans la matrice de Passage \mathbf{P} , et doit également satisfaire les contraintes :

$$\mathbf{W}_j^{(i)} = \delta_{i,j}, \forall i \in F$$

Ce résultat clôt la section sur les temps de séjour et les lois d'entrée. Nous allons maintenant nous intéresser au comportement asymptotique (i.e. sur des horizons de temps très grands) du vecteur des probabilités d'état.

3.6 Régime permanent

Soit $\{X_n\}_{n \in \mathbb{N}}$, une CMTD homogène sur un espace d'état E discret. Le vecteur ligne $\boldsymbol{\pi}^{(n)}$ des *probabilités d'état* au temps n ($n \geq 0$) est défini composante par composante par :

$$\pi_j^{(n)} := P(X_n = j) \text{ , } \forall j \in E$$

La formule de Chapman-Kolmogorov a permis d'établir $\boldsymbol{\pi}^{(n)} = \boldsymbol{\pi}^{(0)} \mathbf{P}^n$ pour $n \geq 0$. On s'intéresse dans cette section au comportement asymptotique de la CMTD et plus particulièrement à la limite :

$$\lim_{n \rightarrow \infty} \boldsymbol{\pi}^{(n)}$$

Deux questions se posent :

Q1 *Cette limite existe-t-elle ?*

La question de l'existence n'est pas triviale. Une chaîne périodique ne possèdera en général pas de limite. Si nous prenons l'exemple de la CMTD périodique à deux états représentée à la figure 3.14, nous obtenons la matrice de transition :

$$\mathbf{P} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

dont les puissances s'écrivent :

$$\mathbf{P}^n = \begin{cases} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} & \text{si } n \text{ impair} \\ \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} & \text{si } n \text{ pair} \end{cases}$$

On obtient donc, pour le vecteur des probabilités d'état au temps n :

$$\boldsymbol{\pi}^{(n)} = \begin{cases} \begin{pmatrix} \pi_2^{(0)} & \pi_1^{(0)} \end{pmatrix} & \text{si } n \text{ impair} \\ \begin{pmatrix} \pi_1^{(0)} & \pi_2^{(0)} \end{pmatrix} & \text{si } n \text{ pair} \end{cases}$$

qui ne possède pas de limite pour $n \rightarrow \infty$, sauf pour $\pi_1^{(0)} = \pi_2^{(0)} = \frac{1}{2}$.



FIGURE 3.14 – CMTD périodique à deux états

Q2 Si la limite existe, comment la calculer ?

Une approche naïve consiste à calculer explicitement

$$\lim_{n \rightarrow \infty} \pi^{(0)} \mathbf{P}^n$$

Supposons, pour simplifier l'exposé, que la matrice de transition \mathbf{P} est unitairement diagonalisable⁹, c'est-à-dire qu'il existe une matrice unitaire \mathbf{U} (i.e ; $\mathbf{U}\mathbf{U}^T = \mathbf{U}^T\mathbf{U} = \mathbf{I}$), composée des vecteurs propres ortho-normés de \mathbf{P} et telle que :

$$\mathbf{P} = \mathbf{U}^T \mathbf{D} \mathbf{U}$$

où \mathbf{D} désigne la matrice diagonale, dont les éléments diagonaux sont les valeurs propres de \mathbf{P} . Pour calculer la limite ci-dessus, il faudrait alors évaluer

$$\pi^{(0)} \mathbf{P}^n = \pi^{(0)} \mathbf{U}^T \mathbf{D}^n \mathbf{U},$$

ce qui peut être très laborieux pour des CMTD complexes, même dans le cas d'une matrice \mathbf{P} diagonalisable.

Afin de répondre à ces deux questions, on dispose du résultat qui suit, relatif à l'existence de la distribution limite et au calcul du vecteur des probabilités stationnaires.

Proposition 14 (Distribution limite et distribution stationnaire)

Si la CMTD $\{X_n\}_{n \in \mathbb{N}}$, définie sur l'espace d'état discret E , est *irréductible* et *apériodique*, alors la distribution limite

$$\pi^{(\infty)} := \lim_{n \rightarrow \infty} \pi^{(n)}$$

existe. Elle est indépendante de la distribution initiale de probabilités (i.e. elle est la même, quelle que soit $\pi^{(0)}$). Deux cas sont possibles :

(i) *tous les états sont transitoires ou récurrents nuls*. On a alors

$$\pi_i^{(\infty)} = 0, \forall i \in E$$

(ii) *tous les états sont récurrents non nuls*. On a alors

$$\pi^{(\infty)} = \pi$$

où π est la *distribution stationnaire* de probabilités définie par :

$$\pi = \pi \mathbf{P} \tag{3.28}$$

avec la condition de normalisation

$$\|\pi\|_1 := \sum_{i \in E} |\pi_i| = \sum_{i \in E} \pi_i = 1 \tag{3.29}$$

9. La matrice des probabilités de transition n'est pas, dans le cas général, diagonalisable. Cependant un exposé plus général faisant appel à sa forme normale de Jordan alourdirait les notations et ne serait pas utile au propos. Ce propos consiste en à montrer que même dans le cas diagonalisable, le calcul explicite direct de $\pi_j^{(n)}$ est laborieux.

Nous admettrons ce résultat, en ce qui concerne l'existence. La valeur de la limite, lorsqu'elle existe, se montre facilement. Comme la chaîne est supposée irréductible, tous les états sont de même nature. Dans le cas où ils sont tous transitoires ou récurrents nuls, la chaîne ne reviendra plus dans un état $j \in E$, à partir d'un certain rang. La suite $\pi_j^{(n)}$ est donc nulle à partir de ce rang. Dans le cas où tous les états sont récurrents non nuls, la récurrence (3.9) donne, par passage à la limite ($n \rightarrow \infty$) à gauche et à droite, l'équation (3.28).

Remarque 9 Le vecteur π des probabilités stationnaires est aussi appelé *distribution stationnaire* de probabilités. Il est défini par les équations (3.28) et (3.29) qui montrent que c'est le vecteur propre à gauche de \mathbf{P} associé à la valeur propre 1, de norme $\|\pi\|_1 = 1$. Lorsqu'il existe (il doit alors être non nul), on dit que la chaîne admet un *régime stationnaire*. La distribution stationnaire décrit le comportement sur le long terme de la CMTD. La distribution stationnaire ne dépend pas de la distribution initiale de probabilité. C'est en fait la distribution des probabilités d'état obtenues lorsque le *régime stationnaire* est atteint, c'est-à-dire lorsque n est suffisamment grand (i.e. que la simulation a commencé depuis suffisamment longtemps) pour que les conditions initiales aient été oubliées (i.e. ne jouent plus qu'un rôle négligeable). Dans le cas d'une chaîne périodique, si la distribution stationnaire existe, π_j est la proportion du temps que la chaîne passe dans l'état j , en moyenne, lorsque $n \rightarrow \infty$, dans le cas d'une distribution uniforme de probabilités initiales.

Exemple 7 On considère la CMTD irréductible et apériodique représentée à la figure 3.15. On a dans ce cas

$$\mathbf{P} = \begin{pmatrix} 0.6 & 0.4 & 0 \\ 0.2 & 0.6 & 0.2 \\ 0 & 0.4 & 0.6 \end{pmatrix}$$

Le calcul de la distribution stationnaire d'après les équations (3.28) et (3.29) donne

$$\pi = (0.25 \quad 0.5 \quad 0.25)$$

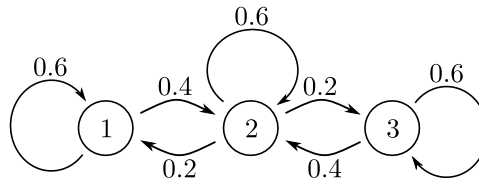


FIGURE 3.15 – CMTD irréductible et apériodique considérée dans l'exemple 7

C'est aussi dans ce cas la valeur de la distribution limite $\pi^{(\infty)}$. Le tableau ci-dessous donne l'évolution des probabilités d'état au temps n , pour $n \in \{0, \dots, 20\}$, pour les trois conditions initiales $X_0 = 1$, $X_0 = 2$ et $X_0 = 3$. On peut y observer que les conditions initiales ont de moins en moins d'influence, au fur et à mesure que n augmente et que la chaîne se rapproche du régime stationnaire.

$\pi^{(0)}$	[1 0 0]	[0 1 0]	[0 0 1]
$\pi^{(1)}$	[0.60000 0.40000 0.00000]	[0.20000 0.60000 0.20000]	[0.00000 0.40000 0.60000]
$\pi^{(2)}$	[0.44000 0.48000 0.08000]	[0.24000 0.52000 0.24000]	[0.08000 0.48000 0.44000]
$\pi^{(3)}$	[0.36000 0.49600 0.14400]	[0.24800 0.50400 0.24800]	[0.14400 0.49600 0.36000]
$\pi^{(4)}$	[0.31520 0.49920 0.18560]	[0.24960 0.50080 0.24960]	[0.18560 0.49920 0.31520]
$\pi^{(5)}$	[0.28896 0.49984 0.21120]	[0.24992 0.50016 0.24992]	[0.21120 0.49984 0.28896]
$\pi^{(6)}$	[0.27334 0.49997 0.22669]	[0.24998 0.50003 0.24998]	[0.22669 0.49997 0.27334]
$\pi^{(7)}$	[0.26400 0.49999 0.23601]	[0.25000 0.50001 0.25000]	[0.23601 0.49999 0.26400]
$\pi^{(8)}$	[0.25840 0.50000 0.24160]	[0.25000 0.50000 0.25000]	[0.24160 0.50000 0.25840]
$\pi^{(9)}$	[0.25504 0.50000 0.24496]	[0.25000 0.50000 0.25000]	[0.24496 0.50000 0.25504]
$\pi^{(10)}$	[0.25302 0.50000 0.24698]	[0.25000 0.50000 0.25000]	[0.24698 0.50000 0.25302]
$\pi^{(11)}$	[0.25181 0.50000 0.24819]	[0.25000 0.50000 0.25000]	[0.24819 0.50000 0.25181]
$\pi^{(12)}$	[0.25109 0.50000 0.24891]	[0.25000 0.50000 0.25000]	[0.24891 0.50000 0.25109]
$\pi^{(13)}$	[0.25065 0.50000 0.24935]	[0.25000 0.50000 0.25000]	[0.24935 0.50000 0.25065]
$\pi^{(14)}$	[0.25039 0.50000 0.24961]	[0.25000 0.50000 0.25000]	[0.24961 0.50000 0.25039]
$\pi^{(15)}$	[0.25024 0.50000 0.24976]	[0.25000 0.50000 0.25000]	[0.24976 0.50000 0.25024]
$\pi^{(16)}$	[0.25014 0.50000 0.24986]	[0.25000 0.50000 0.25000]	[0.24986 0.50000 0.25014]
$\pi^{(17)}$	[0.25008 0.50000 0.24992]	[0.25000 0.50000 0.25000]	[0.24992 0.50000 0.25008]
$\pi^{(18)}$	[0.25005 0.50000 0.24995]	[0.25000 0.50000 0.25000]	[0.24995 0.50000 0.25005]
$\pi^{(19)}$	[0.25003 0.50000 0.24997]	[0.25000 0.50000 0.25000]	[0.24997 0.50000 0.25003]
$\pi^{(20)}$	[0.25002 0.50000 0.24998]	[0.25000 0.50000 0.25000]	[0.24998 0.50000 0.25002]

Le régime stationnaire, lorsqu'il existe, correspond à une sorte d'équilibre. La distribution stationnaire est un point fixe pour l'itération (3.9). L'équation de point fixe (3.28) pour la colonne j donne :

$$\pi_j = \sum_i \pi_i p_{i,j}, \quad \forall j \in E$$

ce qui peut aussi s'écrire

$$\left(\sum_i p_{j,i} \right) \pi_j = \sum_i \pi_i p_{i,j},$$

car $\sum_i p_{j,i} = 1, \forall j$. On a donc

$$\sum_i \pi_j p_{j,i} = \sum_i \pi_i p_{i,j}, \quad \forall j \in E \quad (3.30)$$

Le membre de gauche correspond au *flux moyen de sortie de l'état j* , c'est-à-dire le nombre moyen de transition de l'état j vers tous les autres états

i , par unité de temps, sur le long terme. Le membre de droite correspond au *flux moyen d'entrée dans l'état j* . En régime stationnaire, on a donc l'équilibre en chaque état :

$$\forall j \in E : \text{flux moyen sortant de } j = \text{flux moyen entrant dans } j$$

Dans le cas où la CMTD n'est pas irréductible, mais composée de plusieurs classes d'équivalence, on pourra s'intéresser à la probabilité d'entrer dans l'une ou l'autre de ces classes d'équivalence (qui sont entre autres des sous-chaînes), puis à la distribution stationnaire (ou limite) de probabilités, une fois dans l'une des ces sous-chaînes. Reprenons pour illustrer cela l'exemple de la figure 3.13. Pour cet exemple, la probabilité, partant de l'état 1, de tomber dans la classe $\{6, 7\}$ vaut $f_{1,6}$. En appliquant la formule (3.17) pour le calcul des $f_{i,j}$, on trouve :

$$\begin{aligned} f_{1,6} &= p_{1,6} + \sum_{k \neq 6} p_{1,k} f_{k,6} \\ &= 0.25 + 0.5f_{2,6} + 0.25f_{4,6} \end{aligned}$$

avec $f_{4,6} = 0$. Toujours avec (3.17), on obtient également :

$$\begin{aligned} f_{2,6} &= \sum_{k \neq 6} p_{2,k} f_{k,6} \\ &= 0.5f_{1,6} + 0.5f_{3,6} \end{aligned}$$

avec $f_{3,6} = 0$. On obtient finalement, pour la probabilité de finir dans la classe $\{6, 7\}$ partant de l'état 1, la valeur $f_{1,6} = \frac{1}{3}$. La probabilité de finir dans la classe $\{3, 4, 5\}$ partant de l'état 1, vaut donc $1 - f_{1,6} = \frac{2}{3}$. Une fois dans la classe $\{6, 7\}$ qui est périodique, tous les états sont récurrents non nuls (et périodiques). La distribution stationnaire est solution de :

$$\begin{pmatrix} \pi_6 & \pi_7 \end{pmatrix} = \begin{pmatrix} \pi_6 & \pi_7 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} ; \pi_6 + \pi_7 = 1$$

On obtient $\pi_6 = \pi_7 = \frac{1}{2}$. Une fois dans la classe $\{3, 4, 5\}$ qui est irréductible et apériodique, la distribution stationnaire est solution de :

$$\begin{pmatrix} \pi_3 & \pi_4 & \pi_5 \end{pmatrix} = \begin{pmatrix} \pi_3 & \pi_4 & \pi_5 \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0.5 & 0.5 & 0 \end{pmatrix} ; \pi_3 + \pi_4 + \pi_5 = 1$$

On obtient $\pi_3 = \frac{1}{5}$ et $\pi_4 = \pi_5 = \frac{2}{5}$.

3.7 Exercices

Exercice 11

Les études dans une école d'ingénieurs durent 3 ans¹⁰. A l'issue de chaque année, un étudiant peut redoubler avec la probabilité r , passer dans la classe suivante avec la probabilité p , ou être exclu avec la probabilité e . On supposera dans un premier temps que ces probabilités ne dépendent ni de l'année d'étude de l'étudiant, ni du nombre de redoublements éventuels de celui-ci ... Ces hypothèses pourront être levées ensuite pour proposer un modèle plus réaliste !

1. Modéliser le cursus d'un élève par une chaîne de Markov à cinq états.
2. Quatre ans viennent de s'écouler depuis que Robert est entré à l'école. Que peut on savoir sur l'année d'étude dans laquelle se trouve Robert ?
3. Calculer la probabilité qu'un élève, venant de rentrer dans l'école, arrive à avoir son diplôme. Même question avec un élève de deuxième, puis de troisième année.
4. Calculer la durée moyenne que passe un étudiant à l'école, avant l'obtention du diplôme ou son renvoi, en supposant qu'il entre en première année.
5. Proposez une modélisation plus réaliste du cursus "ingénieur" à L'ESI-SAR dans laquelle un seul redoublement est permis.

Exercice 12

On dispose de 2 machines qui fonctionnent indépendamment l'une de l'autre. Au cours d'une journée chaque machine peut tomber en panne avec la probabilité $1 - p$.

1. On suppose que l'on peut réparer la nuit une seule machine qui marchera ainsi le lendemain. Modéliser le système par une chaîne de Markov (donner son graphe, sa matrice)
2. Même question si la réparation ne se fait pas pendant la nuit mais le lendemain.
3. Calculer dans le deuxième cas (réparations en journée) la distribution stationnaire de probabilités. Quelle est la probabilité pour un ouvrier qui arrive le matin de se retrouver en chômage technique dans ce deuxième cas ?

Exercice 13

Dans une maison constituée de 4 pièces se déroule un cruel épisode de la vie domestique. Une souris se promène pour trouver à manger. Un chat¹¹

10. Toute ressemblance avec des établissements actuels ou ayant existé serait purement fortuite et ne pourrait être que le fruit d'une pure coïncidence...

11. qui lui n'a pas ce genre de soucis, c'est une larve gavée par ses maîtres

décide d'attraper la souris.

La souris démarre sa recherche au hasard dans la maison (mais pas par la cuisine, où la nourriture se trouve), elle reste 10 minutes dans chaque pièce, puis ayant oublié où elle a déjà cherché, change d'endroit pour une des 3 autres pièces. Le chat, en bonne larve qu'il est, ne veut pas courir. Il se place dans la cuisine et attend la souris.

1. Pourquoi peut-on modéliser cette situation par une chaîne de Markov ?
2. Donner le graphe et la matrice de transition de cette chaîne.
3. Classifier les états de la chaîne.
4. Quelle est l'espérance de vie de la souris ?
5. Déterminer le régime stationnaire de la chaîne (i.e. donner la distribution stationnaire). Etait-il possible de donner à priori, sans calcul, cette distribution stationnaire de probabilités ?

Exercice 14

Une entreprise utilise des machines dont l'usure est caractérisée par quatre valeurs, notées $\{1, 2, 3, 4\}$, qui correspondent respectivement aux états des machines *neuves*, *usées*, *très usées* et *inutilisables*. Chaque jour, une machine dans un de ces états rapporte à l'entreprise respectivement 1000 €, 1000 €, 800 € et 0 €. On modélise le processus de vieillissement de la machine par une chaîne de Markov à temps discret (l'unité est le jour) de matrice de transition :

$$\mathbf{P} = \begin{pmatrix} 0.6 & 0.2 & 0.1 & 0.1 \\ 0 & 0.7 & 0.2 & 0.1 \\ 0 & 0 & 0.8 & 0.2 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

1. Calculer la durée de vie moyenne d'une machine neuve. Même question avec une machine très usée.
2. Le remplacement de la machine coûte K € et demande une journée (pendant laquelle la machine ne produit rien). Choisir parmi ces deux politiques laquelle est la plus profitable à long terme pour l'entreprise :
 P1 : remplacer la machine dès qu'elle est *inutilisable*
 P2 : remplacer la machine dès qu'elle est *très usée* ou *inutilisable*

Exercice 15

Une information (de type binaire 0 ou 1) est transmise d'un émetteur R_0 , jusqu'à un récepteur R_n à travers une suite de n relais intermédiaires désignés par R_1, R_2, \dots, R_{n-1} ($n \geq 2$). Chaque relais R_k transmet l'information reçue de R_{k-1} au suivant R_{k+1} avec la probabilité p ($0 < p < 1$). Il transmet donc l'information contraire à celle reçue (erreur) avec la probabilité $1 - p$. Ces probabilités sont indépendantes des informations transmises par les autres relais, en amont ou en aval de la chaîne.

1. Modéliser ce problème par une chaîne de Markov.
2. Calculer p_n la probabilité pour l'information initiale soit correctement transmise de R_0 à R_n .
3. Calculer la limite de p_n , pour $n \rightarrow \infty$

Exercice 16

On considère une suite aléatoire de bits indépendants b_n . Chacun de ces bits est distribué selon la même loi :

$$\begin{aligned} P(b_i = 1) &= p \\ P(b_i = 0) &= 1 - p \end{aligned}$$

pour tout $i \geq 1$. Soit X_n = longueur du dernier bloc de 1 touchant le n^e bit (cela implique notamment que le n^e bit lui-même est un 1). Par exemple, pour la suite

$$0101110 \underbrace{11111}_{\text{bloc de 1}} 100110...$$

on a $X_{12} = 5$, $X_{13} = 6$ et $X_{14} = 0$, alors que pour la suite

$$01011100000 \underbrace{0}_{\text{0}} 110010...$$

on a $X_{12} = 0$, $X_{13} = 1$ et $X_{14} = 2$

1. Montrer que X_n est une chaîne de Markov dont on calculera les probabilités de transition.
2. Montrer que cette chaîne est irréductible et apériodique
3. Calculer la distribution limite de probabilités.

Exercice 17 (Les urnes d'Ehrenfest)

On considère deux urnes A et B contenant initialement b boules chacune. A chaque étape, on choisit une boule parmi les $2b$ et on la change d'urne. On note X_n le nombre de boules dans l'urne A à l'étape n .

1. Montrer que $(X_n)_{n \in \mathbb{N}}$ une chaîne de Markov, et déterminer les probabilités de transition de cette chaîne.
2. On prend $b = 5$. Calculer la distribution stationnaire de la chaîne. Est-elle la loi limite ?
3. Si $b = 5$, déterminer le nombre moyen de transitions avant que l'une des deux urnes ne se vide.

Exercice 18 (Pagerank)

On examine dans cet exercice le fonctionnement de l'algorithme *Pagerank* de Google. On note N le nombre de pages internet existantes, on note r_i le

nombre de liens existant sur la page i .

On modélise l'attitude d'un surfer de la façon suivante : une fois visitée la page i , soit il choisit une adresse Web au hasard (avec la probabilité p), soit il décide de visiter une des pages référencées sur la page i (avec la probabilité $1 - p$). Dans le cas où il choisit une adresse au hasard, on considère que chaque page a la même probabilité $\frac{1}{N}$ d'être choisie. Dans le cas où le surfer choisit un des liens de la page i , il le fait également avec une distribution uniforme. Dans ce cas, chaque page a donc une probabilité $\frac{1}{r_i}$.

On considère dans cet exercice un Web miniature¹² constitué de 5 pages et représenté à la figure 6.12.

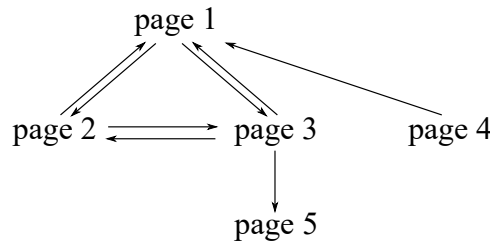


FIGURE 3.16 – Le Web simplifié considéré à l'exercice 18

1. Modéliser la promenade d'un surfer sur le Web de la figure 6.12 par une chaîne de Markov dont on donnera le graphe et la matrice de transition.
2. Pourquoi existe-t-il nécessairement une loi limite $\pi^{(\infty)}$? Comment peut-on la calculer ?
3. Soit $\pi = (\pi_1, \pi_2, \dots, \pi_N)$ la distribution stationnaire de probabilité. La mesure *Pagerank* de la page i est alors simplement définie par

$$\text{Pagerank}(i) := \pi_i$$

Calculer le *Pagerank* pour toutes les pages du Web de la figure 6.12, successivement pour $p = 0.15$, $p = 10^{-2}$ et $p = 0.99$.

4. Imaginer une méthode de calcul itérative de la mesure *Pagerank* qui soit adaptée à un Web de très grande taille mais de structure creuse (la matrice d'incidence du graphe associé est creuse ou, de manière équivalente, une page ne pointe que sur quelques autres pages)

12. Dans le cas du Web mondial, les principes de modélisation et de définition de la mesure *pagerank* restent les mêmes. Les méthodes de modélisation, d'identification et de calcul, par contre, sont adaptées à des CMTD de très grande taille. On pourra pour une introduction rapide au sujet se reporter par exemple à la page Wikipedia en anglais sur le Pagerank.

Exercice 19

Soit $n \in \mathbb{N}^*$. On pose $E = \{0, 1, \dots, n\}$.

On considère la chaîne de Markov (X_n) sur E de matrice de transition P donnée, pour $0 \leq i \leq n-1$ par

$$P(i, j) = \begin{cases} p & \text{si } j = (i + 1) \\ q & \text{si } j = 0 \\ 0 & \text{sinon} \end{cases}$$

avec $p, q \in]0, 1[$ et $p + q = 1$. L'état n est absorbant.

1. Faire le graphe de (X_n) et classer les états de la chaîne.
2. On note T le temps d'entrée en n . Que vaut $E_x(T)$, pour $x \in E$?
3. Une pièce donne pile avec la probabilité p . On la lance successivement plusieurs fois. Soit $n \in \mathbb{N}^*$ fixé.
 - (a) Montrer que $P(\text{obtenir } n \text{ piles consécutifs en un nombre fini de lancers}) = 1$.
 - (b) Combien de lancers sont nécessaires en moyenne?

Exercice 20

On considère la chaîne de Markov (X_n) définie sur l'espace d'état $E = \{1, 2, \dots, 5\}$, dont la matrice de transition est donnée par

$$\mathbf{P} = \begin{pmatrix} \frac{1}{2} & 0 & \frac{1}{12} & \frac{1}{4} & \frac{1}{6} \\ 0 & \frac{1}{4} & 0 & 0 & \frac{3}{4} \\ \frac{1}{3} & \frac{1}{12} & \frac{1}{2} & 0 & \frac{1}{12} \\ 0 & \frac{1}{3} & \frac{1}{4} & \frac{1}{4} & \frac{1}{6} \\ 0 & \frac{1}{3} & 0 & 0 & \frac{2}{3} \end{pmatrix}$$

1. Classer les états de la chaîne. On notera dans la suite C , la classe formée par les états récurrents.
2. On note $T = \inf\{n \geq 0, X_n \in C\}$ le temps d'entrée dans C . Calculer $E_x(T)$, l'espérance de T partant de l'état x , pour tout $x \in \bar{C} = E \setminus C$.
3. Calculer la loi d'entrée dans C , c'est-à-dire les probabilités $P_x(X_T = y)$ si d'entrer dans la classe C par l'état $y \in C$, partant de $x \in \bar{C}$.

Exercice 21

On considère un codage bipolaire, qui associe à la suite de symboles a_k la suite b_k de la façon suivante :

$$\begin{aligned} a_k = 0 & \Rightarrow b_k = 0 \\ a_k = 1 & \Rightarrow b_k = 1 \text{ ou } -1 \text{ alternativement} \end{aligned}$$

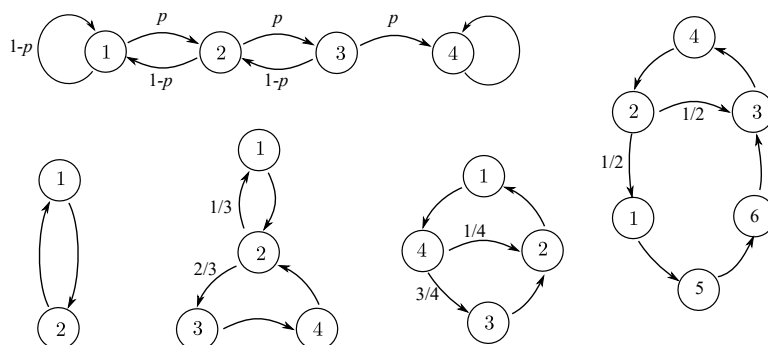


FIGURE 3.17 – Les chaînes de Markov pour l'exercice 22

On suppose que la suite (a_k) est une suite de symboles indépendants, avec $P(a_k = 1) = p$ et $P(a_k = 0) = 1 - p$

1. Montrer que (b_k) n'est pas une chaîne de Markov.
2. Proposer un modèle markovien du système en 4 états. Tracer son graphe
3. Calculer la loi limite de b_k quand $k \rightarrow +\infty$

Exercice 22

Pour les cinq chaînes de Markov dont les graphes sont représentés à la figure 6.18, déterminer la matrice de transitions \mathbf{P} , les classes d'équivalence, la nature des états de ces classes (récurrents ou transitoires), ainsi que les éventuels états absorbants ou périodiques (dans le cas d'états périodiques, donner leur période).

Chapitre 4

Chaînes de Markov à temps continu

4.1 Introduction et définition

Nous allons dans ce chapitre considérer des processus stochastiques à espace d'état discret E (i.e. une chaîne), mais cette fois à *temps continu*. Nous noterons $(X(t))_{t \geq 0}$ la famille de variables aléatoires à valeurs dans E , qui représente l'évolution au cours du temps t de cette chaîne. Nous noterons $(x(t))_{t \geq 0}$ une réalisation particulière de la chaîne. Il s'agira en général d'une application constante par morceau de \mathbb{R}^+ dans $E := \mathbb{N}$. Une telle réalisation est représentée à la figure 4.1.

Les *chaînes de Markov à temps continu* (ou *CMTC*, en abrégé) constituent une sous-classe particulière des chaînes à temps continu : elles sont caractérisées par l'absence de mémoire. Les probabilités de transitions entre états après le temps t ne dépendront pas des transitions réalisées avant le temps t . Les processus de Poisson - étudiés dans le premier chapitre du cours - sont des exemples particuliers des tels processus sans mémoire et de chaînes de Markov en temps continu.

La décision de modéliser un processus stochastique par une CMTC, plutôt que par une CMTD, sera souvent motivée par la nécessité (ou non) d'avoir des informations sur le temps : temps entre deux transitions, temps d'attente dans une file, etc. Un modèle sous la forme d'une CMTD fournira une estimation des temps en nombre de transitions, sans permettre pour autant de caractériser les durées correspondantes (on ne sait pas combien de temps s'écoule entre deux transitions). Au contraire, un modèle sous la forme de CMTC permet de représenter un processus stochastique où les événements peuvent survenir à n'importe quels moments (*systèmes à événements discrets* ou *SED*, en abrégé) et d'analyser les durées entre ces événements.

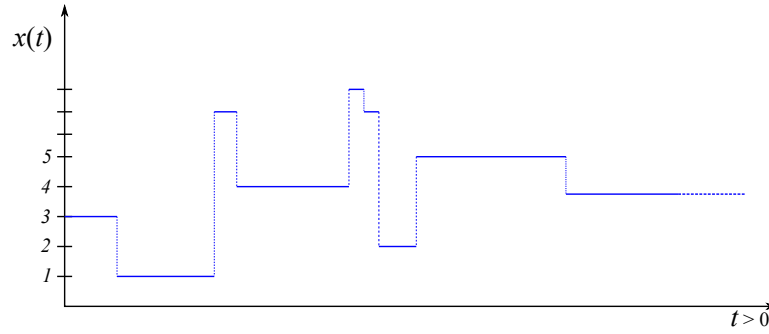


FIGURE 4.1 – Une réalisation particulière d’une chaîne de Markov à temps continu (CMTC). L’espace d’état discret E sera choisi (sans restriction) comme l’ensemble des entiers naturels \mathbb{N} , par soucis de simplicité des notations. Les sauts entre états peuvent survenir à n’importe quels instants.

Définition 16 (chaîne de Markov à temps continu)

Soit $(X(t))_{t \geq 0}$, une famille de variables aléatoires à valeurs dans E , un espace d’état discret. Alors, par définition, $(X(t))_{t \geq 0}$ est une *chaîne de Markov à temps continu* si et seulement si

$$\begin{aligned} P(X(t_n) = x_n | X(t_{n-1}) = x_{n-1}; \dots; X(t_1) = x_1) = \\ P(X(t_n) = x_n | X(t_{n-1}) = x_{n-1}) \\ \forall n \in \mathbb{N}, \forall (x_1, \dots, x_n) \in E^n \end{aligned} \quad (4.1)$$

et quels que soient les instants d’échantillonnages

$$0 \leq t_1 < t_2 < \dots < t_{n-1} < t_n \quad (4.2)$$

Dans le cas d’une chaîne en temps continu, on ne dispose généralement pas d’un historique complet de l’état, c’est à dire d’une observation continue $x(s)$ pour tout $s < t$, mais seulement d’une suite d’observations à certains instants choisis $t_1 < t_2 < \dots < t_{n-1} < t_n$. Dans une chaîne de Markov (en temps continu), la connaissance détaillée de l’ensemble de ces observations passées ne fournit pas davantage d’information que la connaissance de la dernière observation uniquement. La seule différence formelle entre la définition 16 et celle d’une chaîne de Markov en temps discret est la présence des instants d’échantillonnage arbitraires (4.2). Pour le reste, la caractérisation de la propriété markovienne (absence de mémoire) est exactement la même dans les deux cas.

Dans la suite, nous nous intéresserons essentiellement aux CMTC homogènes, c’est-à-dire celles pour lesquelles le modèle (i.e. les probabilités de transition entre états) ne varie pas au cours du temps. Cette hypothèse n’est pas toujours réaliste mais elle permet de résoudre analytiquement la plupart des problèmes que nous poserons sur les CMTC. Comme dans le

cas des CMTD, une mise à jour *en ligne* des probabilités de transition est toujours possible, pour tenir compte de variations lentes dans le processus et adapter en conséquence son modèle.

Définition 17 (CMTC homogène) Soit $(X(t))_{t \geq 0}$ une chaîne de Markov à temps continu à valeurs dans E . Elle est dite *homogène* si et seulement si la probabilité

$$P(X(t_n) = x_n | X(t_{n-1}) = x_{n-1})$$

ne dépend pas des instants t_n et t_{n-1} , mais seulement de l'écart

$$\Delta t_n := t_n - t_{n-1}$$

entre les deux instants d'observation.

Comme précédemment, afin de simplifier les notations, on considérera dans la suite l'espace d'état $E = \{1, 2, \dots, i-1, i, i+1, \dots\}$ dans le cas infini dénombrable ou $E = \{1, 2, \dots, N\}$ dans le cas fini. Avec cette convention de notation, il est possible de définir, pour une CMTC homogène, les probabilités de transition durant un intervalle de temps de longueur t de la manière suivante.

Définition 18 (Probabilités de transition) Soit $(X(t))_{t \geq 0}$ une chaîne de Markov à temps continu *homogène* à valeurs dans E . Les probabilités de transition entre états, pendant un intervalle de temps t sont définies par :

$$p_{i,j}(t) := P(X(s+t) = j | X(s) = i), \forall t > 0, \forall i, j \in E \quad (4.3)$$

Elles sont indépendantes du temps $s > 0$ où elles sont calculées et ne dépendent que de la durée t de l'intervalle. Nous ferons également l'hypothèse :

$$p_{i,j}(0) = \lim_{t \rightarrow 0} p_{i,j}(t) = \delta_{i,j} \quad (4.4)$$

avec

$$\delta_{i,j} = \begin{cases} 1 & \text{si } i = j \\ 0 & \text{sinon} \end{cases}$$

L'hypothèse (4.4) est relativement naturelle. Elle revient à supposer la continuité de fonction $p_{i,j}(t)$ en $t = 0$ ou, de manière plus intuitive, à supposer que la probabilité qu'il y ait un changement d'état sur une période infiniment courte est nulle.

Exemple 8 (Processus de Poisson) Les processus de Poisson, étudiés en détail précédemment, sont un premier exemple de chaînes de Markov

en temps continu. Considérons un tel processus de Poisson $(N(t))_{t \geq 0}$ de paramètre λ . Par définition de la probabilité conditionnelle, on a :

$$\begin{aligned} \tilde{P} &:= P(N(t_n) = k_n | N(t_{n-1}) = k_{n-1}; \dots; N(t_1) = k_1) \\ &= \frac{P(N(t_n) = k_n; N(t_{n-1}) = k_{n-1}; \dots; N(t_1) = k_1)}{P(N(t_{n-1}) = k_{n-1}; \dots; N(t_1) = k_1)}, \end{aligned} \quad (4.5)$$

$\forall n \in \mathbb{N}, \forall k_1, \dots, k_n \in \mathbb{N}, \forall 0 \leq t_1 < \dots, t_n$

Le membre de droite dans (4.5) peut s'écrire de manière équivalente :

$$\begin{aligned} \tilde{P} &= \frac{P(N(t_n) - N(t_{n-1}) = k_n - k_{n-1}; \dots; N(t_2) - N(t_1) = k_2 - k_1; N(t_1) = k_1)}{P(N(t_{n-1}) = k_{n-1}; \dots; N(t_1) = k_1)} \\ &= P(N(t_n) - N(t_{n-1}) = k_n - k_{n-1}) \cdot \frac{P(\dots; N(t_2) - N(t_1) = k_2 - k_1; N(t_1) = k_1)}{P(N(t_{n-1}) = k_{n-1}; \dots; N(t_1) = k_1)} \\ &= P(N(t_n) - N(t_{n-1}) = k_n - k_{n-1}) \cdot \frac{P(N(t_{n-1}) = k_{n-1}; \dots; N(t_1) = k_1)}{P(N(t_{n-1}) = k_{n-1}; \dots; N(t_1) = k_1)} \\ &= P(N(t_n) - N(t_{n-1}) = k_n - k_{n-1}) \end{aligned}$$

par l'hypothèse des accroissements indépendants¹. Or, on sait que cette dernière probabilité vaut :

$$\begin{aligned} \tilde{P} &= P(N(t_n) - N(t_{n-1}) = k_n - k_{n-1}) \\ &= \frac{(\lambda(t_n - t_{n-1}))^{k_n - k_{n-1}}}{(k_n - k_{n-1})!} e^{-\lambda(t_n - t_{n-1})} \\ &= P(N(t_n) = k_n | N(t_{n-1}) = k_{n-1}) \end{aligned} \quad (4.6)$$

Nous venons donc de montrer que $(N(t))_{t \geq 0}$ est une chaîne de Markov au sens de la définition 16. De plus, l'équation (4.6) montre que les probabilités de transition ne dépendent que de l'état initial k_{n-1} , de l'état final k_n et de la durée $\Delta t_n := t_n - t_{n-1}$. Il s'agit donc d'une CMTC homogène au sens de la définition 17. Enfin, nous venons de calculer les probabilités :

$$p_{i,j}(t) := P(N(s+t) = j | N(s) = i) = \begin{cases} \frac{(\lambda t)^{j-i}}{(j-i)!} e^{-\lambda t} & \text{si } j \geq i \\ 0 & \text{sinon} \end{cases} \quad (4.7)$$

qui ne sont rien d'autres que les probabilités de transition au sens de la définition 18. Elles vérifient bien sûr la condition de continuité (4.4).

Nous allons maintenant examiner un exemple plus simple de CMTC sur un espace d'état E fini. Celui-ci nous servira à illustrer les différentes notions abordées tout au long du chapitre.

1. Dans un processus de Poisson, les incréments (arrivées) sur $[t_1, t_{n-1}]$ et sur $[t_{n-1}, t_n]$ sont supposés être des événements aléatoires indépendants

Exemple 9 (Modèle étudiant) Dans cet exemple, un enseignant consciencieux cherche à modéliser le comportement quotidien des étudiants d'une célèbre école d'ingénieurs valentinoise. Il s'intéresse à leur emploi du temps et, en particulier, à la répartition du temps que passent quotidiennement ces étudiants à dormir (état 1, la résidence étudiante), à boire (état 6, le bar), à étudier (état 5, l'école), ou dans les transports. Les états 2, 3, 4, 7 et 8 correspondent respectivement à l'attente au pôle bus, au trajet aller en empruntant la ligne 3, au trajet aller en empruntant la ligne 4, à l'attente à l'arrêt Briffaut et au trajet retour via la ligne 3. Il est admis que les élèves ingénieur ne fréquentent pas d'autres lieux et - en particulier - fréquentent tous le même bar à Valence.

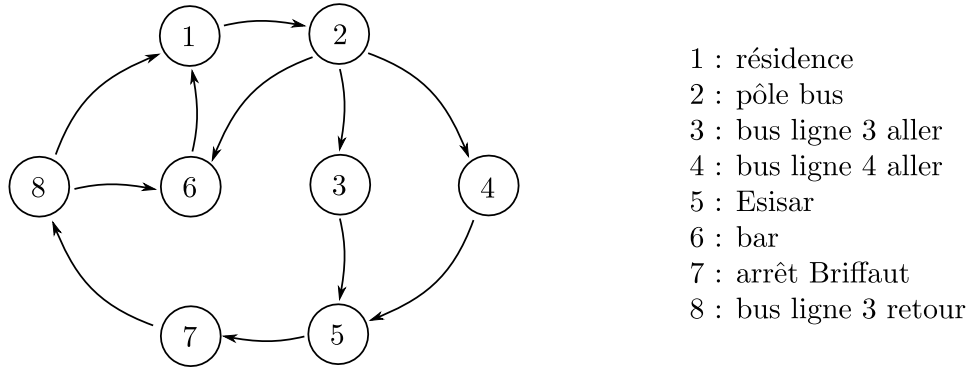


FIGURE 4.2 – Modèle d'un élève : répartition du temps passé à dormir (résidence), à boire (bar), à se déplacer (bus, voiture) et à l'école

Quelques hypothèses simplificatrices sont faites pour pouvoir représenter les journées des étudiants comme une marche aléatoire markovienne en temps continu sur la graphe de la figure 4.2 et calculer explicitement les solutions aux différents problèmes posés :

- (i) les temps de séjour dans les différents états (résidence, Esisar, bar, transport aller ou retour) sont des exponentielles indépendantes les unes des autres, chacune avec son propre paramètre λ_i , $i \in \{1, 3, 4, 5, 6, 8\}$
- (ii) les bus se présentent aux arrêts (pôle bus, Briffaut) selon des processus de Poisson, chacun avec un paramètre spécifique, respectivement λ_{23} et λ_{24} pour les arrivées des bus 3 et 4 à l'arrêt pôle bus², et λ_{73} pour l'arrivée du bus 3 à l'arrêt Briffaut.

2. Il faut imaginer que les temps de trajet des bus 3 et 4 sont différents, mais aussi leurs fréquences respectives, ce qui peut justifier l'intérêt de considérer les deux processus. Par exemple, le bus 4 peut passer moins souvent, mais prendre en moyenne moins de temps pour rejoindre l'arrêt Briffaut.

- (iii) lorsqu'un élève attend le bus à l'arrêt pôle bus, le temps d'attente qu'il supporte avant de se décourager et d'aller au bar est distribué selon une exponentielle de paramètre λ_{26} . Lorsqu'un élève, de retour de l'Esisar, descend du bus 3, il y a une probabilité p ($0 < p < 1$) qu'il rejoigne la résidence pour aller manger, réviser et se reposer, et une probabilité $1 - p$ qu'il aille au bar.

Avec les hypothèses précédentes, les probabilités de transition, d'un état vers un autre sur un intervalle de temps donné, ne dépendent que de l'état de départ et de l'état d'arrivée, ainsi que de la longueur de cet intervalle de temps. Elles ne varient pas au cours du temps. Les hypothèses pour modéliser cette situation par une CMTC homogène sont donc réunies. Examinons, pour nous en convaincre, le calcul de ces probabilités de transition. On a :

$$\begin{aligned} p_{1,2}(dt) &:= P(X(t+dt) = 2 | X(t) = 1) \\ &= P(T_1 < dt) \\ &= 1 - e^{-\lambda_1 dt} \end{aligned} \quad (4.8)$$

où $T_1 \sim \text{Exp}(\lambda_1)$ désigne la variable aléatoire *temps de séjour* dans la résidence. Lorsqu'un élève est dans l'état 8, il y a deux possibilités de sorties. On a par exemple

$$\begin{aligned} p_{8,1}(dt) &:= P(X(t+dt) = 1 | X(t) = 8) \\ &= P(T_8 < dt; \text{l'élève choisit la résidence}) \\ &= (1 - e^{-\lambda_8 dt}) \cdot p \end{aligned} \quad (4.9)$$

De manière complémentaire, si l'élève choisit à la descente du bus d'aller au bar, on a :

$$\begin{aligned} p_{8,6}(dt) &:= P(X(t+dt) = 6 | X(t) = 8) \\ &= P(T_8 < dt; \text{l'élève choisit le bar}) \\ &= (1 - e^{-\lambda_8 dt}) \cdot (1 - p) \end{aligned} \quad (4.10)$$

Lorsqu'un élève attend au pôle bus (état 2), il y a trois processus simultanés en compétition : les arrivées de bus n°3, les arrivées de bus n°4 et les décisions de ne plus attendre et de rejoindre le bar. Il est facile de caractériser $T_2 := \min(T_{23}, T_{24}, T_{26})$ où T_{23} , T_{24} et T_{26} désignent respectivement les intervalles de temps entre deux arrivées de bus n°3, de bus n°4 ou la durée d'attente d'un étudiant avant qu'il ne renonce. Il s'agit de trois variables aléatoires exponentielles et :

$$\begin{aligned} P(T_2 \leq dt) &:= P(\min(T_{23}, T_{24}, T_{26}) \leq dt) \\ &= 1 - P(\min(T_{23}, T_{24}, T_{26}) > dt) \\ &= 1 - P(T_{23} > dt) P(T_{24} > dt) P(T_{26} > dt) \\ &= 1 - e^{-\lambda_{23} dt} e^{-\lambda_{24} dt} e^{-\lambda_{26} dt} \\ &= 1 - e^{-(\lambda_{23} + \lambda_{24} + \lambda_{26}) dt} \end{aligned} \quad (4.11)$$

Il s'agit donc d'une variable exponentielle $T_2 \sim \text{Exp}(\lambda_2)$ avec $\lambda_2 := \lambda_{23} + \lambda_{24} + \lambda_{26}$. Lorsqu'on quitte l'état 2, on rejoindra les états 3, 4 ou 6 avec différentes probabilités. On a par exemple :

$$\begin{aligned}
 p_{2,6}(dt) &:= P(X(t+dt) = 6 | X(t) = 2) \\
 &= P(T_2 \leq dt; T_{26} \leq \min(T_{23}, T_{24})) \\
 &= P(T_2 \leq dt | T_{26} \leq \min(T_{23}, T_{24})) \cdot P(T_{26} \leq \min(T_{23}, T_{24})) \\
 &= P(T_{26} \leq dt) \cdot P(T_{26} \leq \min(T_{23}, T_{24})) \\
 &= \left(1 - e^{-\lambda_{26} dt}\right) \cdot P(T_{26} \leq \min(T_{23}, T_{24}))
 \end{aligned}$$

La variable $T := \min(T_{23}, T_{24})$ est une variable aléatoire exponentielle de paramètre $\lambda := \lambda_{23} + \lambda_{24}$. On a donc :

$$\begin{aligned}
 P(T_{26} \leq \min(T_{23}, T_{24})) &= P(T_{26} \leq T) \\
 &= \int_0^\infty \int_0^y \lambda_{26} e^{-\lambda_{26} x} dx \lambda e^{-\lambda y} dy \\
 &= \frac{\lambda_{26}}{\lambda_{26} + \lambda} \\
 &= \frac{\lambda_{26}}{\lambda_{26} + \lambda_{23} + \lambda_{24}}
 \end{aligned}$$

On obtient donc finalement :

$$p_{2,6}(dt) = \frac{\lambda_{26} (1 - e^{-\lambda_{26} dt})}{\lambda_{26} + \lambda_{23} + \lambda_{24}} \quad (4.12)$$

et des résultats similaires pour $p_{2,3}(dt)$ et $p_{2,4}(dt)$.

Il est donc possible de calculer toutes les probabilités de transition entre états pendant un intervalle de temps dt . L'ensemble de ces probabilités caractérisent complètement la CMTC qui modélise le comportement de l'élève. A l'aide d'un tel modèle markovien en temps continu, nous pourrions par exemple répondre à des questions telles que :

- Q1 partant d'un état donné au temps 0 (par exemple de la résidence), quelle est la probabilité d'être dans un état quelconque (par exemple à l'Esisar) au temps $t > 0$ (on parlera de *probabilité d'état* au temps t) ?
- Q2 quelle proportion du temps passe en moyenne un élève dans les transports, au bar, etc. ?
- Q3 combien de fois en moyenne renonce-t-il à aller à l'Esisar pour aller au bar, alors qu'il est à l'arrêt *pôle bus* ?

Remarque 10 (rebouclage direct) Que ce soit dans l'exemple des processus de Poisson ou dans celui du modèle de l'élève, nous avons considéré des exemples de CMTC sans rebouclage direct. La transition depuis un état

$i \in E$, lorsqu'elle a lieu, se fait nécessairement vers un état autre que i . Les modifications à apporter à la théorie pour inclure le cas des CMTC avec rebouclage direct seront examinées plus loin.

Dans l'exemple du modèle de l'élève, nous avons décrit la transition, entre deux états i et j , comme un *temps de séjour exponentiel* dans l'état i , suivi d'une *transition probabiliste* vers l'état j , identique à celle qui survient dans une CMTD. Il est naturel de mettre en oeuvre cette représentation, par exemple dans le cas des transitions entre l'état 8 et les états 1 ou 6, quand l'élève en sortant du bus choisit d'aller à la résidence ou au bar (respectivement avec une probabilité p et $1-p$). Dans le cas où l'élève attend au pôle bus, la situation est plus compliquée. Il y a trois processus aléatoires *en parallèle* et la transition effective sera déterminée par la *première arrivée* dans l'un de ces trois processus. Nous avons pu malgré tout calculer les *probabilités de transition* correspondantes (équations (4.11) et (4.12)) mais ce calcul s'est avéré plus compliqué. Nous verrons dans la section suivante qu'il existe deux représentations équivalentes pour les CMTC, qui font appel soit au *probabilités de transition*, soit aux *taux de transition*. Elles permettent, l'une ou l'autre, de modéliser différents types de processus stochastiques de manière (plus ou moins) naturelle.

4.2 Caractérisations des CMTC

Partant de la définition, nous allons caractériser le comportement dynamique d'une CMTC :

- soit comme une succession de temps séjours de distributions exponentielles dans les états visités, suivis de transitions vers d'autres états caractérisées par une distribution discrète des probabilités de transition (section 4.2.1),
- soit comme une succession, à chaque état traversé par la chaîne, de compétitions entre plusieurs processus d'arrivées concurrents (simultanés, en temps continu), la première arrivée dans l'un des processus en compétition déterminant la transition effective de la chaîne vers le nouvel état (section 4.2.2).

Nous établirons l'équivalence entre ces deux représentations, ainsi que les formules permettant de passer de l'une à l'autre. Nous illustrerons les résultats sur le modèle de l'élève (section 4.2.3). Enfin, nous terminerons cette section par l'analyse des rebouclages directs et de leurs effets sur ces deux représentations des CMTC (section 4.2.4).

4.2.1 Temps de séjours et probabilités de transition

Considérons une CMTC homogène $(X(t))_{t \geq 0}$ à valeurs dans E , au sens de la définition 17. Le temps de séjour dans un état $i \in E$ est une variable

aléatoire continue que nous noterons T_i . Sa fonction de répartition s'écrit :

$$\begin{aligned}
 F_{T_i}(dt) &:= P(T_i \leq dt), \forall dt > 0 \\
 &= P(T_i \leq dt | T_i \geq 0) \\
 &= P(T_i \leq t + dt | T_i \geq t), \forall t, dt > 0 \\
 &= \frac{P(t \leq T_i \leq t + dt)}{P(T_i \geq t)} \\
 &= \frac{F_{T_i}(t + dt) - F_{T_i}(t)}{1 - F_{T_i}(t)}
 \end{aligned} \tag{4.13}$$

d'une part, et $F_{T_i}(\xi) = 0, \forall \xi < 0$, d'autre part. En divisant les deux membres de l'équation (4.13) par $dt > 0$, on obtient

$$(1 - F_{T_i}(t)) \frac{F_{T_i}(dt)}{dt} = \frac{F_{T_i}(t + dt) - F_{T_i}(t)}{dt}$$

Par passage à la limite $dt \rightarrow 0^+$, on obtient

$$(1 - F_{T_i}(t)) F'_{T_i}(0) = F'_{T_i}(t) \tag{4.14}$$

où

$$F'_{T_i}(t) = \frac{d}{dt}(F_{T_i}(t)) := \lim_{dt \rightarrow 0} \frac{F_{T_i}(t + dt) - F_{T_i}(t)}{dt}$$

L'équation différentielle (4.14), avec la condition initiale $F_{T_i}(0) = P(T_i \leq 0) = 0$, possède l'unique solution

$$F_{T_i}(t) = \left(1 - e^{-F'_{T_i}(0)t}\right)$$

sur le domaine $t \geq 0$. Nous avons donc obtenu

$$F_{T_i}(t) = \begin{cases} 1 - e^{-F'_{T_i}(0)t} & \text{si } t \geq 0 \\ 0 & \text{si } t < 0 \end{cases} \tag{4.15}$$

qui n'est rien d'autre que la fonction de répartition d'une variable aléatoire exponentielle $\text{Exp}(\mu_i)$ de paramètre $\mu_i := F'_{T_i}(0)$. Ce résultat est résumé dans la proposition suivante.

Proposition 15 (Temps de séjour) Soit $(X(t))_{t \geq 0}$, une chaîne de Markov à temps continu homogène à valeurs dans E . Alors, le temps de séjour de la chaîne dans un état $i \in E$ est une variable aléatoire exponentielle.

Nous noterons dans la suite $T_i \sim \text{Exp}(\mu_i)$ le temps de séjour de la chaîne dans l'état i lorsqu'elle s'y trouve. Le paramètre $\mu_i > 0$ doit bien sûr être calculé en fonction des données du problème considéré (voir la section 4.2.3 pour des exemples à ce sujet). Après être resté un temps T_i dans un état $i \in E$, il y a transition de la chaîne vers un autre état $j \in E$ avec $j \neq i$

(le cas du rebouclage direct où $i = j$ sera examiné à la section 4.2.4). La distribution discrète de probabilités d'une transition vers le nouvel état j , pour les différentes valeurs possible de $j \in E$, ne dépend que de l'état i à l'instant présent (au temps t) et pas du chemin parcouru pour y arriver (i.e. de la trajectoire passée) par l'hypothèse markovienne. Ces *probabilités de transition sachant que l'on quitte l'état i* sont donc associées à une chaîne de Markov en temps discret, incluse (ou associée) à la CMTC d'origine, et décrite par le même graphe. Nous venons ainsi d'établir la caractérisation des CMTC qui suit.

Proposition 16 (Caractérisation I) Soit $(X(t))_{t \geq 0}$ une chaîne à temps continu homogène (stationnaire) à valeurs dans un espace d'état discret E . C'est une chaîne de Markov à temps continu (CMTC) si et seulement si :

- (i) les temps de séjour T_i dans les états $i \in E$ ont tous une distribution exponentielle, c'est-à-dire

$$\exists \mu_i > 0 : T_i \sim \text{Exp}(\mu_i), \forall i \in E$$

- (ii) les transitions entres états, lorsqu'elles ont lieu, sont probabilistes. On notera $p_{i,j}$ (avec $0 < p_{i,j} < 1$ et $\sum_{j \neq i} p_{i,j} = 1$), la probabilité conditionnelle d'aller de i à j , sachant que l'on quitte i .

Cette caractérisation est illustrée schématiquement sur le graphe de la figure 4.3. Les grandeurs μ_i et $p_{i,j}$ de la proposition 16 (caractérisation I) sont liées

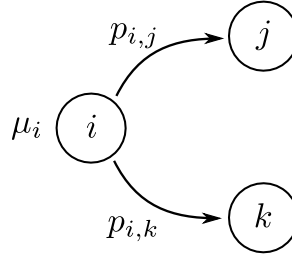


FIGURE 4.3 – Graphe associé à une CMTC représentant sa caractérisation en termes de temps de séjour $T_i \sim \text{Exp}(\mu_i)$ et de probabilités discrètes de transition $p_{i,j}$ (cas sans rebouclages directs, i.e. avec $p_{i,i} = 0, \forall i \in E$). Seules les transition de i vers j et de i vers k sont représentées ici.

aux probabilités de transition $p_{i,j}(dt)$ de la définition 18 (CMTC homogène). On a en effet

$$\begin{aligned} p_{i,j}(dt) &= P(X(t+dt) = j | X(t) = i) \\ &= P(T_i < t+dt | T_i \geq t) \cdot p_{i,j} \\ &= (1 - e^{-\mu_i dt}) \cdot p_{i,j} \end{aligned} \tag{4.16}$$

4.2.2 Taux de transition

Une deuxième caractérisation possible des CMTC repose sur les taux de transition. Ceux-ci peuvent être interprétés comme des taux de variation par unité de temps (ou flux) des probabilités $p_{i,j}(t)$ et sont définis de la manière suivante.

Définition 19 (taux de transition) Soit $(X(t))_{t \geq 0}$, une chaîne de Markov à temps continu homogène à valeurs dans E et

$$p_{i,j}(dt) = P(X(dt) = j | X(0) = i)$$

les probabilités de transition associées (voir définition 18). Alors, le *taux de transition* de l'état i vers l'état j ($i \neq j$) est défini par :

$$\begin{aligned} \lambda_{i,j} &:= p'_{i,j}(0) \\ &= \lim_{dt \rightarrow 0} \frac{p_{i,j}(dt) - p_{i,j}(0)}{dt} \\ &= \lim_{dt \rightarrow 0} \frac{p_{i,j}(dt)}{dt} \text{ pour } i \neq j \end{aligned} \quad (4.17)$$

On définit de plus

$$\begin{aligned} \lambda_{i,i} &:= p'_{i,i}(0) \\ &= \lim_{dt \rightarrow 0} \frac{p_{i,i}(dt) - p_{i,i}(0)}{dt} \\ &= \lim_{dt \rightarrow 0} \frac{p_{i,i}(dt) - 1}{dt} \text{ pour } i = j \\ &= \lim_{dt \rightarrow 0} \frac{-\sum_{j \neq i} p_{i,j}(dt)}{dt} \\ &= -\sum_{j \neq i} p'_{i,j}(0) \\ &= -\sum_{j \neq i} \lambda_{i,j} \end{aligned} \quad (4.18)$$

A partir de cette définition des taux de transition, nous pouvons écrire³ :

$$\begin{aligned} p_{i,j}(dt) &= P(X(t+dt) = j | X(t) = i) \\ &= \lambda_{i,j}dt + o(dt), \text{ pour } dt \rightarrow 0 \end{aligned} \quad (4.19)$$

Le développement limité de $p_{i,j}(dt)$ au voisinage de $dt = 0$ coïncide donc avec celui de l'exponentielle jusqu'au premier ordre :

$$p_{i,j}(dt) = \left(1 - e^{-\lambda_{i,j}dt}\right) + o(dt), \text{ pour } dt \rightarrow 0$$

3. $f(t) := o(dt)$ représente une fonction *négligeable* devant dt au voisinage de 0, c'est-à-dire telle que $\lim_{dt \rightarrow 0} \frac{f(t)}{dt} = 0$.

On peut donc interpréter la transition de i vers j comme un processus continu dans lequel le *temps de transition* est distribué selon une variable aléatoire exponentielle de paramètre $\lambda_{i,j}$. Nous venons donc d'établir le caractère *plausible*⁴ de la caractérisation qui suit pour les CMTC, en termes de taux de transition.

Proposition 17 (Caractérisation II) Soit $(X(t))_{t \geq 0}$ une chaîne de Markov à temps continu homogène à valeurs dans l'espace d'état discret E . Alors, partant d'un état $i \in E$, le *temps de transition* vers j (ou respectivement et simultanément vers un autre état k) suit une loi exponentielle de paramètre $\lambda_{i,j}$ (respectivement $\lambda_{i,k}$).

Afin de déterminer vers quel état aura lieu la prochaine transition de la chaîne, partant de l'état i , il faut déterminer, parmi l'ensemble des temps de transition vers les autres états atteignables depuis i (e.g. j et k dans la proposition ci-dessus), lequel aura la réalisation la plus courte. L'état correspondant sera le prochain état visité par la chaîne. Nous pouvons donc interpréter la CMTC comme un ensemble de processus de Poisson concourants (simultanés), défini en chacun de ces états, les transitions effectives d'un état à un autre étant celles qui correspondent à la première arrivée dans les processus concourants.

La caractérisation des CMTC en termes de *taux de transition* $\lambda_{i,j}$, plutôt qu'en termes de *temps de séjour* μ_i et *probabilités de transition* $p_{i,j}$, est illustrée schématiquement sur le graphe de la figure 4.4.

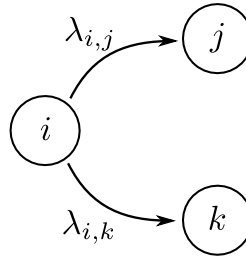


FIGURE 4.4 – Graphe associé à une CMTC représentant sa caractérisation en termes de *taux de transition* $\lambda_{i,j}$. Seules les transitions (arrêtes) qui correspondent à des taux de transition non nuls sont représentées.

Exemple 10 (taux de transition dans un PP) Afin de renforcer notre intuition de cette interprétation des CMTC en termes de taux de transition, examinons le cas simple d'un processus de Poisson de paramètre λ , où l'état

4. La démonstration du caractère exponentiel de la distribution des *temps de transition* de i vers j repose sur (4.19), d'une part, et sur un raisonnement analogue à la démonstration de la proposition 15, d'autre part.

$X(t) = N(t)$ représente le nombre d'arrivées au temps $t > 0$. Dans cet exemple, l'espace d'état est $E = \mathbb{N}$. Comme il s'agit d'un processus de comptage, les transitions ne sont possible qu'entre un état $i \in \mathbb{N}$ et les états $j \in \mathbb{N}$ avec $j \geq i$. On a donc $\lambda_{i,j} = 0$ pour tout $j < i$. Nous avons par ailleurs déjà établi les probabilités de transitions $p_{i,j}(dt)$ pour un processus de Poisson (voir équation (4.7)). On obtient à l'aide de ce résultat et de la définition des taux de transition :

$$\begin{aligned} \lambda_{i,j} &= \lim_{dt \rightarrow 0} \frac{p_{i,j}(dt)}{dt} \\ &= \lim_{dt \rightarrow 0} \frac{(\lambda dt)^{j-i}}{(j-i)! dt} e^{-\lambda dt}, \quad \text{si } j > i \\ &= \begin{cases} \lambda & \text{si } j = i + 1 \\ 0 & \text{si } j > i + 1 \end{cases} \end{aligned} \quad (4.20)$$

ainsi que

$$\lambda_{i,i} = - \sum_{j \neq i} \lambda_{i,j} = -\lambda \quad (4.21)$$

$$(4.22)$$

Dans un processus de Poisson, le taux de transition d'un état i vers le suivant ($i+1$) vaut donc simplement λ . Tous les autres sont nuls. Nous avons en effet déjà noté que les probabilités qu'il y ait deux arrivées pendant un intervalle de temps dt sont négligeables (devant les probabilités qu'il n'y en ai pas ou qu'il n'y en ai qu'une). On peut représenter un processus de Poisson de paramètre λ à l'aide des taux de transition comme illustré à la figure 4.5.

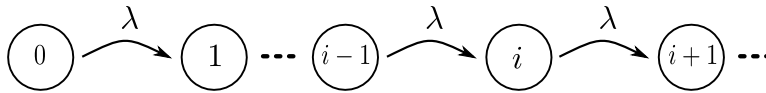


FIGURE 4.5 – Graphe associé à un processus de Poisson de paramètre λ . On représente les taux de transition (deuxième caractérisation).

Dans certaines situations, il sera plus aisé d'établir le modèle d'une CMTC (ou d'une partie de celle-ci) en utilisant la première caractérisation (temps de séjours et probabilités de transition), alors que dans d'autres situations (ou pour une autre partie de la CMTC), une caractérisation par les taux de transition sera plus facile à mettre en oeuvre. Heureusement, il est facile de passer d'une caractérisation à l'autre. En effet, en comparant les expressions (4.16) et (4.19) qui donnent toutes les deux les probabilités de transition

pendant un temps dt , on obtient :

$$\begin{aligned} p_{i,j}(dt) &= P(X(t+dt) = j | X(t) = i) \\ &= \lambda_{i,j} dt + o(dt) \\ &= (1 - e^{-\mu_i dt}) \cdot p_{i,j} \end{aligned}$$

D'où, par identification des développements limités au premier ordre :

$$\lambda_{i,j} = \mu_i p_{i,j}, \quad \forall j \neq i \quad (4.23)$$

D'autre part,

$$\begin{aligned} \lambda_{i,i} &= - \sum_{j \neq i} \lambda_{i,j} \\ &= - \sum_{j \neq i} \mu_i p_{i,j} \\ &= \mu_i \left(- \sum_{j \neq i} p_{i,j} \right) \\ &= \mu_i (p_{i,i} - 1) \end{aligned}$$

Donc, dans le cas d'une chaîne sans rebouclage direct ($p_{i,i} = 0$), on a :

$$\begin{aligned} \mu_i &= \sum_{j \neq i} \lambda_{i,j} \\ p_{i,j} &= \frac{\lambda_{i,j}}{\mu_i}, \quad \forall j \neq i \end{aligned} \quad (4.24)$$

Les équations (4.23) et (4.24) permettent de passer de la représentation par temps de séjour et probabilités de transition, à la représentation par taux de transition. Ces deux représentations équivalentes sont données à la figure 4.6. Nous utiliserons l'une ou l'autre de ces représentations pour construire le modèle de l'élève dans le paragraphe suivant.

4.2.3 Le modèle de l'élève revisité

Reprenons l'exemple 9 du modèle de l'élève. Certaines transitions sont immédiates à caractériser en terme de temps de séjour. Par exemple, les temps de séjours dans les états 1 (résidence), 3 (temps de transport dans le bus 3, trajet aller), 4 (temps de transport dans le bus 4, aller), 5 (Esisar) et 6 (bar) sont distribués suivant des exponentielles indépendantes les unes des autres, de paramètres respectifs λ_1 , λ_3 , λ_4 , λ_5 et λ_6 . Les temps de séjours correspondants sont suivis par des transitions déterministes (de probabilité 1), respectivement vers les états 2, 5, 5, 7 et 1, car il n'y a à chaque fois qu'une

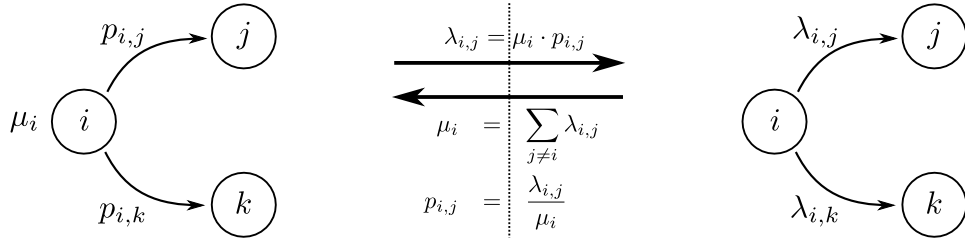


FIGURE 4.6 – Deux représentations équivalentes d’une même CMTC qui correspondent respectivement à sa caractérisation par temps de séjour et probabilités de transition (à gauche) ou à sa caractérisation par les taux de transition (à droite)

seuls destination possible (un seul état atteignable). La formule (4.23) donne alors

$$\begin{aligned}\lambda_{1,2} &= \lambda_1 ; \lambda_{3,5} = \lambda_3 ; \lambda_{4,5} = \lambda_4 \\ \lambda_{5,7} &= \lambda_5 ; \lambda_{6,1} = \lambda_1\end{aligned}$$

Lorsqu’un élève se trouve dans l’état 7 (arrêt Briffaut), il attend l’arrivée d’un bus qui se produit selon un processus de Poisson de paramètre $\lambda_{7,3}$ avant de monter dans ce bus (état 8). On a donc, d’après l’équation (4.20) qui donne les taux de transition pour un processus de Poisson :

$$\lambda_{7,8} = \lambda_{7,3}$$

Lorsqu’un élève se trouve dans l’état 2 (arrêt pôle bus), il y a trois processus concourants simultanés : les arrivées des bus n°3, selon un processus de Poisson de paramètre $\lambda_{2,3}$, les arrivées des bus n°4 selon un processus de Poisson de paramètre $\lambda_{2,4}$ et l’arrivée du découragement de l’étudiant après un temps exponentiel de paramètre $\lambda_{2,6}$, donc de manière équivalente selon un processus de Poisson de même paramètre $\lambda_{2,6}$. Les transitions depuis l’état 2 vers les états 3, 4 ou 6 sont donc données directement en terme de taux de transition. On a :

$$\lambda_{2,3} = \lambda_{2,3} ; \lambda_{2,4} = \lambda_{2,4} ; \lambda_{2,6} = \lambda_{2,6}$$

Enfin, un élève dans l’état 8 (transport dans le bus n°3, pour le trajet depuis l’Esisar) y séjournera pendant un temps de transport distribué selon une exponentielle de paramètre λ_8 , avant de se rendre au bar (état 6) avec une probabilité $p_{8,6} = (1 - p)$ ou à la résidence avec une probabilité $p_{8,1} = p$. On a donc pour les taux de transition, en suivant l’équation (4.23) :

$$\lambda_{8,1} = \lambda_8 p ; \lambda_{8,6} = \lambda_8 (1 - p)$$

Nous venons de caractériser complètement le modèle de CMTC de l’étudiant en spécifiant l’ensemble des taux de transitions non nuls. Il est représenté à la figure 4.7.

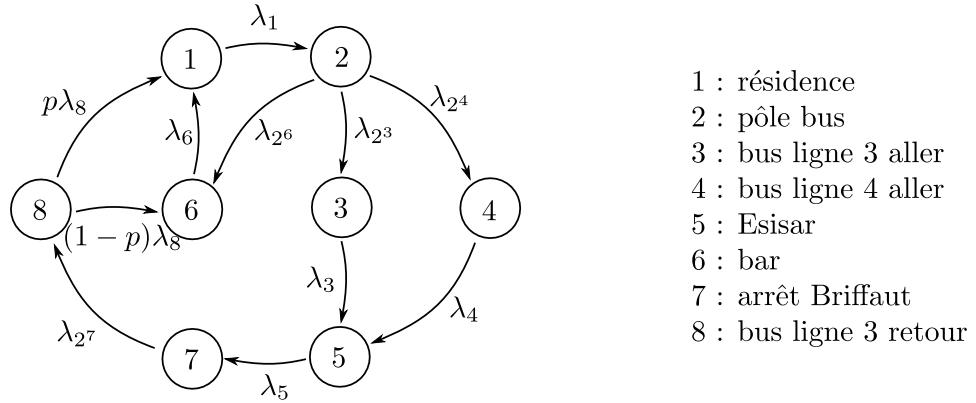


FIGURE 4.7 – Le modèle de l'étudiant de l'exemple 9 caractérisé par l'ensemble des taux de transition non nuls

4.2.4 Le cas des rebouclages directs

Examinons maintenant le cas des CMTC *avec rebouclages directs*. Après un temps de séjour $T_i \sim \text{Exp}(\mu_i)$ dans l'état $i \in E$, la chaîne peut donc aller vers un autre état $j \neq i, j \in E$ avec une probabilité $p_{i,j}$ ou revenir à l'état i avec une probabilité $p_{i,i} > 0$. En utilisant la première caractérisation, on représentera cette situation de manière schématique à l'aide du graphe de la figure 4.8. Il est possible d'interpréter cette situation comme une CMTC sans

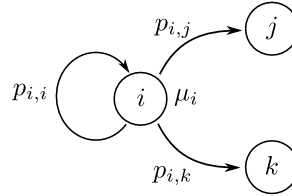


FIGURE 4.8 – CTMC avec rebouclage direct ($p_{i,i} > 0$), première caractérisation

rebouclage, mais avec des probabilités modifiées $\tilde{p}_{i,j}$ et des temps de séjour modifiés \tilde{T}_i . Il suffit à cet effet de remarquer que si après un temps de séjour T_i on revient à l'état i , alors cela rallongera d'autant le temps de séjour dans l'état i , le "deuxième séjour" consécutif étant simplement considéré comme une prolongation du premier. Ce temps de séjour équivalent sans rebouclage, \tilde{T}_i , satisfait l'équation

$$F_{\tilde{T}_i}(t) := P(\tilde{T}_i \leq t) = P(T_i \leq t)(1 - p_{i,i}) + p_{i,i}P(T_i + \tilde{T}_i \leq t), \forall t > 0$$

On donc :

$$f_{\tilde{T}_i}(t) = (1 - p_{i,i})f_{T_i}(t) + p_{i,i} \left(f_{T_i} * f_{\tilde{T}_i} \right)(t) \quad (4.25)$$

En notant $\hat{f}_{T_i}(s)$ la transformée de Laplace de la densité de probabilité $f_{T_i}(s)$, on a d'une part :

$$\hat{f}_{T_i}(s) = \frac{\mu_i}{\mu_i + s}$$

et d'autre part, en prenant la transformée de Laplace de (4.25)

$$\hat{f}_{\tilde{T}_i}(s) = (1 - p_{i,i})\hat{f}_{T_i}(s) + p_{i,i}\hat{f}_{T_i}(s)\hat{f}_{\tilde{T}_i}(s)$$

On obtient ainsi

$$\hat{f}_{\tilde{T}_i}(s) = \frac{(1 - p_{i,i})\mu_i}{(1 - p_{i,i})\mu_i + s}$$

qui montre que le temps de séjour modifié (avec rebouclage) est lui aussi une variable exponentielle

$$\tilde{T}_i \sim \text{Exp}(\tilde{\mu}_i)$$

avec

$$\tilde{\mu}_i := (1 - p_{i,i})\mu_i$$

Dans l'interprétation d'une CMTC avec rebouclage comme une CMTC sans rebouclage, il faut non seulement modifier les temps de séjour, mais également les probabilités de transition. En appelant $\tilde{p}_{i,j}$ les probabilités de transition de la CMTC équivalence sans rebouclage, il vient :

$$\tilde{p}_{i,j} = p_{i,j} + p_{i,i}\tilde{p}_{i,j}$$

c'est-à-dire

$$\tilde{p}_{i,j} = \frac{p_{i,j}}{1 - p_{i,i}}$$

Nous avons donc montré le résultat suivant.

Proposition 18 (CMTC avec rebouclages directs)

Une CMTC homogène sur E avec rebouclages directs, caractérisée par des temps de séjour $T_i \sim \text{Exp}(\mu_i)$ et des probabilités de transition $p_{i,j}$ ($i, j \in E$) est *équivalente* à une CMTC sans rebouclage, avec des temps de séjour $\tilde{T}_i \sim \text{Exp}(\tilde{\mu}_i)$ et des probabilités de transition $\tilde{p}_{i,j}$ ($i, j \in E$) avec :

$$\tilde{\mu}_i = (1 - p_{i,i})\mu_i \quad (4.26)$$

$$\tilde{p}_{i,j} = \frac{p_{i,j}}{1 - p_{i,i}} = \frac{p_{i,j}}{\sum_{k \neq i} p_{i,k}} \quad (4.27)$$

A l'aide de ce résultat, nous pouvons examiner l'effet d'un rebouclage direct $p_{i,i}$ sur la deuxième caractérisation des CMTC, c'est-à-dire sur la valeur des

taux de transitions. Les taux de transition qui correspondent à la CMTC avec rebouclage direct s'écrivent :

$$\begin{aligned}\tilde{\lambda}_{i,j} &= \tilde{\mu}_i \tilde{p}_{i,j} \text{ pour } i \neq i \\ &= \mu_i p_{i,j} \\ &= \lambda_{i,j}\end{aligned}$$

Les taux de transition $\lambda_{i,j}$ pour $i \neq j$ ne sont pas affectés par le rebouclage direct $p_{i,i} > 0$. Le taux de transition $\lambda_{i,i}$ défini par

$$\tilde{\lambda}_{i,i} = - \sum_{j \neq i} \tilde{\lambda}_{i,j} = \lambda_{i,i}$$

n'est bien sûr pas non plus modifié par la présence d'un rebouclage direct. Le taux de transition $\lambda_{i,j}$ est un flux *net* de probabilité entre l'état i et l'état j . La distinction des cas avec ou sans rebouclage direct n'a en fait de sens que dans le cas de la première caractérisation. Les représentations équivalentes des CMTC avec rebouclage sont illustrées à la figure 4.9.

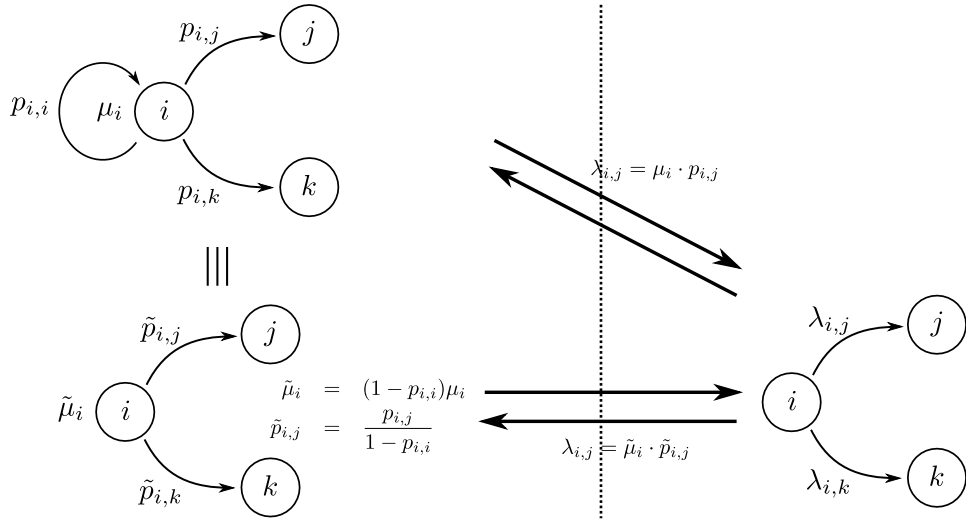


FIGURE 4.9 – Les deux représentations d'une CMTC avec rebouclage direct ($p_{i,i} > 0$). La représentation avec les taux de transition (à droite) n'est pas affectée par la présence d'un rebouclage. Dans la représentation de gauche, l'ensemble des temps de séjour et des probabilités de transition doivent être recalculés dans le cas où un rebouclage direct est ajouté.

Remarque 11 (avantages des taux de transition) La description d'une CMTC à l'aide des taux de transition présente certain avantages :

- (i) Comme on vient de le voir, elle intègre intrinsèquement la possibilité de rebouclages directs : il n'est pas nécessaire de traiter ce cas (avec rebouclage) de manière spécifique.
- (ii) Dans une représentation avec les taux de transition, il est possible de faire évoluer le modèle, par exemple en ajoutant des états ou des transitions, sans avoir à recalculer l'ensemble des taux de transitions. Seuls les taux qui concernent directement l'état ou la transition modifiés doivent être recalculés. Cette représentation est donc beaucoup plus modulaire et évolutive, ce qui constitue un avantage décisif pour traiter des problèmes complexes
- (iii) Comme nous le verrons dans la section suivante, le calcul des régimes transitoire et permanent - ainsi que tous les autres calculs qui en découlent - font appel à la matrice des taux de transition (appelée *générateur infinitésimal* de la CMTC)

4.3 Régimes transitoire et permanent

Les notions de régimes transitoire et permanent (ou stationnaire) que nous avons définies pour les chaînes à temps discret, s'étendent aux chaînes de Markov à temps continu. La matrice de passage \mathbf{P} sera remplacée par une matrice des taux de transition, appelée *générateur infinitésimal*, alors que la récurrence de type point fixe qui détermine l'évolution dans le temps du vecteur des probabilités d'état sera remplacé par une équation différentielle ordinaire.

4.3.1 régime transitoire

Définition 20 (probabilité d'état) Soit $(X(t))_{t \geq 0}$, une chaîne de Markov à temps continu homogène à valeurs dans un espace d'état discret E . On définit le vecteur des probabilités d'état au temps t , noté $\pi(t)$, comme le vecteur ligne dont les composantes sont

$$\pi_j(t) := P(X(t) = j), \forall j \in E$$

Cherchons à caractériser l'évolution au cours du temps de ce vecteur des probabilités totales $\pi(t)$. Par application de la formule des probabilités totales, on obtient, quels que soient $i \in E$ et $h > 0$:

$$\begin{aligned}
 \pi_i(t+h) &:= P(X(t+h) = i) \\
 &= \sum_{k \in E} P(X(t+h) = i | X(t) = k) \cdot P(X(t) = k) \\
 &= \sum_{k \in E} p_{k,i}(h) \cdot \pi_k(t)
 \end{aligned} \tag{4.28}$$

où $p_{k,i}(h)$ sont les probabilités de transition définies en (4.3). En adoptant des notations vectorielles, (4.28) peut être écrite sous la forme :

$$\boldsymbol{\pi}(t+h) = \boldsymbol{\pi}(t)\mathbf{P}(h) \quad (4.29)$$

où $\mathbf{P}(h)$ désigne la matrice des probabilités de transition définie, composante par composante, par :

$$\mathbf{P}_{i,j}(h) := P(X(h) = j | X(0) = i) , \forall i, j \in E \quad (4.30)$$

En dérivant l'équation (4.29) par rapport à h , on obtient :

$$\boldsymbol{\pi}'(t+h) = \boldsymbol{\pi}(t)\mathbf{P}'(h)$$

avec

$$\boldsymbol{\pi}'(\xi) := \frac{d\boldsymbol{\pi}}{d\xi}(\xi) \text{ et } \mathbf{P}'(\eta) := \frac{d\mathbf{P}}{d\eta}(\eta)$$

En prenant la limite pour $h \rightarrow 0$, on obtient :

$$\begin{aligned} \boldsymbol{\pi}'(t) &= \lim_{h \rightarrow 0} \boldsymbol{\pi}'(t+h) \\ &= \boldsymbol{\pi}(t) \lim_{h \rightarrow 0} \mathbf{P}'(h) \\ &= \boldsymbol{\pi}(t)\mathbf{P}'(0) \end{aligned}$$

La matrice $\mathbf{P}'(0)$ qui apparait dans cette dernière expression n'est autre que la matrice des taux de transition de la définition 19, donnés par les équations (4.17) et (4.18). Nous venons donc de montrer la proposition qui suit.

Proposition 19 (régime transitoire) Soit $(X(t))_{t \geq 0}$, une chaîne de Markov à temps continu homogène à valeurs dans un espace d'état discret E . Soit \mathbf{A} le *générateur infinitésimal* de cette CMTC, défini comme la matrice des taux de transition :

$$\mathbf{A}_{i,j} := \lambda_{i,j} , \forall i, j \in E \quad (4.31)$$

On a alors :

$$\dot{\boldsymbol{\pi}}(t) = \boldsymbol{\pi}(t)\mathbf{A} \quad (4.32)$$

où

$$\dot{\boldsymbol{\pi}}(t) := \frac{d\boldsymbol{\pi}}{dt}(t)$$

L'équation différentielle linéaire (4.32) est appelée *équation de Kolmogorov*. Comme dans le cas discret, elle doit être complétée par la donnée de la distribution initiale de probabilité $\boldsymbol{\pi}(0)$ définie, composante par composante, par

$$\pi_i(0) = (\boldsymbol{\pi}_0)_i = P(X(0) = i) , \forall i \in E$$

pour pouvoir être intégrée et fournir une solution unique qui aura la forme

$$\boldsymbol{\pi}(t) = \boldsymbol{\pi}(0)e^{\mathbf{A}t}$$

avec

$$e^{\mathbf{A}t} := \sum_{n \geq 0} \frac{(\mathbf{A}t)^n}{n!}$$

La distribution initiale de probabilité traduit notre connaissance, à priori, de l'état initial du système. En l'absence de toute connaissance à priori, on choisira une distribution uniforme. Dans le cas où l'état initial est connu et vaut par exemple $i_0 \in E$, on choisira une distribution initiale

$$\pi_i(0) = \delta_{i,i_0}, \forall i \in E$$

En général l'équation différentielle de Kolmogorov ne peut être résolue que pour des CMTC de taille réduite, de structure (topologie) très particulière ou seulement numériquement. Cependant, l'évaluation des performances d'un processus stochastique ne nécessite en général pas la connaissance de la solution (probabilités d'état au temps t), mais seulement de son comportement sur le long terme (asymptotique), c'est-à-dire la connaissance du régime stationnaire, aussi appelé régime permanent.

4.3.2 régime permanent

Nous nous intéressons maintenant, comme dans le cas discret, au régime stationnaire - ou encore permanent - de la CMTC. Il s'agit par définition d'une distribution de probabilités d'état $\boldsymbol{\pi}$, à la fois indépendante du temps t et indépendante de la distribution initiale $\boldsymbol{\pi}(0)$. Elle doit donc en particulier être une distribution d'équilibre pour l'équation différentielle de Kolmogorov (4.32), c'est-à-dire une solution de

$$\frac{d\boldsymbol{\pi}}{dt} = 0 = \boldsymbol{\pi}\mathbf{A} \quad (4.33)$$

et satisfaire les propriétés d'une distribution de probabilités :

$$0 \leq \pi_i \leq 1, \forall i \in E$$

$$\sum_{i \in E} \pi_i = 1 \quad (4.34)$$

Comme dans le cas discret, lorsque la limite

$$\boldsymbol{\pi}^\infty := \lim_{t \rightarrow \infty} \boldsymbol{\pi}(t) \quad (4.35)$$

existe, elle satisfait nécessairement les conditions (4.33) et (4.34). La distribution limite $\boldsymbol{\pi}^\infty$, lorsqu'elle existe, est donc nécessairement une distribution

d'équilibre stable pour l'équation différentielle de Kolmogorov, c'est-à-dire aussi une distribution stationnaire permanente.

Dans le cas des chaînes de Markov en temps continu, il n'est pas possible d'observer dans une réalisation particulière un comportement périodique. En effet, les transitions pouvant survenir à tout moment et les intervalles entre deux transitions étant des variables aléatoires continues, il n'y a aucune raison d'observer une régularité temporelle dans ces transitions. Lorsque la distribution stationnaire existe, on observera sur le long terme un comportement sur le long terme en apparence chaotique. Le chaîne sautera d'un état à l'autre, passant en moyenne (sur un horizon de temps infini, pas à chaque séjour) dans chacun des états $i \in E$ visités un temps proportionnel à π_i . La chaîne est alors dite *ergodique*.

Par contre, la distinction entre états récurrents et états transitoires garde tout son sens. Un état est en effet *récurrent* si partant de cet état, la chaîne y revient presque sûrement (i.e. avec une probabilité 1). Si nous considérons la première caractérisation des CMTC, nous pouvons voir que cette propriété est *indépendante des temps de séjour* dans chacun des états, mais ne dépend que de la topologie du graphe associé et des probabilités discrètes de transitions entre les états. Ainsi, un état sera récurrent ou transitoire selon qu'il est récurrent ou transitoire pour la *chaîne de Markov discrète associée* (on dit aussi *incluse*) à la CMTC. La distinction entre états *récurrents nuls* et *récurrents non nuls* garde aussi son sens (dans le cas bien sûr où la chaîne de Markov est définie sur un espace d'état est infini). En effet, les temps de séjour étant distribués selon des variables exponentielles d'espérance *finie*, la chaîne reviendra sur un état récurrent en temps fini si elle y revient en un nombre fini de transitions, c'est-à-dire si cet état est récurrent non nul pour la CMTD incluse.

De même, les notions de *chemin*, d'*état atteignable* depuis un autre, d'états qui *communiquent* et de *classes d'équivalence* et de *chaînes irréductibles* se déduisent directement de ces mêmes propriétés sur la CMTD incluse, c'est-à-dire, dans ce cas, de la topologie de son graphe (qui est bien sûr identique au graphe de la CMTC).

L'ensemble des propriétés des états d'une CMTC se déterminent donc à partir du graphe associé et des probabilités de transition $p_{i,j}$ de la CMTD incluse. C'est pourquoi les propriétés ci-dessous, valables pour les chaînes en temps discret, restent valable pour les chaînes de Markov en temps continu :

- (i) Dans une CMTC irréductible (ou dans une classe d'équivalence), tous les états sont de même nature : ils sont tous transitoires, tous récurrents nuls ou récurrents non nuls en même temps
- (ii) Tous les états d'une CMTC irréductible finie sont récurrents non nuls

Par un raisonnement similaire à celui développé pour les CMTD, on peut montrer la proposition suivante.

Proposition 20 (Distributions limite et stationnaire)

Soit $(X(t))_{t \geq 0}$, une chaîne de Markov à temps continu homogène à valeurs dans un espace d'état discret E . Si cette chaîne est irréductible, alors la limite

$$\boldsymbol{\pi}^\infty := \lim_{t \rightarrow \infty} \boldsymbol{\pi}(t) \quad (4.36)$$

existe nécessairement. Elle est indépendante de la distribution initiale $\boldsymbol{\pi}(0)$. De plus :

- (i) si tous les états sont transitoires ou tous récurrents nuls (E doit être infini), alors

$$\pi_j^\infty = 0, \forall j \in E$$

- (ii) si tous les états sont récurrents non nuls (c'est en particulier le cas si E est fini), alors

$$\boldsymbol{\pi}^\infty = \boldsymbol{\pi}$$

où $\boldsymbol{\pi}$ est la distribution stationnaire solution de

$$\boldsymbol{\pi} \mathbf{A} = 0 \quad (4.37)$$

avec la condition de normalisation

$$\|\boldsymbol{\pi}\|_1 := \sum_{i \in E} |\pi_i| = \sum_{i \in E} \pi_i = 1 \quad (4.38)$$

Le système d'équations (4.39) satisfaites par la distribution stationnaire de probabilités $\boldsymbol{\pi}$ peut être développé composante par composante. On obtient :

$$\begin{aligned} (\boldsymbol{\pi} \mathbf{A})_j &= \sum_{i \in E} \pi_i \mathbf{A}_{i,j} \\ &= \sum_{i \neq j} \pi_i \mathbf{A}_{i,j} + \pi_j \mathbf{A}_{j,j} \\ &= \sum_{i \neq j} \pi_i \lambda_{i,j} + \pi_j \lambda_{j,j} \\ &= \sum_{i \neq j} \pi_i \lambda_{i,j} - \pi_j \sum_{i \neq j} \lambda_{j,i} \\ &= \sum_{i \neq j} \pi_i \lambda_{i,j} - \sum_{i \neq j} \pi_j \lambda_{j,i} \\ &= 0 \end{aligned} \quad (4.39)$$

On peut donc conclure :

$$\underbrace{\sum_{i \neq j} \pi_i \lambda_{i,j}}_{\text{flux entrant dans } j} = \underbrace{\sum_{i \neq j} \pi_j \lambda_{j,i}}_{\text{flux sortant de } j}, \quad \forall j \in E \quad (4.40)$$

Le système d'équations (4.39) qui exprime l'équilibre global de la CMTC est donc équivalent aux équations (4.40) qui traduisent l'équilibre des flux entrant et sortant dans chacun des états $j \in E$.

Nous allons maintenant appliquer ce que nous venons d'apprendre au sujet des distributions stationnaires sur le modèle de l'élève de l'exemple 9.

Exemple 11 (encore le modèle de l'élève) Nous avons calculé dans la section 4.2.3 l'ensemble des taux de transition qui correspondent au modèle de l'élève décrit à la figure 4.2. Ces taux de transition peuvent être regroupés dans le générateur infinitésimal de la CMTC qui s'écrit ici :

$$\mathbf{A} = \begin{pmatrix} -\lambda_1 & \lambda_1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -\lambda_{23} - \lambda_{24} - \lambda_{26} & \lambda_{23} & \lambda_{24} & 0 & \lambda_{26} & 0 & 0 \\ 0 & 0 & -\lambda_3 & 0 & \lambda_3 & 0 & 0 & 0 \\ 0 & 0 & 0 & -\lambda_4 & \lambda_4 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -\lambda_5 & 0 & \lambda_5 & 0 \\ \lambda_6 & 0 & 0 & 0 & 0 & -\lambda_6 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -\lambda_{27} & \lambda_{27} \\ p\lambda_8 & 0 & 0 & 0 & 0 & (1-p)\lambda_8 & 0 & -\lambda_8 \end{pmatrix}$$

Nous pouvons à l'aide de ce modèle répondre aux questions que nous avons précédemment formulées sur ce problème et qui concerne le régime transitoire.

Par exemple, si nous savons que l'élève est au temps $t = 0$ dans la résidence (i.e. $X(0) = 1$), nous choisissons comme distribution initiale de probabilités :

$$\boldsymbol{\pi}(0) = (1 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0)$$

Le vecteur $\boldsymbol{\pi}(t) = \boldsymbol{\pi}(0)e^{\mathbf{A}t}$ des probabilités d'état au temps t sera donc donné par la première ligne de la matrice $e^{\mathbf{A}t} := \sum_{n \geq 0} \frac{(\mathbf{A}t)^n}{n!}$ que nous pouvons évaluer numériquement pour tout choix particulier des valeurs numériques des taux de transition. L'expression symbolique de la solution en fonction de ces paramètres est par contre impossible à obtenir dans le cas général.

Nous pouvons aussi, à l'aide du générateur \mathbf{A} , répondre à toutes les questions qui concernent le comportement de la CMTC sur le long terme, c'est-à-dire en dernier ressort la distribution stationnaire de probabilités. La CMTC du modèle de l'élève étant finie et irréductible (i.e. tous les états communiquent), tous les états sont donc récurrents non nuls. La distribution stationnaire existe, est unique et égale à la distribution limite. La CMTC exhibera un comportement ergodique sur le long terme. En écrivant une à une les huit équations de la relation d'équilibre global $\boldsymbol{\pi}\mathbf{A} = 0$, on obtient

pour la distribution stationnaire de probabilités :

$$\begin{aligned}
-\lambda_1 \pi_1 + \lambda_6 \pi_6 + p \lambda_8 \pi_8 &= 0 \\
\lambda_1 \pi_1 - (\lambda_{2^3} + \lambda_{2^4} + \lambda_{2^6}) \pi_2 &= 0 \\
\lambda_{2^3} \pi_2 - \lambda_3 \pi_3 &= 0 \\
\lambda_{2^4} \pi_2 - \lambda_4 \pi_4 &= 0 \\
\lambda_3 \pi_3 + \lambda_4 \pi_4 - \lambda_5 \pi_5 &= 0 \\
\lambda_{2^6} \pi_2 - \lambda_6 \pi_6 + (1-p) \lambda_8 \pi_8 &= 0 \\
\lambda_5 \pi_5 - \lambda_{2^7} \pi_7 &= 0 \\
\lambda_{2^7} \pi_7 - \lambda_8 \pi_8 &= 0
\end{aligned}$$

En exprimant cette fois l'équilibre des flux entrant et sortant en chaque noeud du graphe, on obtient :

$$\begin{aligned}
\lambda_6 \pi_6 + p \lambda_8 \pi_8 &= \lambda_1 \pi_1 \\
\lambda_1 \pi_1 &= (\lambda_{2^3} + \lambda_{2^4} + \lambda_{2^6}) \pi_2 \\
\lambda_{2^3} \pi_2 &= \lambda_3 \pi_3 \\
\lambda_{2^4} \pi_2 &= \lambda_4 \pi_4 \\
\lambda_3 \pi_3 + \lambda_4 \pi_4 &= \lambda_5 \pi_5 \\
\lambda_{2^6} \pi_2 + (1-p) \lambda_8 \pi_8 &= \lambda_6 \pi_6 \\
\lambda_5 \pi_5 &= \lambda_{2^7} \pi_7 \\
\lambda_{2^7} \pi_7 &= \lambda_8 \pi_8
\end{aligned}$$

Les huit équations sont bien sûr identiques deux à deux, à une réorganisation des termes près. Pour trouver la distribution stationnaire, elles doivent être complétées par la condition de normalisation :

$$\sum_{i=1}^8 \pi_i = 1$$

On peut, par exemple, exprimer toutes les probabilités stationnaires en fonc-

tion de π_2 . On obtient ainsi :

$$\begin{aligned}
 \pi_3 &= \frac{\lambda_{23}}{\lambda_3} \pi_2 \\
 \pi_4 &= \frac{\lambda_{24}}{\lambda_4} \pi_2 \\
 \pi_5 &= \frac{\lambda_{23} + \lambda_{24}}{\lambda_5} \pi_2 \\
 \pi_7 &= \frac{\lambda_{23} + \lambda_{24}}{\lambda_{27}} \pi_2 \\
 \pi_8 &= \frac{\lambda_{23} + \lambda_{24}}{\lambda_8} \pi_2 \\
 \pi_1 &= \frac{\lambda_{23} + \lambda_{24} + \lambda_{26}}{\lambda_1} \pi_2 \\
 \pi_6 &= \frac{(1-p)(\lambda_{23} + \lambda_{24}) + \lambda_{26}}{\lambda_6} \pi_2
 \end{aligned}$$

La condition de normalisation permet alors de calculer π_2 . On obtient :

$$\begin{aligned}
 \pi_2 = & \left[\frac{\lambda_{23} + \lambda_{24} + \lambda_{26}}{\lambda_1} + 1 + \frac{\lambda_{23}}{\lambda_3} + \frac{\lambda_{24}}{\lambda_4} + \frac{\lambda_{23} + \lambda_{24}}{\lambda_5} + \dots \right. \\
 & \left. \dots \frac{(1-p)(\lambda_{23} + \lambda_{24}) + \lambda_{26}}{\lambda_6} + \frac{\lambda_{23} + \lambda_{24}}{\lambda_{27}} + \frac{\lambda_{23} + \lambda_{24}}{\lambda_8} \right]^{-1}
 \end{aligned}$$

Les probabilités stationnaires π_i donnent la proportion du temps que l'élève passe dans chacun des états $i \in E$, en moyenne et sur le long terme. On peut répondre à d'autres questions à l'aide de cette distribution stationnaire de probabilités. Par exemple, nous savons que l'élève passera en moyenne une proportion du temps π_2 à l'arrêt *pôle bus*. La probabilité $p_{2,6}$, lorsqu'il est à l'arrêt pôle bus, qu'il renonce à aller à l'Esisar pour aller au bar peut se calculer à partir des taux de transition. On obtient en appliquant les formules de passage entre les deux caractérisations :

$$p_{2,6} = \frac{\lambda_{2,6}}{\lambda_{2,3} + \lambda_{2,4} + \lambda_{2,6}} = \frac{\lambda_{26}}{\lambda_{23} + \lambda_{24} + \lambda_{26}}$$

La proportion du temps que l'élève passera à l'état 2 avant de renoncer à aller à l'Esisar s'écrit donc

$$\frac{\lambda_{26} \pi_2}{\lambda_{23} + \lambda_{24} + \lambda_{26}}$$

4.3.3 méthode des coupes

Le calcul de la distribution stationnaire peut parfois être simplifié en utilisant une équation d'équilibre duale, au sens de la dualité dans les graphes. Cette méthode de calcul est appelée *méthodes des coupes*.

On considère, comme précédemment, une chaîne de Markov, $(X(t))_{t \geq 0}$, à temps continu homogène à valeurs dans un espace d'état discret E . Une *coupe* dans l'espace d'état E (ou du graphe associé à la CMTC) est toute partition de E en deux sous-ensembles E_1 et E_2 , non vides et disjoints :

$$\begin{aligned} E &= E_1 \cup E_2 \\ E_1 \cap E_2 &= \emptyset \end{aligned}$$

Nous allons décomposer l'équation (4.40) qui établit, en régime stationnaire, l'équilibre en chaque noeud j , selon cette partition. Cet équilibre des flux entrant et sortant au noeud j peut s'écrire :

$$\sum_{i \in E} \pi_i \lambda_{i,j} = \sum_{i \in E} \pi_j \lambda_{j,i}, \quad \forall j \in E$$

En décomposant E selon la partition $E = E_1 \cup E_2$, on obtient :

$$\sum_{i \in E_1} \pi_i \lambda_{i,j} + \sum_{i \in E_2} \pi_i \lambda_{i,j} = \sum_{i \in E_1} \pi_j \lambda_{j,i} + \sum_{i \in E_2} \pi_j \lambda_{j,i}$$

En sommant sur tous les indices $j \in E_2$, il vient :

$$\sum_{j \in E_2} \left[\sum_{i \in E_1} \pi_i \lambda_{i,j} + \sum_{i \in E_2} \pi_i \lambda_{i,j} \right] = \sum_{j \in E_2} \left[\sum_{i \in E_1} \pi_j \lambda_{j,i} + \sum_{i \in E_2} \pi_j \lambda_{j,i} \right]$$

c'est-à-dire, après simplification et permutation⁵ des indices de sommation :

$$\begin{aligned} \sum_{i \in E_1} \underbrace{\sum_{j \in E_2} \pi_i \lambda_{i,j}}_{\text{flux de } i \text{ vers } E_2} &= \sum_{j \in E_2} \underbrace{\sum_{i \in E_1} \pi_j \lambda_{j,i}}_{\text{flux de } j \text{ vers } E_1} \end{aligned}$$

Nous venons d'établir la proposition qui suit.

Proposition 21 (méthode des coupes)

Soit $(X(t))_{t \geq 0}$ une chaîne de Markov à temps continu homogène à valeurs dans E et $\{E_1, E_2\}$ une partition de l'espace d'état E . Alors, en régime stationnaire, le flux total de probabilités de E_1 vers E_2 est égal au flux total de probabilités de E_2 vers E_1 :

$$\sum_{i \in E_1} \sum_{j \in E_2} \pi_i \lambda_{i,j} = \sum_{j \in E_2} \sum_{i \in E_1} \pi_j \lambda_{j,i} \quad (4.41)$$

5. La permutation est autorisée car les sommes sont absolument convergentes

En définissant de nouvelles coupes sur le graphe associé à la CMTC, on obtient à chaque fois de nouvelles équations de la forme (4.41). Dans le cas particulier où l'espace d'état est fini, c'est à dire où $E = \{1, \dots, N\}$, $N - 1$ coupes indépendantes suffisent, avec la condition de normalisation (4.38), pour déterminer la valeur de toutes les probabilités du régime stationnaire. Les équations obtenues par la méthode des coupes ne sont pas les mêmes que celles obtenues en exprimant l'équilibre en chaque noeud du graphe. Mais, elles s'obtiennent comme combinaisons linéaires de ces dernières et sont équivalentes (i.e. elles fournissent la même solution). Dans un certain nombre de cas, les équations obtenues en choisissant judicieusement les coupes sont plus simples à résoudre que les équations d'équilibre aux noeuds. C'est le cas dans le processus de naissance et de mort examinés ci-dessous.

Exemple 12 (processus de naissances et de morts)

Les *processus de naissance et de mort* sont des cas particuliers de processus de Markov en temps continu où les transitions d'état sont de deux types seulement : les *naissances* où l'état passe de n à $n + 1$ et les *morts* où l'état passe de n à $n - 1$. Ces processus ont de nombreuses applications en dynamique des populations et dans la théorie des files d'attente. Un tel processus est représenté à la figure 4.10. Le modèle du processus de naissance

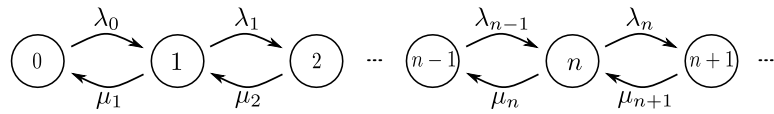


FIGURE 4.10 – *processus de naissance et de mort*

et de mort est défini par les *taux de naissance* $(\lambda_n)_{n \geq 0}$ et les *taux de mortalité* $(\mu_n)_{n \geq 0}$. Le *générateur infinitésimal* (infini) associé à ce modèle s'écrit donc :

$$\mathbf{A} = \begin{pmatrix} -\lambda_0 & \lambda_0 & 0 & \dots \\ \mu_1 & -(\lambda_1 + \mu_1) & \lambda_1 & \dots \\ 0 & \mu_2 & -(\lambda_2 + \mu_2) & \ddots \\ \vdots & \vdots & \ddots & \ddots \end{pmatrix} \quad (4.42)$$

L'équilibre au noeud n s'écrit :

$$\lambda_{n-1}\pi_{n-1} + \mu_{n+1}\pi_{n+1} = (\lambda_n + \mu_n)\pi_n, \quad \forall n \geq 1 \quad (4.43)$$

Il s'agit d'une récurrence à trois termes en les inconnues π_n , lesquelles doivent satisfaire de plus la condition de normalisation

$$\|\pi\|_1 := \sum_{n \geq 0} \pi_n = 1 \quad (4.44)$$

ainsi que l'équilibre au noeud 0 :

$$\lambda_0 \pi_0 = \mu_1 \pi_1 \quad (4.45)$$

Examinons maintenant les équations obtenues par la méthode des coupes, à l'aide des différentes coupes représentées à la figure 4.11. L'équation qui

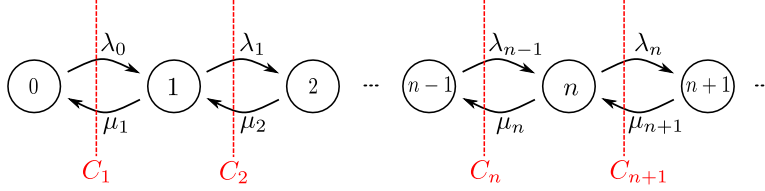


FIGURE 4.11 – Les coupes indépendantes (représentées en rouge) utilisées pour calculer, par la méthode des coupes, la distribution stationnaire de probabilités dans le cas du processus de naissance et de mort

correspond à la coupe C_1 , qui sépare $E_1 := \{0\}$ de $E_2 := \{1, 2, \dots\}$, est exactement l'équation (4.45). Mais les équations qui correspondent aux coupes C_n , qui sépare $E_1 := \{0, \dots, n-1\}$ de $E_2 := \{n, n+1, \dots\}$, s'écrivent sous la forme :

$$\lambda_{n-1} \pi_{n-1} = \mu_n \pi_n, \quad \forall n \geq 1 \quad (4.46)$$

Il s'agit d'une récurrence à deux termes, beaucoup plus simple à résoudre que la récurrence à trois termes obtenues par la méthode des équilibres en chaque noeud. On obtient directement :

$$\pi_n = \pi_0 \prod_{k=0}^{n-1} \frac{\lambda_k}{\mu_{k+1}}, \quad \forall n \geq 1 \quad (4.47)$$

et la condition de normalisation donne :

$$\pi_0 = \frac{1}{1 + \sum_{n=1}^{\infty} \prod_{k=0}^{n-1} \frac{\lambda_k}{\mu_{k+1}}} \quad (4.48)$$

La convergence de la série au dénominateur, dans l'équation (4.48), est donc une condition nécessaire et suffisante d'existence de la distribution stationnaire de probabilités (aussi appelée *condition d'ergodicité*) pour le processus de naissance et de mort. Une condition nécessaire et suffisante pour la convergence de cette série est l'existence d'un indice $k_0 >$ tel que :

$$k \geq k_0 \Rightarrow \frac{\lambda_k}{\mu_k} < 1 \quad (4.49)$$

L'exemple des processus de naissance et de mort considéré ci-dessus, en particulier les résultats (4.47), (4.48) et (4.49), seront utilisés très souvent dans le chapitre du cours consacré aux files et réseaux de files d'attente.

4.4 Exercices

Exercice 23 (la file simple)

On considère un processeur auquel arrivent des requêtes, ces arrivées étant modélisées par un processus de Poisson de paramètre λ . Le temps que met le processeur à traiter chaque requête est modélisé par une variable aléatoire de distribution exponentielle $Exp(\mu)$, de paramètre μ . Le processeur traite la première requête qui lui arrive. Pendant ce temps, les autres requêtes qui arrivent sont rangées dans *une file d'attente*. Lorsque le processeur a terminé la traitement de la requête en cours, il commence le traitement de la requête la plus ancienne dans la file d'attente (discipline de file “*first in, first out*” (FIFO), ou *premier arrivé, premier servi*).

1. Modéliser ce problème par une chaîne de Markov à temps continu (CMTC) : donner le graphe associé avec les taux de transitions, écrire le générateur infinitésimal \mathbf{A} et donner la caractérisation terme de temps de séjour et de probabilités de transition.
2. Déterminer le régime stationnaire.
3. Quel sont - en régime stationnaire - le nombre moyen de requêtes dans le système (buffer et processeur ensemble), la taille moyenne de la file (i.e. le nombre de requêtes en attente dans le buffer) et le taux d'utilisation du processeur (en pourcentage du temps total) ?

Exercice 24 (maintenance)

Un parking souterrain est éclairé par K ampoules allumées 24 heures sur 24, 7 jours sur 7. Dans cet exercice, on s'interroge sur la qualité des ampoules à acheter (investissement) et sur la fréquence des visites de à effectuer (maintenance). La durée de vie des ampoules est une variable aléatoire exponentielle de paramètre λ (plus λ est faible, plus les ampoules coûtent cher !). Le concierge du parking visite de temps en temps le parking, les dates de ses visites étant espacées par un temps de loi $Exp(\mu)$. A chacune de ses visites, le concierge remplace toutes les ampoules en panne par des ampoules neuves. Il ne touche pas aux autres ⁶.

1. Modéliser l'évolution temporelle des conditions d'éclairage du parking (nombre d'ampoules qui fonctionnent) par une CMTC. Donner le graphe associé, préciser les taux de transition et écrire le générateur infinitésimal.
2. Quelle est la probabilité qu'un usager du parking se retrouve dans le noir complet quand il va chercher sa voiture ? On supposera que l'installation fonctionne depuis suffisamment longtemps pour que sa

6. une action de maintenance préventive ne peut pas être optimisée à l'aide d'un modèle sous forme de CMTC où l'usure progressive des ampoules n'est pas prise en compte (absence de mémoire)

configuration initiale n'ait plus d'effet sur son fonctionnement actuel (i.e. on suppose que le régime stationnaire est établi).

3. A l'aide de la méthode des coupes, calculer le nombre moyen d'ampoules qui fonctionnent (en régime stationnaire)

Exercice 25

Une salle est équipée de n ordinateurs. La durée de fonctionnement de chacun des ordinateurs suit une loi $Exp(\mu)$. La durée de remise en état de chaque ordinateur suit une loi $Exp(\lambda)$. Le réseau est maintenu en service par n techniciens. Toutes les durées sont indépendantes. On note $X(t)$ le nombre d'ordinateurs qui fonctionnent à l'instant t .

1. Donner le graphe et les taux de transition pour $X(t)$
2. Donner la distribution stationnaire de probabilités
3. Donner, en régime stationnaire, le nombre moyen d'ordinateurs en fonctionnement.
4. Donner le graphe si la salle ne possède plus qu'un technicien

Exercice 26 (croissance logistique)

Nous nous intéressons dans cet exercice à l'évolution dynamique de la taille d'une population. Des modèles par chaînes de Markov sont fréquemment utilisés pour établir ces bilans de population, que ce soit en biologie (e.g. croissance bactérienne, modèles proies-prédateurs, etc.), en démographie humaine ou en physique (e.g. désintégration de particules radioactives). Dans notre cas, nous allons nous intéresser à un processus de naissances et de morts, pour lequel le générateur infinitésimal est tridiagonal : à partir d'une population de taille k , seules les transitions vers les tailles $k - 1$ et $k + 1$ sont possibles. Nous allons en particulier nous intéresser au *modèle logistique*. Dans celui-ci, la taille de la population peut évoluer entre une taille minimale \underline{X} et une taille maximale \bar{X} (avec $0 \leq \underline{X} < \bar{X}$). Dans le modèle logistique, le taux de naissance⁷ d'une population de taille X est donné par :

$$\lambda = \alpha (\bar{X} - X)$$

où α est un paramètre réel positif qui représente la capacité de reproduction de l'espèce pour une population de taille donnée. Le taux de natalité λ diminue donc au fur et à mesure que la population croît en taille, jusqu'à atteindre une valeur nulle pour $X = \bar{X}$ (valeur de saturation). C'est une manière de représenter la raréfaction des ressources nécessaires à la survie de la population quand la taille de celle-ci augmente. De même, le taux de mortalité d'une population de taille X , dans le modèle logistique, est donné par :

$$\mu = \beta (X - \underline{X})$$

7. Le taux de naissance se définit comme la variation par unité de temps de la taille de la population rapportée à la taille de la population

où β est un paramètre réel positif. Le taux de mortalité diminue au fur et à mesure que la population diminue en taille, jusqu'à atteindre une valeur nulle pour $X = \underline{X}$ (valeur de saturation). C'est une manière de représenter l'abondance de ressources et la plus grande facilité à survivre quand la taille de la population est faible.

1. Modéliser par une CMTC l'évolution dans le temps de la taille de la population du modèle logistique. Donner le graphe associé et les taux de transition.
2. Calculer la distribution de probabilité limite de cette population
3. On considère le cas particulier où $\underline{N} = 0$ et où lorsque la taille de la population devient nulle, aucune reproduction n'est plus possible. Déterminer quelle est la probabilité P_k d'extinction de l'espèce si à l'instant initial la population est de taille k ⁸. Calculer ensuite le temps moyen T_k d'extinction de l'espèce en partant d'une population initiale de taille k .

Exercice 27

On considère le réseau de transmission par paquets, représenté à la figure 6.30, constitué d'un émetteur E , de deux noeuds relais en série notés respectivement 1, 2 et d'un récepteur R . Les noeuds 1 et 2 ont une capacité

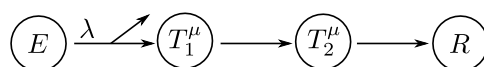


FIGURE 4.12 – réseau de transmission avec deux noeuds relais en série

limitée à un seul paquet (pas de buffer). Les paquets sont émis par E selon un processus de Poisson de paramètre λ . Lorsqu'un paquet arrive au noeud 1 alors que ce dernier est déjà occupé, le paquet est perdu. Un paquet ne peut être émis du noeud 1 vers le noeud 2 que lorsque le noeud 2 est vide. Le temps d'émission des noeuds 1 et 2 sont distribués selon des variables exponentielles de paramètre μ . Tant qu'un paquet est en cours d'émission, on considère qu'il occupe une place sur le noeud de départ. Le noeud 1 est informé instantanément de l'état du noeud 2 (pour décider de l'émission éventuelle d'un paquet). Lorsque R reçoit un paquet, il le traite instantanément et peut immédiatement recevoir un autre paquet éventuellement envoyé par le noeud 2 (le noeud R représente la *sortie* du réseau). Le temps de propagation sur les liaisons est négligeable devant le temps d'émission.

1. Modéliser l'état du réseau de transmission par une chaîne de Markov à temps continu à quatre états (définir précisément la signification

8. Passer par la caractérisation de la CMTC en termes de temps de séjour et de probabilités de transition. Résoudre le problème directement sur la CMTD induite par la CMTC

- de chacun des quatre états). Donner le graphe et le générateur infinitésimal (matrice A des taux de transition).
2. En utilisant les relations d'équilibre en chaque noeud, déterminer le profil stationnaire de probabilités. S'agit-il d'une distribution limite? Pourquoi?
 3. Donner la probabilité de rejet des paquets
 4. Représenter le graphe de la CMTC associée au réseau de la figure 6.31 et indiquer les taux associés à chacune des transitions. Si les noeuds 2 et 3 sont vides, un noeud qui quitte le noeud 1 est routé de manière équiprobable vers les noeuds 2 ou 3. Si un seul des deux noeuds 2 ou 3 est vide, alors un paquet quittant le noeud 1 est routé vers le paquet libre. Si les deux noeuds 2 et 3 sont occupés, les paquets sont bloqués en amont au noeud 1.

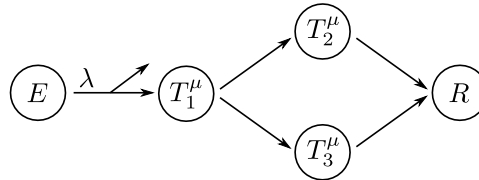


FIGURE 4.13 – réseau de transmission avec un noeud relais en série avec deux autres noeuds relais en parallèle

Chapitre 5

Files et réseaux de files d'attente

5.1 Introduction et définitions

Nous allons dans ce chapitre modéliser et analyser des systèmes de partage de ressources. Une ressource est un composante (physique, logique, humaine, etc.) que les clients du système doivent obtenir pour réaliser une activité. Le client qui fait une demande d'accès à cette ressource doit attendre qu'elle soit disponible. Il entre ensuite *en service* et mobilise donc la ressource pour la durée de ce service. Il libère ensuite la ressource qui redevient disponible pour les autres clients en attente.

Dans les systèmes que nous considérons, les clients se présentent de manière aléatoire (i.e. selon un processus d'arrivée stochastique, par exemple un processus de Poisson) et la durée de service du client par le serveur est également une variable aléatoire. L'état sera généralement défini comme le nombre de clients présents dans le système à l'instant $t > 0$. Les files ainsi définies sont donc des exemples de chaînes à temps continu.

La théorie des files et réseaux de files d'attentes est très utilisée pour modéliser, analyser et concevoir de nombreux systèmes, comme par exemple :

- *les systèmes informatiques* - les clients peuvent être des processus ou des requêtes E/S, les ressources peuvent être des processeurs ou des périphériques, l'activité du serveur correspond à un temps de traitement ou à un temps de lecture/écriture
- *les réseaux de communication* - les clients peuvent être des messages ou des paquets, les serveurs peuvent être les noeuds du réseau ou les canaux de transmission (ou les deux) et l'activité de ces *serveurs* correspond au temps de traitement, d'émission ou de transmission des paquets
- *les systèmes de production* - les clients peuvent être des pièces (ou des palettes de pièces), les serveurs peuvent être des machine-outils

ou des convoyeurs, les temps de service peuvent correspondre à de l'usinage, du chargement, du transport, etc.

- *les systèmes de service* - les clients sont alors des individus et les serveurs des personnes ressources ou des services. On peut ranger dans cette catégorie aussi bien des files simples (guichet à la poste, files dans les supermarchés, etc.) que des systèmes sociaux de grande taille avec une topologie complexe et des services variés

Nous étudierons dans ce chapitre des exemples de files simples et des réseaux de files d'attente, c'est-à-dire des files simples reliées entre elles par une topologie d'interconnexion et où les clients sont soumis à un routage spécifique. Pour de tels systèmes, nous chercherons à établir les conditions de stabilité (éviter que la taille des files n'explose) ainsi que les paramètres de performance (débit moyen, nombre moyen de clients dans la file, temps moyen de service, etc.). Nous chercherons à analyser l'influence de la structure des files simples (nombre de serveurs, discipline de service, etc.), ainsi que la celle de la topologie et du routage des réseaux de files d'attente, sur ces paramètres de performance

Le chapitre est organisé de la manière suivante. Dans une première partie, nous allons décrire les files simples (section 5.1.1) et différentes catégories de réseaux de files d'attente (section 5.1.2). Nous proposerons ensuite quelques exemples pour motiver l'étude de ces systèmes (section 5.2). Nous étudierons ensuite les paramètres de performance de manière générale, pour les files et les réseaux de files d'attente. Nous commencerons par les définir (section 5.3.1). Ensuite nous étudierons la stabilité des files (section 5.3.2), nécessaire pour que ces paramètres aient un sens. Enfin, nous énoncerons la loi de Little qui relie ces paramètres entre eux (section 5.3.3). Dans la section 5.4, nous caractériserons différents exemples de files simples, avec des arrivées poissonniennes ou non, des temps de service exponentiels ou non, un ou plusieurs serveurs, des capacités limitées ou non, différentes disciplines de service, etc. Enfin, dans la section ??, nous donnerons des résultats généraux concernant un type particulier de réseaux de files d'attente : les réseaux à forme produit.

5.1.1 La file simple

Dans une file simple, l'activité d'un client se résume à attendre la disponibilité de la ressource (i.e. d'un serveur délivrant cette ressource), puis à attendre ensuite la réalisation du service associé pendant un temps de service (temps pendant lequel le serveur délivre effectivement la ressource à ce client). La file simple est représentée schématiquement à la figure 5.1. Dans de nombreuses applications, l'activité d'un client nécessite l'accès à une succession de services. Il est alors nécessaire d'utiliser un réseau de files d'attente simples (voir la section suivante).

Pour caractériser une file simple, il faut définir le processus d'arrivée des clients, le temps de service, la structure de la file (nombre de serveurs, ca-

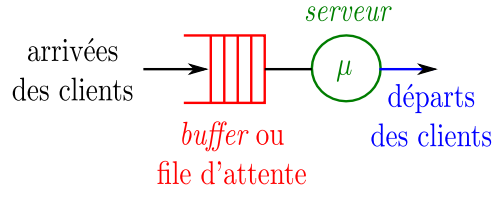


FIGURE 5.1 – représentation schématique d'une file simple.

pacité) et la discipline de service.

Processus d'arrivée

Le processus d'arrivée des clients sera décrit comme un *processus de comptage* $(N(t))_{t \geq 0}$, c'est-à-dire une chaîne à temps continu, à valeur dans \mathbb{N} , non décroissante et telle que $N(0) = 0$. On définira l'instant d'arrivée du client n par

$$\begin{aligned} A_n &:= \inf \{t \mid N(t) = n\}, \text{ pour } n \geq 1 \\ A_0 &:= 0 \end{aligned} \quad (5.1)$$

et les intervalles entre les arrivées par

$$T_n := A_n - A_{n-1}, \text{ pour } n \geq 1 \quad (5.2)$$

Dans la plupart des exemples, le processus d'arrivée sera un *processus de renouvellement*, c'est-à-dire un processus de comptage pour lequel les variables aléatoires $\{T_n\}_{n \geq 1}$ sont indépendantes et ont toutes la même loi (i.e. la même densité de probabilité). Cette loi suffit donc à caractériser entièrement le processus de renouvellement. On notera dans ce cas $\bar{T}_a := E[T_n]$ l'intervalle *moyen* entre deux arrivées de clients et

$$\lambda := \frac{1}{\bar{T}_a} = \frac{1}{E[T_n]} \quad (5.3)$$

le *taux moyen d'arrivée* des clients dans la file. Le processus de Poisson, par exemple, est un processus de renouvellement où les intervalles T_n sont tous des variables exponentielles $\text{Exp}(\lambda)$.

temps de service

Le temps de service pour le client n est une variable aléatoire qui sera notée S_n . La plupart du temps, on considérera des serveurs pour lesquels les temps de service consécutifs sont décrits par des variables aléatoires $\{S_n\}_{n \geq 1}$

indépendantes et de même loi. On notera dans ce cas $\bar{T}_s := E[S_n]$ la durée moyenne des temps de service et

$$\mu := \frac{1}{\bar{T}_s} = \frac{1}{E[S_n]} \quad (5.4)$$

le *taux moyen* de service ou parfois simplement le *taux de service*. La distribution la plus simple est la distribution exponentielle. La propriété d'absence de mémoire de la loi exponentielle facilite souvent les calculs mais peut ne pas être très réaliste dans certaines applications. D'autres distributions sont donc également utilisées pour modéliser les temps de service : uniforme, constante (déterministe), normale, Gamma, etc.

L'instant de départ D_n correspond toujours à la fin de service du client n . Cependant, les instants de départ ne permettent en général pas de retrouver les durées de service :

$$S_n \neq D_n - D_{n-1}$$

Il se peut en effet qu'un client, qui quitte le système, laisse le serveur inoccupé jusqu'à l'arrivée du client suivant. Dans l'exemple la figure 5.2 ci après, le nombre total de clients dans le système $Q(t)$ est nul sur $t \in [D_3, A_4]$. En conséquence $S_4 \neq D_4 - D_3$.

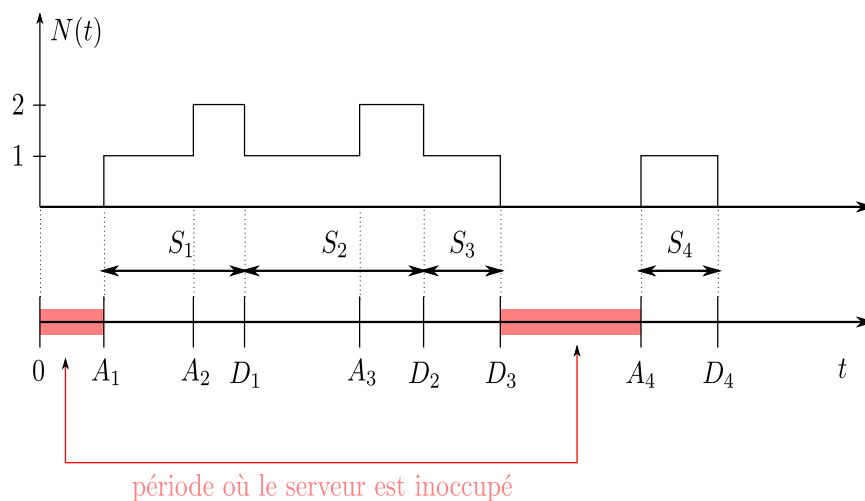


FIGURE 5.2 – Dans une file simple, le serveur peut être inoccupé à certains moments. Les instants de départs, D_n , ne permettent donc pas, à eux seuls, de retrouver les durées de service S_n

structure de la file d'attente

Une file peut avoir plusieurs serveurs qui opèrent en parallèle. Cette configuration est représentée à la figure 5.3 pour le cas où il y a C ($C \geq 2$)

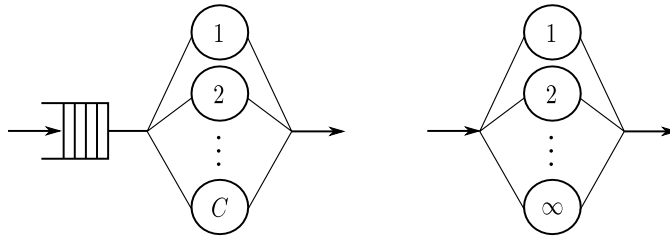


FIGURE 5.3 – représentation schématique d’une file avec C serveurs en parallèle, à gauche, ou avec une “infinité” de serveurs (i.e. plus de serveurs que de clients, à droite).

serveurs en parallèle. Dès qu’un client entre dans le système la station, il est pris en charge par un serveur si il y a un serveur disponible ou il est placé dans la file d’attente, en attendant la libération d’un des serveurs. Il est nécessaire de définir la durée de service de chacun des serveurs. En général (et donc aussi par défaut), on suppose que les serveurs sont identiques et que les durées de services sont des variables aléatoires indépendantes, mais de même loi pour tous les serveurs. Un cas particulier est la file IS (*infinite servers*) dans laquelle il n’y a pas de buffer. Ce type de file modélise en général un retard aléatoire pur (le temps de traitement du client) dans un réseau plus complexe. Il permet de représenter des systèmes pour lesquels le nombre de serveurs est toujours supérieur au nombre de clients qui peuvent s’y trouver.

La capacité d’une file peut être finie ou non. Nous appellerons K la capacité de la file (on considérera arbitrairement $K = \infty$ lorsque la capacité de la file n’est pas bornée). Si il y a déjà K clients dans le système lorsqu’un nouveau client arrive (le système est “plein”), alors ce dernier est perdu. Les files à capacité limitée sont représentée (en général) comme à la figure 5.4.

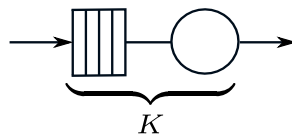


FIGURE 5.4 – représentation schématique d’une file à capacité K limitée

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). Dans le cas où il n'y a qu'un serveur, ce sont les mêmes disciplines. Dans le cas où il y a plusieurs serveurs en parallèle, il n'est pas possible de décider à priori quel client sortira en premier de son serveur (on aura donc une discipline FCFS), sauf à forcer un client servi à attendre un autre arrivé avant lui, avant de sortir (pour restaurer l'ordre d'arrivée et la discipline FIFO)
- LIFO (last in, first out) ou LCFS (last come, first served). Cela correspond à une pile, dans laquelle le dernier client arrivé est posé sur le haut de la pile et sera le premier retiré de la pile dès qu'un serveur se libère. Les disciplines LIFO et LCFS ne sont pas équivalentes dans le cas multi-serveurs.
- RANDOM. Le client dans le buffer qui est servi lorsqu'un serveur se libère est choisi aléatoirement dans la file, en général avec une distribution de probabilité (discrète) uniforme
- RR (Round-Robin, ou cyclique). Tous les clients de la file d'attente entrent en service à tour de rôle, effectuent un quantum Q de leur temps de service et sont replacés dans la file, jusqu'à ce que leur service ait pu être totalement accompli à l'aide d'un nombre suffisant de quantai. C'est une discipline de service souvent utilisée pour modéliser la gestion des processus par le processeur dans un système informatique
- 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 dans la file. Par exemple, si le temps de service du serveur pour un client est une exponentielle $\text{Exp}(\mu)$, et qu'il y a n clients présents dans la station à l'instant considéré, alors le taux effectif de service pour chacun des client est $\frac{\mu}{n}$ (ce qui ne signifie pas que les n clients seront libérés en même temps).

Notation de Kendall

En théorie des files d'attente, la notation de Kendall est une notation qui permet de décrire un système à l'aide de six paramètres. Elle porte le nom du mathématicien David George Kendall, qui l'a introduite en 1953. Elle est constituée par une suite de six symboles standardisés qui prend la forme

$$T/S/C/K/m/Z.$$

où

- T désigne la distribution des temps inter-arrivées. On notera par exemple $T = GI$ pour des distributions indépendantes de même loi

- générale, $T = M$ pour un processus d'arrivées poissonien, D pour des arrivées déterministes, etc.
- S indique la loi de probabilité de la durée du service ; on utilise les mêmes symboles que précédemment.
 - C indique le nombre de serveurs
 - K est la capacité totale du système, c'est-à-dire le nombre de serveurs (C), plus le nombre de places en attente.
 - m indique la taille de la population totale de clients, lorsque celle-ci est à priori finie (et connue)
 - Z désigne la discipline de service, par exemple FIFO, LIFO, RANDOM, etc.

Très souvent, les trois derniers symboles de la notation sont omis. Dans ce cas, il faut considérer par défaut $K = \infty$, $m = \infty$ et $Z = FCFS$. La file $M/M/1$ désigne ainsi la file simple avec des arrivées poissoniennes, un temps de service exponentiel et une discipline de service FIFO (équivalente à FCFS dans ce cas. La file $M/M/N/N$ (avec $N \in \mathbb{N}$), par exemple, désigne une file markovienne avec N serveurs et à capacité totale N . Il s'agit donc d'une file sans buffer où les clients qui se présentent alors qu'il y a déjà N clients dans le système sont perdus. En général les systèmes avec une population totale m finie sont représentés comme un réseau fermé constitué de deux sous-systèmes : la file proprement dite et l'extérieur.

Remarque 12 (files multi-classes) Une file d'attente peut être parcourue par différentes classes de clients (voir figure 5.5). Ces différentes classes pourront se distinguer en général par des processus d'arrivées différents, des temps de service différents et un ordonnancement dans la file d'attente fonction de leur classe. Pour définir une file multiclassées, il faut donc 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 (*avec préemption*) et HOL (sans préemption ou *hold on line*). Par exemple, dans une file à un serveur, parcourue par deux classes de clients, les clients de la classe 1, prioritaires, seront placés *devant* les clients de la classe 2. Pour caractériser la discipline de service prioritaire, il faut préciser que faire lorsqu'un client de la classe 1 arrive alors qu'un client de la classe 2 est en service. Si le service est préemptible (PR), le client de classe 2 est instantanément replacé dans la file d'attente afin de laisser le serveur disponible pour le client de classe 1. Il faut encore distinguer les cas où le temps de service déjà effectué (pour le client de classe 2) est mémorisé (valorisable) ou non. Dans le cas d'un service sans préemption (HOL), le client de classe 1 doit attendre la fin de service du client 2 en cours avant d'être servi lui-même (avant les autres clients de type 2).

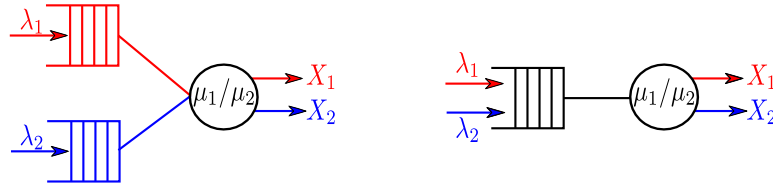


FIGURE 5.5 – deux représentations schématiques équivalentes d'une file multi-classes. Les clients de classe 1 et 2 arrivent avec des taux respectifs λ_1 et λ_2 . Ils sont placés en file d'attente (buffers séparés ou commun) et servis selon une discipline de service à priorité. Les temps de services peuvent être distincts pour les deux types de clients. Les débits moyen de clients des deux classes, notés X_1 et X_2 , sont en général différents.

5.1.2 Les réseaux de files d'attente

Un réseau de files d'attente est un ensemble de files simples (parfois appelées *station*), au sens de la section précédente, reliées par une *topologie d'interconnexion*, à travers laquelle les clients (éventuellement de plusieurs classes distinctes) sont routés de manière déterministe ou probabiliste, stationnaire ou dynamique. Les réseaux peuvent être ouverts ou fermés, à capacité limitée ou non.

Un réseau est dit *ouvert* si les clients peuvent entrer dans le réseau depuis l'extérieur ou le quitter. En général, les clients peuvent arriver dans un réseau ouvert, ou en sortir, à chacun de ses noeuds (files simples). Un réseau est dit *fermé* quand les clients ne peuvent ni y entrer, ni le quitter. Le nombre de clients dans un réseau fermé est donc constant. Un réseau dans lequel un nouveau client entre dès qu'un client le quitte est considéré également par extension comme un réseau fermé.

Les réseaux de files d'attente peuvent également être *multi-classes*. Les différentes classes de clients peuvent se distinguer tant par leur temps de service et leurs disciplines de services dans chacune des files simples du réseau, que par un routage spécifique. Il est également possible qu'un client puisse changer de classe en quittant une file pour en rejoindre une autre. Si les clients d'une classe donnée ne peuvent ni entrer ni sortir du réseau, cette classe de client est dite fermée. Dans le cas contraire, elle est dite ouverte. Le réseau peut être fermé pour certaines classes de clients et ouverts pour d'autres classes. On parle alors de *réseau mixte* (voir figure 5.6).

Dans un réseau, certaines files peuvent être à *capacité limitée*. Aucun client ne peut plus entrer dans le buffer de cette file lorsque la capacité est atteinte. Si le système est ouvert, certains clients à l'entrée du réseau peuvent être perdus. Certains clients également peuvent être *bloqués* au sein du réseau, par exemple dans des files *en amont* du buffer plein. On distingue principalement deux types de blocage : le blocage *avant service* et le *blocage*

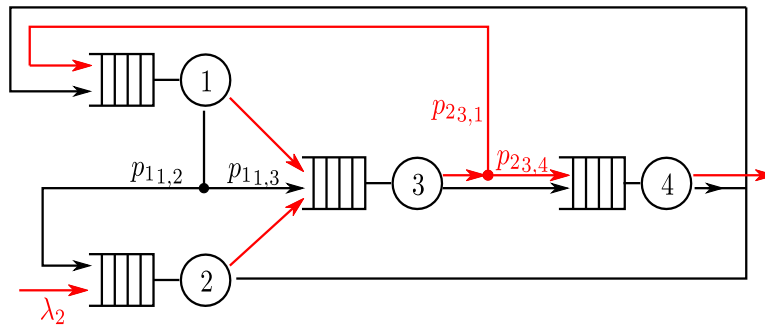


FIGURE 5.6 – représentation schématique d'un réseau mixte avec deux classes de clients : la classe 1 (en noir) est fermée et la classe 2 (en rouge) est ouverte

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.

Le réseau dans son ensemble peut être lui même à capacité limitée : soit qu'il soit fermé (population de clients constante), soit qu'il soit ouvert mais à *contrainte de population* (voir figure 5.7). Cette contrainte de population peut être réalisée par un mécanisme de type sémaphore et matérialisée par des *jetons* nécessaires aux clients extérieurs pour entrer dans le réseau, qui les accompagnent dans leur parcours au sein du réseau et sont libérés par les clients qui en sortent. Les clients en attente d'un jeton libre peuvent être rangés dans un buffer extérieur (tout comme les jetons excédentaires en attente de clients). Une alternative est que les clients en excès (par rapport à la contrainte de population) soient perdus. Dans ce cas, le réseau est équivalent à un réseau fermé.

Pour définir un réseau de files d'attente, il est nécessaire de caractériser :

- chacune des M files simples qui constituent le réseau
- le processus (global) d'arrivée des clients dans le réseau depuis l'extérieur
- le *routing* des clients en chacun des noeuds du réseau

Le nombre moyen de clients qui arrivent dans le système par unité de temps peut être insuffisant pour caractériser l'arrivée des clients. Il faut en plus préciser lorsqu'un client arrive dans le réseau, dans quelle file il se rend.

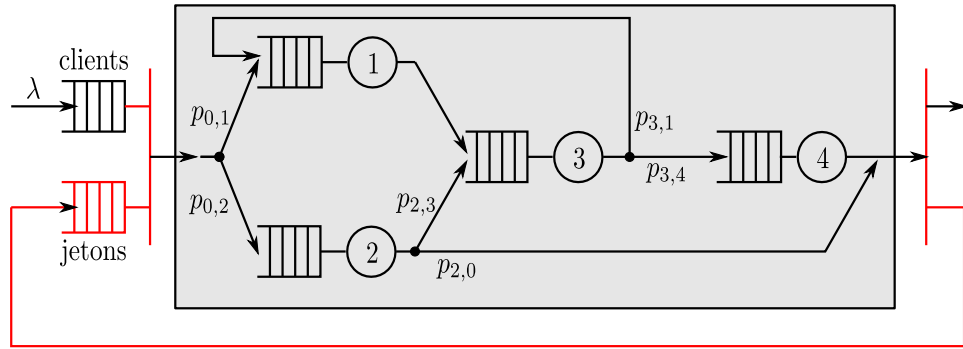


FIGURE 5.7 – représentation par un modèle de type sémaphore (en rouge) pour un réseau ouvert à contrainte de population (boîte grisée). Le nombre de jetons disponibles est égal à la taille maximale de la population. Un client, si il veut entrer dans le réseau, doit prendre un jeton dans le buffer jeton, puis le restituer à la sortie du réseau.

On caractérisera la plupart du temps ce routage d'entrée comme un *routage probabiliste*, par exemple en donnant les probabilités $p_{0,i}$ pour qu'un client qui arrive dans le réseau se rende à la station i , pour toutes les stations i du réseau. Si les arrivées globales dans le système sont poissonniennes de taux λ , alors les processus d'arrivées des clients (uniquement ceux venant de l'extérieur) à la station i sont poissonniens et de taux $p_{0,i}$ (voir la décomposition des processus de Poisson, dans le premier chapitre).

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). A nouveau, le routage des clients est très souvent caractérisé de façon probabiliste. Nous noterons (dans ce contexte) $p_{i,j}$ la probabilité pour qu'un client qui quitte la station i se rende à la station j et $p_{i,0}$ la probabilité pour qu'un client qui quitte la station i quitte le système. Les probabilités $p_{i,j}$ sont telles que

$$\sum_{j=0}^M p_{i,j} = 1$$

où M est le nombre total de files simples dans le réseau. Il existe cependant d'autres types de routages :

- les *routages dynamiques* : par exemple, un client quittant une station choisira, parmi toutes les destinations possibles, la station qui comporte le moins de clients. La route choisie dépend donc de l'état courant des différentes files, c'est-à-dire de l'état global du réseau
- les *routages déterministes* : par exemple, le routage cyclique où les clients quittant une station choisiront à tour de rôle chacune des stations parmi toutes les destinations possibles

5.2 Exemples

Nous allons survoler quelques exemples de réseaux de files d'attente de différents types : ouverts, fermés, etc. Différents domaines seront abordés tels que les modèles de systèmes informatiques, de processeurs simples ou multiples, des architectures clients-serveurs, des systèmes de communication, des systèmes de votes ou des systèmes de production. Les applications de la théorie des réseaux de files d'attente sont innombrables ...

5.2.1 Machine repairman model

Un réseau simple et très utilisé est le modèle du réparateur (*machine repairman model* en anglais) représenté à la figure 5.8. Un exemple d'application est celui d'un système informatique central où les m machines du modèle représentent des terminaux, susceptible d'adresser des requêtes au réparateur qui représente l'ordinateur central. Un autre exemple est celui d'un atelier où m machines fonctionnent en parallèle, indépendamment les unes des autres, et sont réparées par un réparateur unique lorsqu'elles tombent en panne. Dans ce système, les machines sont représentées par une file à m serveurs dans laquelle il n'y a pas de buffer (les machines n'attendent pas, lorsqu'elles sont réparées, pour être remise en service). Ce modèle représente simplement le temps de service aléatoire des machines entre deux pannes (il peut être différent pour chaque machine). Quand une machine tombe en panne, elle envoie une requête au réparateur. Selon la discipline de service choisie, cette demande peut être placée dans la file en attendant que d'autres requêtes, sollicitées par d'autres machines, soit traitées. Le modèle du réparateur est un réseau fermé constitué de deux files, qui comporte M clients qui sont les machines, en attente d'être réparées (dans le buffer), en train d'être réparées (dans le réparateur) ou en activité.

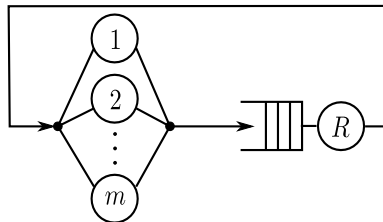


FIGURE 5.8 – représentation du *machine repairman model* par un réseau fermé de deux files simples

5.2.2 Systèmes informatiques

Les réseaux de files d'attente sont très utilisés pour modéliser et analyser les performances de systèmes informatiques, que ce soit des modèles développés des les années 70' pour représenter des architectures mono-processeur avec terminaux et périphériques entrées-sorties, des systèmes mono-processeur à degré de multiprogrammation développés par la suite ou, plus récemment des modèles multi-processeurs avec des architectures parallèles ou distribuées. Les réseaux de files d'attente ont été également utilisés pour évaluer les performances relatives de différents noyaux Unix dans la gestion des tâches.

modèle serveur central

Le système considéré comporte une unité centrale qui exécute des processus et un ensemble d'unités entrée-sortie (mémoire, terminaux, etc.) notées ES_i . Les entités qui circulent dans le réseau sont des processus. Ces processus peuvent être dans différents états : *prêt* en attente de libération de l'unité centrale, *élu* et en cours d'exécution sur l'unité centrale ou *en attente* pour l'accès à un périphérique ES_i . Le système est modélisé par le *réseau ouvert* de files d'attente représenté à la figure 5.9. Les clients du réseau de

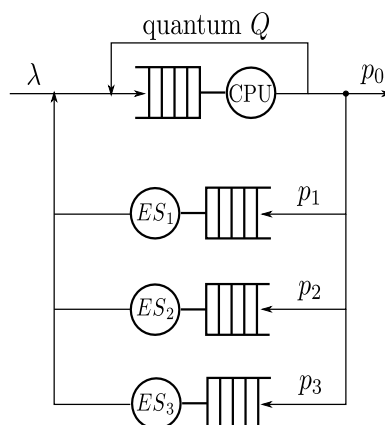


FIGURE 5.9 – modèle simple d'un ordinateur avec une discipline de service Round-Robin de quantum Q pour le CPU.

la figure 5.9 sont les processus. Les différentes ressources sont représentées par des files simples : une pour le CPU et une pour chacune des unités ES_i . Les processus arrivent dans le réseau, générés par des utilisateurs *extérieurs* avec un taux d'arrivées λ . Les clients présents dans le buffer du CPU sont les processus *prêts* et le client en service dans le processus est le client *élu*. La discipline de service du CPU est une discipline de type *Round-Robin* (RR) avec un quantum Q . A la fin de leurs services, les clients peuvent quitter le système avec une probabilité p_0 (lorsque leur exécution est terminée) ou

être dirigés vers les files ES_i avec une probabilité p_i , lorsque leur traitement nécessite un accès en lecture/écriture sur la ressource correspondante.

Souvent, le modèle ne fait pas la distinction entre processus *prêts* et *élus*. Dans ce cas, on considère que les processus qui ne sont pas en attente dans les buffer des files ES_i sont en exécutions simultanées sur le processeur. La discipline de service du CPU est alors une discipline de type *Processor Sharing* (PS) avec un taux moyen de service μ qui correspond à l'inverse du temps moyen d'exécution d'un processus sur le CPU entre deux accès successifs de ce processus aux files ES_i . Ce modèle appelé *serveur central* est représenté à la figure 5.10. Dans celui-ci on considère que les disciplines de service des files ES_i sont de type FIFO et que les temps de lecture/écriture respectifs sont des exponentielles de paramètre μ_i . Par souci de simplicité, on suppose la plupart du temps que les requêtes extérieures (nouveaux processus) arrivent dans le système selon un processus de Poisson de paramètre λ .

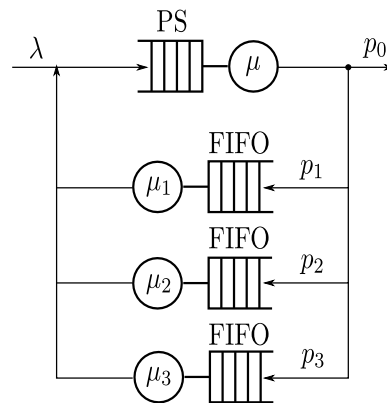


FIGURE 5.10 – modèle *serveur central* d'un ordinateur avec une discipline de service Processor Sharing (PS) pour le CPU.

système à degré de multiprogrammation

Dans un système réel, la limitation de la capacité mémoire impose de limiter le nombre total de processus qui peuvent se trouver simultanément en cours d'exécution ou en attente dans le système. Ce nombre maximum de client, noté K , est appelé *degré de multi-programmation*. Lorsque cette limite est atteinte, les nouveaux processus sont stockés sur le disque et ne pourront entrer dans le système que lorsqu'un processus en cours d'exécution se terminera et quittera le système. Le modèle obtenu comporte donc un buffer externe de processus extérieurs *en attente* d'un jeton pour entrer dans le système et un buffer de jetons libres en attente de clients externes. Le réseau correspondant est représenté à la figure 5.11.

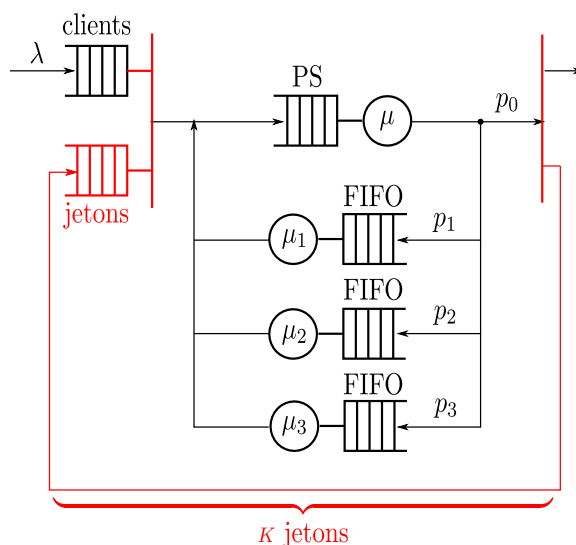


FIGURE 5.11 – Système informatique a degré de multiprogramation représenté à l'aide d'un mécanisme de type sémaphore (jetons).

Afin de déterminer les performance du système, il est fréquent de l'étudier à saturation, c'est-à-dire lorsqu'il y a en permanence des processus en attente sur le bufer extérieur. Cette analyse permet de déterminer la capacité de traitement du système, c'est-à-dire le nombre maximum de processus peuvent sortir du système par unité de temps (i.e. le *débit maximum* du système). Supposer, dans le modèle à multiprogramation de degré K , qu'il y a toujours un processus dans le buffer extérieur, revient en fait à considérer le réseau fermé contenant K processus qui est représenté à la figure 5.12.

De nombreux autres modèles de systèmes informatiques ont été proposés, à des fins de conception ou d'analyse de performance. Les systèmes multiprocesseurs en particulier ont fait l'objet de nombreux travaux consacré à leur représentation comme des réseaux de files d'attente et à l'analyse de performance qui pouvait être faite sur ces représentations. Le lecteur pourra se référer par exemple à [Bolch et al., 2006, pp. 704-707]¹ pour une présentation détaillée des systèmes multiprocesseurs fortement ou faiblement couplés. De modèles par réseaux multi-classes de files d'attente ont également été utilisés pour étudier les performances de noyaux Unix (voir [Bolch et al., 2006, pp. 724-733]).

1. G. Bolch; S. Greiner, H. de Meer and K. Trivedi, *Queuing networks and Markov chains*, Wiley, 2006 (2nd edition)

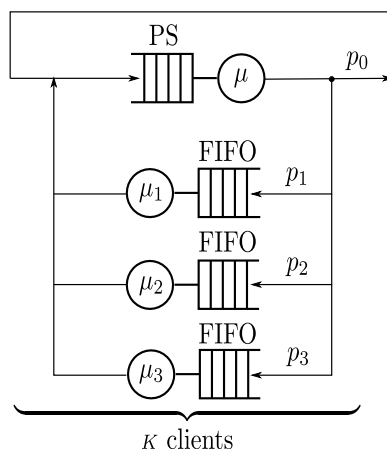


FIGURE 5.12 – Système informatique à degré de multiprogrammation fonctionnant à saturation avec K processus.

5.2.3 réseaux de communication

Un autre domaine d'application de la théorie des files d'attente est la modélisation, l'analyse de performance et la conception des réseaux de communication. Nous considérerons ici un réseau de transmission à *commutation de paquets*, constitué d'un ensemble de noeuds de commutations (*commutateurs*), reliés par un ensemble de liaisons (*canaux de transmission*). Des paquets sont générés par un *émetteur* avec un taux λ . Ces paquets doivent transiter à travers le réseau afin d'atteindre un noeud de celui-ci désigné comme le *récepteur*. Le cheminement des paquets à travers le réseau – en général plusieurs chemins ou *routes* sont possibles – peut être déterminé à priori (routage statique) ou en temps réel (routage dynamique). Dans un mécanisme de routage statique, chaque noeud du réseau est en charge de transmettre les paquets qui lui parviennent vers un noeud destination, choisi indépendamment de l'état courant du réseau (notamment de la charge des noeuds qui lui sont connectés). Localement, le routage statique peut néanmoins être probabiliste, mais dans ce cas les probabilités de routage à travers les différents canaux possibles sont indépendantes de l'état du système. Dans le cas du routage dynamique, le noeud de commutation décidera à l'arrivée d'un paquet, en fonction de la destination de celui-ci et de la perception qu'il a de l'état du réseau (de la charge de ses voisins, par exemple) à cet instant, sur quelle liaison et vers quel noeud destination il enverra ce paquet. Dans ce cas, les probabilités de routage sur les différentes liaisons dépendent – en général – de l'état du système, notamment du nombre de clients dans les différentes files des noeuds voisins du réseau. La figure 5.13 le principe d'un réseau à commutation de paquets avec routage statique probabiliste. La topologie des réseaux réels est bien

sûr – en général – beaucoup plus complexe, avec de très nombreux noeuds de degrés divers (dont des hubs, notamment) et possibilités de chemins (et des cycles!). La décision de routage induit pour chaque paquet un temps

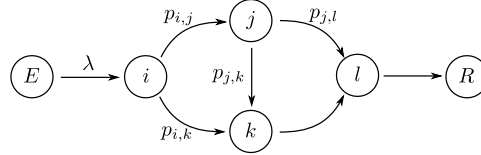


FIGURE 5.13 – représentation schématique de la topologie d'un réseau de transmission à commutation de paquets. Dans ce exemple, seulement deux routes sont possible de l'émetteur E au récepteur R : $E - i - j - l - R$ et $E - i - j - k - l - R$

de traitement au niveau du noeud de commutation. Le paquet est ensuite rangé dans la file de sortie du noeud qui correspond à la liaison choisie. Dès que la liaison est libre, le premier paquet en attente dans cette file peut être transmis, ce qui nécessite un temps d'émission auquel s'ajoute un temps de propagation sur le canal de transmission. Le modèle complet d'un noeud du réseau de commutation est représenté à la figure 5.14. Dans celui-ci le buffer en amont du serveur de traitement des paquets (qui choisit la prochaine destination) est en général avec une discipline de service de type FIFO, de même que la file d'attente en amont du serveur d'émission. Le canal de transmission lui-même est représenté par une file IS (Infinite Server) qui modélise le temps de propagation des paquets (pas d'attente dans le canal).

Dans la plupart des réseaux, les temps de traitement des paquets et les temps de propagation dans les canaux sont négligeables devant les temps d'émission des paquets². Le modèle précédent est alors simplifié en conséquence. Le seul serveur représenté est celui qui représente le temps de transmission de chaque liaison, lequel se réduit au temps d'émission du noeud aval. Le modèle simplifié comportera donc une file simple par liaison (arrête) dans la topologie (graphe) du réseau et un routage par noeud de commutation. Le buffer de chacune des files représente les paquets en attente d'émission et le serveur de ces files représente le travail de transmission (émission) de ces paquets. Les clients du système sont les paquets. Ceux-ci peuvent être de tailles variables, ce qui induit des temps de transmission variables. Cette variabilité est prise en compte par la distribution aléatoires des temps de service, le plus souvent selon une variable aléatoire exponentielle, qui peut également être différente pour chaque liaison. Les buffers des noeuds de commutation ont en pratique une capacité limitée. Un noeud ne commencera donc à émettre

2. Dans les réseaux ATM, cependant, les temps de traitement, d'émission et de propagation sont du même ordre de grandeur et doivent tous être pris en compte dans la modélisation

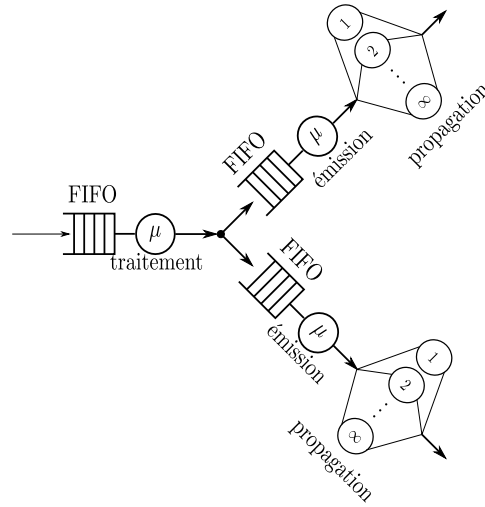


FIGURE 5.14 – modèle d'un noeud de commutation (et des canaux de transmission aval) dans un réseau de transmission à commutation de paquets

que si il est sûr qu'il reste une place dans le buffer de réception du noeud aval. Pour tenir compte de cette limitation, des files à capacité limitée sont considérées et un mécanisme de blocage avant service est ajouté à la discipline de service. La représentation simplifiée d'un réseau à commutation de paquets est donnée à la figure 5.15, pour l'exemple précédent du réseau dont la topologie est donnée à la figure 5.13. Les réseaux peuvent également être

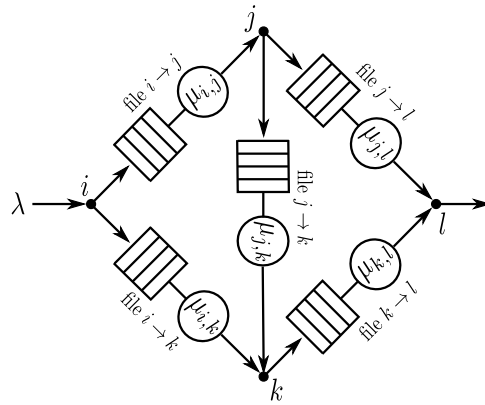


FIGURE 5.15 – modèle par réseau de files d'attente du réseau de transmission à commutation de paquets associé à la topologie de la figure 5.13

modélisé à plus haut niveau. On trouvera par exemple dans [Bolch et al.,

2006, pp. 709-720]³ la modélisation par réseau de files d'attente et l'analyse de performance pour un réseau d'entreprise typique où les clients considérés sont les requêtes que les différentes machines s'adressent entre elles. Les réseaux de files d'attente sont également utilisés pour modéliser des architectures clients-serveurs dans lesquels les processus du client et du serveurs s'exécutent sur différentes machines qui sont connectées entre elles à travers un réseau local (voir [Bolch et al., 2006, pp. 707-709]).

5.2.4 systèmes de production flexibles

Les systèmes de productions complexes de l'industrie moderne sont aujourd'hui souvent modélisés, puis analysés et optimisés, à l'aide de logiciels intégrant modèles par réseaux de files d'attente et de méthodes de simulation et d'optimisation numériques. Nous allons examiner à titre d'illustrations deux exemples de ce type de systèmes de production flexibles pour comprendre comment mettre en oeuvre des modèles par réseaux ouverts ou fermés de files d'attente pour de tels procédés.

réseaux ouverts

Considérons l'exemple d'un système de production mécanique composé des sous-systèmes (stations) suivants :

- une station qui charge les pièces à usiner sur les palettes (Lo, pour *load*)
- deux tours d'usinage (La, pour *lathes*)
- trois fraiseuses identiques (M, pour *milling machines*)
- un système de convoyage qui effectue le transfert des pièces (chargées sur les palettes) entre les stations, composé de deux véhicules autonomes (T, pour *transfer*)
- une station qui décharge les pièces usinées, en fin de production (U, pour *unload*)

Les machines identiques sont modélisées comme des files à plusieurs serveurs avec un buffer d'entrée unique et une discipline de services FCFS. On peut alors proposer pour l'atelier le modèle par réseau ouvert de files d'attente représenté à la figure 5.16. Nous supposons que ce système de production peut produire différents types de pièces et pour cette raison les machines sont utilisées de différentes façons et dans différents ordres. Supposons, par exemple, que 60% des pièces qui quitte les fraiseuses sont des pièces finies, alors que les 40% de pièces restantes doivent être traitées par les tours. On définira alors des probabilités de transfert $q_{M,La} = 0.4$, $q_{M,U} = 0.6$ et $q_{M,M} = 0$. On peut supposer par exemple que les pièces qui entrent pour la première fois dans le système de production doivent passer dans des

3. G. Bolch; S. Greiner, H. de Meer and K. Trivedi, *Queuing networks and Markov chains*, Wiley, 2006 (2nd edition)

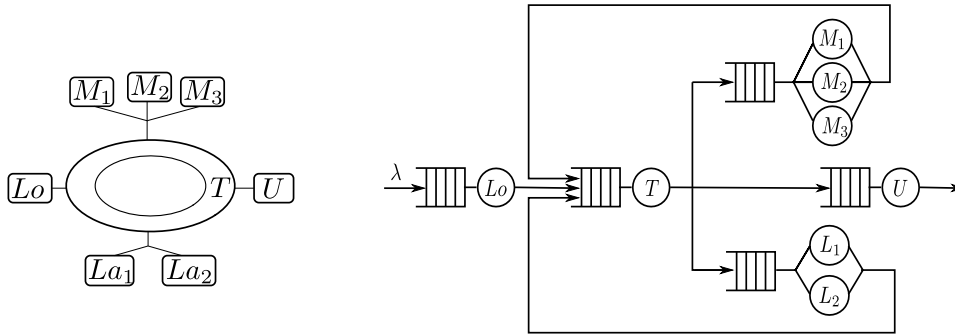


FIGURE 5.16 – Un système de production flexible (à gauche) et son modèle par réseau de files d'attente (à droite)

proportions égales soit par une fraiseuse, soit par un tour. On choisira alors $q_{Lo,La} = 0.5$, $q_{Lo,U} = 0$ et $q_{Lo,M} = 0.5$. Enfin, nous considérons, toujours à titre d'exemple, que 70% des pièces qui sortent d'un tour doivent passer par une fraiseuse et que les 30% de pièces restantes sont des pièces finies. On a alors $q_{La,La} = 0$, $q_{La,U} = 0.3$ et $q_{La,M} = 0.7$. A partir de ces probabilités de transfert entre les stations que nous venons de poser, il faut calculer les probabilités de routage correspondantes. Celles-ci dépendent également des taux d'arrivées λ_i pour $i \in \{Lo, M, La\}$. On a par conservation des pièces :

$$\lambda_T = \lambda_{Lo} + \lambda_M + \lambda_{La}$$

On peut donc définir les probabilités de routage, à partir des probabilités de transfert, en utilisant la formule des probabilités totales. On a dans ce cas :

$$\begin{aligned} p_{T,M} &:= \frac{\lambda_{Lo}}{\lambda_T} q_{Lo,M} + \frac{\lambda_{La}}{\lambda_T} q_{La,M} \\ p_{T,La} &:= \frac{\lambda_{Lo}}{\lambda_T} q_{Lo,La} + \frac{\lambda_M}{\lambda_T} q_{M,La} \\ p_{T,U} &:= \frac{\lambda_{La}}{\lambda_T} q_{La,U} + \frac{\lambda_M}{\lambda_T} q_{M,U} \end{aligned}$$

Les taux d'arrivées λ_i pour $i \in \{Lo, M, La\}$ seront quant à eux calculés en supposant le régime stationnaire établi, à l'aide des équations d'équilibre développées dans ce chapitre. Ce type de modèle permet de calculer les paramètres de performance du système. On pourra par exemple décider, si la file moyenne du système de convoyage est exagérément longue (et ralentit le système de production) d'ajouter un véhicule de convoyage. D'autres grandeurs intéressantes peuvent être par exemple le nombre moyen de pièces en production dans le système ou le temps moyen réponse du système (temps mis par une pièce pour le traverser). Même un modèle aussi simple que celui présenté permet d'analyser le fonctionnement et de décider, le cas échéant,

d'actions correctives (par exemple ajouter une machine d'un type particulier ou un véhicule, en cas de un goulot d'étranglement).

Remarque 13 Dans la pratique, le stock avant chaque machine est de capacité finie. Le nombre de pièces en attente devant une machine est donc limité et la file correspondante est à capacité finie. La discipline de service des stations dans les systèmes de production comprend en général un mécanisme de *blocage après service*. Il est généralement inutile d'attendre qu'une place se libère dans le stock aval d'une machine pour que celle-ci commence son travail. Si à l'issue de celui-ci, la pièce ne peut être transférée (le stock aval étant toujours plein), la machine arrêtera son travail en attendant qu'une place se libère dans le stock aval.

réseaux fermés

Dans la pratique, les pièces sont la plupart du temps transportées par un convoyeur qui comporte un nombre déterminé de palettes. Une pièce brute est fixée sur une palette à l'entrée du système (dès qu'il y en a une de disponible) et y restera attachée pendant toute la durée du processus de production, lors de du traitement par les différentes machines, jusqu'à ce qu'elle soit finalement déchargée dans le stock de sortie de l'atelier. A ce moment, la palette est libérée et redevient disponible pour une autre pièce brute en entrée. Il est bien sûr possible de représenter les palettes comme les jetons d'un mécanisme de sémaphore, limitant en permanence le nombre de pièce en traitement dans l'atelier. Si – ce qui est souvent le cas – on peut supposer que les pièces brutes sont toujours disponibles en nombre suffisant à l'entrée de l'atelier, dès qu'une palette se libère, elle est instantanément rechargée d'une pièce brute et remise en circulation. On sera alors dans le cas d'un réseau fermé contenant un nombre constant de clients (égal au nombre de palettes). On aboutit ainsi, pour l'atelier de la section précédente (mais avec un nombre fixé N de palettes), au modèle fermé représenté à la figure 5.17. Pour être réaliste, il faudra de plus – dans certains cas – remplacer les files simples de ce réseau par des files à capacités finies, avec un mécanisme de blocage après service.

Nous allons maintenant considérer un exemple de système de production modélisé naturellement par un réseau fermé multi-classe. Il s'agit d'un système de production similaire à celui de la section précédente, mais où les différents types de produits qui peuvent être usinés dans l'atelier sont représentés explicitement comme autant de classes de clients différentes, chacune avec un parcours spécifique dans l'atelier. Dans l'exemple que nous allons considérer, ce système de production est composé des sous-systèmes (stations) suivants :

- deux stations qui chargent et déchargent les pièces à usiner sur les palettes (LU)

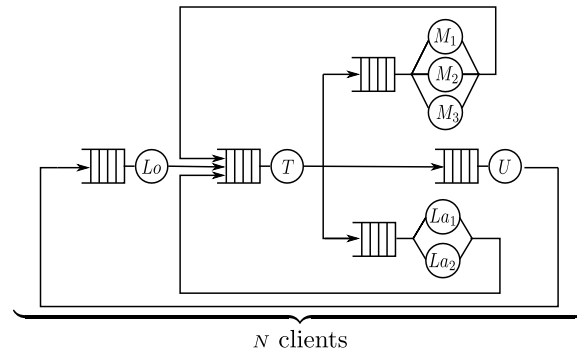


FIGURE 5.17 – le modèle par un réseau fermé de files d’attente du système de production précédent, dans le cas où la population totale de pièces en circulation dans l’atelier est limitée par le nombre total N de palettes

- deux tours d’usinage munis d’outils identiques (La)
- trois machines-outils distinctes (équipées d’outils distincts) (M)
- un système de transport central composé de huit convoyeurs (T)

On suppose que trois types de produits différents peuvent être fabriqués. Chaque type de produit (classe de client) doit circuler sur une palette qui lui est spécifique. On suppose qu’il y a K_i palettes disponibles en tout pour les clients de type i (les trois classes seront donc fermées). Nous supposons $K_1 = 4$, $K_2 = 1$ et $K_3 = 2$. Il y a ainsi moins de palettes que de convoyeurs et le système de transport peut être représenté par une file IS (Infinie Servers). Les pièces de la première classe ne nécessitent après un chargement dans la station LU , qu’un seul passage, soit sur un des deux tours, soit sur la machine outil M_3 , avant d’être déchargée par la station LU dans le stock de pièces produites (de classe 1). Les pièces de classe 2 et 3 suivent un parcours similaire mais doivent nécessairement passer respectivement par les machines M_2 et M_3 avant d’être déchargées (et pas par un des tours). Le modèle de cet atelier est représenté à la figure 5.18. On notera en particulier la caractère fermé des trois classes. La moitié des clients qui sortent du serveur LU sont des produits finis qui libèrent leur palette en sortant et sont aussitôt remplacés par des pièces brutes à usiner.

Le lecteur intéressé trouvera dans [Bolch et al., 2006, pp. 745-759]⁴ les détails sur la modélisation de ces exemples de systèmes de production, ainsi que les calculs et analyses des paramètres de performance correspondants.

4. G. Bolch; S. Greiner, H. de Meer and K. Trivedi, *Queuing networks and Markov chains*, Wiley, 2006 (2nd edition)

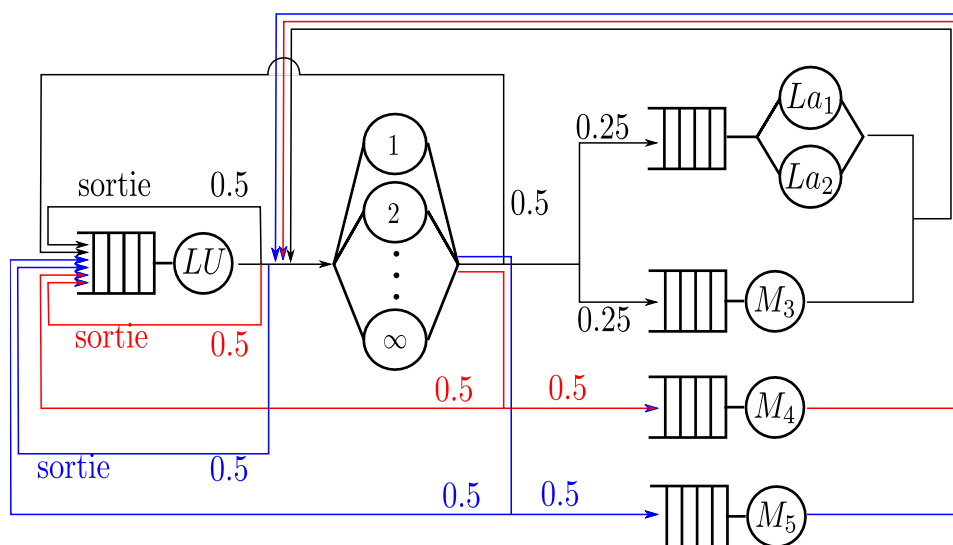


FIGURE 5.18 – un modèle par réseau fermé multi-classe. Les trois classes sont représentées par les couleurs noire, rouge et bleue. Elles sont toutes les trois fermées

5.3 Paramètres de performance

Nous allons dans cette section définir les *paramètres de performance* qui sont les plus utilisés, notamment dans l'analyse du régime permanent des files simples et des réseaux de files d'attente. L'existence de ces paramètres de performance est conditionnée à la *stabilité* des files étudiées (essentiellement, le fait que le nombre de clients dans ces files n'explose pas). Nous énoncerons certains résultats généraux de stabilité et reporterons à la section 5.4 la preuve détaillée de ces résultats de stabilités pour les différents type de files simples. Enfin, dans la troisième partie de cette section, nous établirons la *Loi de Little* qui relie entre eux les différents paramètres de performance⁵ et permet ainsi d'en déduire certains à partir des autres.

5.3.1 Définition des paramètres opérationnels

Nous allons analyser le comportement d'un système quelconque. Ce système est une boîte noire. Il peut aussi bien être constitué d'une seule file simple, que d'un réseau de files à la topologie complexe avec un routage sophistiqué. Les paramètres opérationnels que nous allons définir s'appliqueront dans tous les cas. Le système *boîte noire* considéré et certains de ces paramètres opérationnels sont représenté à la figure 5.19. On s'intéresse

5. aussi bien les paramètres de performance de files simples, que ceux de réseaux complexes de files simples, indépendamment du type des files considérées

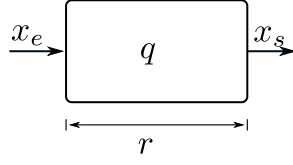


FIGURE 5.19 – *boîte noire* représentant une file simple ou un réseau complexe de files et certains de ses paramètres opérationnels

à une *réalisation particulière* de l'évolution de ce processus stochastique et en particulier de l'évolution du nombre de client $n(t)$ dans le système pour $t \in [0, T]$. T est l'horizon ou la durée totale de l'observation du système. Dans la mesure où nous nous intéressons à une réalisation particulière du processus stochastique, les valeurs considérées seront notées avec des lettres minuscules⁶. Une telle réalisation est représentée à la figure 5.20. L'instant d'arrivée du client k est noté a_k , alors que son instant de départ est noté d_k et sa durée de séjour totale dans le système est notée $r_k := d_k - a_k$. On définira également les autres paramètres opérationnels suivant :

- $t(n, T)$, comme le temps total pendant lequel le système contient n clients exactement
- $p(n, T)$, comme la proportion du temps pendant laquelle le système contient n clients, c'est-à-dire

$$p(n, T) := \frac{t(n, T)}{T}$$

- $a(T)$, comme le nombre total de clients arrivés sur la période $[0, T]$
- $d(T)$, comme le nombre total de clients partis sur la période $[0, T]$

Dans l'exemple de la figure 5.20, on a en particulier $a(T) = 6$ et $d(T) = 5$. À partir de ces paramètres, il est possible de définir des *paramètres de performance opérationnels* en régime transitoire, pour l'intervalle de temps $[0, T]$. On définira en particulier :

- le *débit moyen d'entrée* - il s'agit du nombre moyen de clients arrivés dans le système par unité de temps

$$x_e(T) := \frac{a(T)}{T}$$

- le *débit moyen de sortie* - il s'agit du nombre moyen de clients ayant quitté le système par unité de temps

$$x_s(T) := \frac{d(T)}{T}$$

6. Il s'agira de constantes ou de fonctions du temps au sens déterministe. Nous verrons dans la section 5.3.2 que les paramètres opérationnels qui en découlent sont identiques aux paramètres de performance stochastiques – liés aux distributions stationnaires des variables aléatoires correspondantes – quand le processus observé est *ergodique*

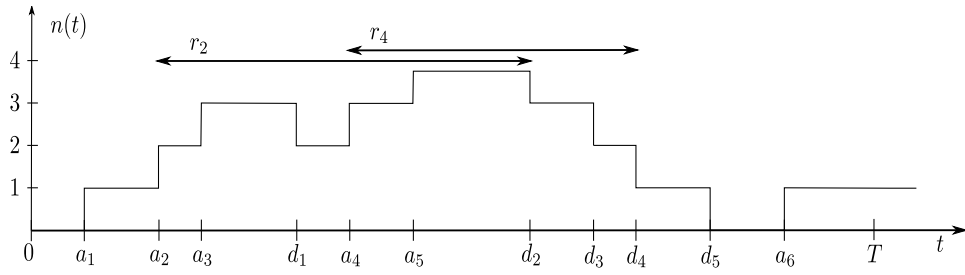


FIGURE 5.20 – Un exemple de réalisation particulière sur l'intervalle d'observation $[0, T]$ des arrivées et départs de clients dans un système

- le *nombre moyen de clients* - il s'agit de la moyenne temporelle de $n(t)$ sur la période $[0, T]$

$$q(T) := \frac{1}{T} \sum_{n \geq 0} n t(n, T) = \sum_{n \geq 0} n p(n, T)$$

- le *temps moyen de séjour* (aussi appelé *temps moyen de réponse*) d'un client - il s'agit de la moyenne arithmétique des temps de séjour des clients arrivés dans le système sur la période $[0, T]$

$$r(T) := \frac{1}{a(T)} \sum_{k=1}^{a(T)} r_k$$

Dans le cas d'une file simple, avec un seul serveur, on peut également définir le *taux d'utilisation* de ce serveur :

$$u(T) := \sum_{n \geq 1} p(n, T) = 1 - p(0, T)$$

Ces définitions s'appliquent quel que soit le système considéré : file simple ou réseau de files d'attente. Dans le cas d'un réseau de files d'attente, il s'agira des paramètres opérationnels de performance de l'ensemble du réseau (on peut bien sûr définir les paramètres opérationnels de performance de chacune des files qui le constituent). Ces paramètres opérationnels globaux n'ont d'intérêt que si le réseau est ouvert. En effet, dans le cas d'un réseau fermé, le nombre N de clients est constant. On a alors, quel que soit $T > 0$:

$$\begin{aligned} x_e(T) &= x_s(T) \\ q(T) &:= N \\ r(T) &:= T \end{aligned}$$

Dans le cas d'un réseau multi-classe, on pourra s'intéresser aux paramètres opérationnels de performance par classe ou toutes classes confondues.

Les quantités précédentes caractérisent le régime transitoire. Toutefois, dans l'analyse des performances opérationnelles d'un système donné – que ce soit à des fins de conception ou d'optimisation – on s'intéresse le plus souvent au fonctionnement sur le long terme (ou encore en *régime permanent*). On définit les paramètres opérationnels de performance en régime permanent de la manière suivante :

$$\begin{aligned} x_e &:= \lim_{T \rightarrow \infty} x_e(T) \\ x_s &:= \lim_{T \rightarrow \infty} x_s(T) \\ q &:= \lim_{T \rightarrow \infty} q(T) \\ r &:= \lim_{T \rightarrow \infty} r(T) \\ u &:= \lim_{T \rightarrow \infty} u(T) \end{aligned}$$

Encore faut-il que ces limites existent ! C'est ce qui motive notre étude de la stabilité des files et réseaux de files d'attente ...

5.3.2 Stabilité et ergodicité

stabilité

La stabilité est une caractéristique du comportement du système stochastique *en régime permanent* (lorsque $T \rightarrow \infty$). On s'intéresse en particulier au nombre de clients dans le système et au fait que ce nombre ne croît pas indéfiniment sur des horizons de temps long (instabilité). Ainsi, les systèmes fermés, ou ceux à contrainte de population, seront stables *par construction*, leur nombre de clients étant constant. Nous ne nous intéressons qu'à des processus sans destruction ni création de clients : nous excluons donc tous les mécanismes par lesquels les clients peuvent être décomposés en sous-clients (mécanismes de type *fork*) ou agrégés en super-clients (mécanisme de type *join*). *L'analyse de la stabilité n'aura donc de sens que pour des files et réseaux ouverts, en régime permanent, sans perte ni création de client au sein du système.*

Définition 21 (stabilité) Un système (i.e. une file simple, une partie d'une file simple ou un réseau de files d'attente) est dit *stable* si le débit asymptotique moyen de sortie des clients du système est égal au débit asymptotique moyen d'entrée des clients dans le système, c'est-à-dire si :

$$x_s := \lim_{T \rightarrow \infty} x_s(T) = \lim_{T \rightarrow \infty} x_e(T) =: x_e$$

Cela implique en particulier que *les deux limites existent*.

D'après les relations précédentes, on a donc nécessairement dans ce cas

$$\frac{x_s(T)}{x_e(T)} = \frac{d(T)}{a(T)} \xrightarrow{T \rightarrow \infty} 1$$

Sur le long terme, *les sorties et entrées du système doivent être équilibrées en nombre*. En particulier, cela entraîne que le nombre de clients dans le système, $n(t)$, reste fini pour tout $t > 0$.

Nous considérons dans un premier temps le cas d'une file simple "générale" $G/G/1$ où les temps inter-arrivées et les temps de services sont distribués selon deux lois quelconques, d'espérances respectives $t_a := \lambda^{-1}$ (temps moyen entre deux arrivées) et $t_s := \mu^{-1}$ (temps moyen de service). Les paramètres λ et μ désignent donc respectivement le taux moyen d'arrivée des clients et le taux moyen de service. On peut montrer pour cette file *générale* le résultat suivant (voir la section 5.4 sur l'étude du régime stationnaire des files simples).

Proposition 22 (stabilité des files simples à un serveur)

Une file $G/G/1$, avec un taux d'arrivée λ et un taux de service μ , est *stable* si $\lambda < \mu$

Ce résultat s'interprète facilement. Si $\lambda < \mu$, le temps moyen t_a entre deux arrivées est supérieur au temps moyen de service d'un client t_s . Sur le long terme (i.e. pour $T \rightarrow \infty$), le nombre moyen de départs $d(T)$ va coïncider avec le nombre moyen d'arrivées $a(T)$, puisque le serveur est capable de servir tous les clients qui se présentent dans la file. Lorsque l'horizon de temps T augmente, les deux grandeurs $a(T)$ et $d(T)$ sont de plus en plus grandes. A mesure que T augmente, $d(T)$ s'écarte donc relativement (en proportion) de moins en moins de $a(T)$. On obtient donc à la limite :

$$\lim_{T \rightarrow \infty} \frac{d(T)}{a(T)} = \lim_{T \rightarrow \infty} \frac{x_s(T)}{x_e(T)} = 1$$

c'est-à-dire la stabilité $x_e = x_s$. Au contraire, si $\lambda > \mu$, le temps moyen t_a entre deux arrivées est inférieur au temps moyen de service d'un client t_s . Les clients s'accumulent dans la file et le serveur ne parvient jamais à rattrapper son retard. Ce retard s'accroît nécessairement avec le temps. Le nombre de clients dans le système $n(t)$ est non borné.

Le cas limite $\lambda = \mu$ est intéressant. Pour celui-ci, les comportements des files déterministes et stochastiques sont distincts. Dans le cas d'une file déterministe $D/D/1$, les temps de service $t_s = \mu^{-1}$ et les temps entre les arrivées $t_a = \lambda^{-1}$ sont constants. La figure 5.21 illustre le cas $\lambda < \mu$ où le nombre $n(t)$ de clients dans la file vaudra successivement 0 ou 1, les départs et les arrivées se succédant de manière périodique. Dans le cas limite $\lambda = \mu$, les arrivées coïncident avec les départs. Le nombre de clients dans le système sera constant et égal à 1. La file $D/D/1$ est donc stable. Ce n'est par contre pas le cas des files stochastiques. Nous verrons dans l'étude des files simples en régime stationnaire (section 5.4) que la file $M/M/1$, par exemple, est instable pour le cas limite $\lambda = \mu$. Cette instabilité peut être interprétée de

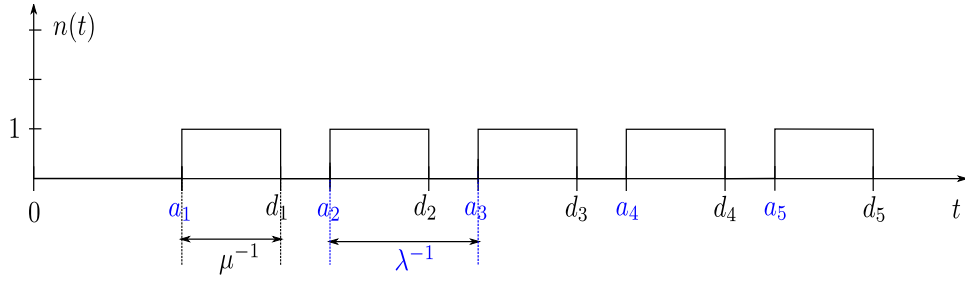


FIGURE 5.21 – Le nombre $n(t)$ de clients au cours du temps pour une file déterministe $D/D/1$ dans le cas stable $\lambda < \mu$.

la manière suivante. Dans la file $M/M/1$, dans le cas $\lambda = \mu$, le temps entre deux arrivées sera une fois sur deux supérieur au temps de service. Dans le cas où il n'y a pas de client dans le buffer, le serveur devra alors attendre l'arrivée du client suivant. Il ne sera donc pas utilisé au maximum de ses capacités. Sur des horizons de temps longs, ce défaut dans le taux d'utilisation du serveur va empêcher celui-ci de traiter tous les clients qui arrivent et de résorber la file qui va croître progressivement.

Considérons maintenant le cas où la file simple contient C serveurs identiques, ils peuvent tous contribuer à servir les clients qui arrivent ou qui attendent dans le buffer. Le taux de service maximal de la file est dans ce cas $C\mu$ (si $n(t) \geq C$, tous les serveurs sont supposés occupés) avec μ^{-1} , le temps moyen de service de chacun des serveurs. La file sera donc stable si le taux moyen d'arrivée λ reste inférieur à ce taux maximal de service $C\mu$, atteint dès qu'il y a au moins C clients dans la file.

Proposition 23 (stabilité des files simples à C serveurs)

Une file $G/G/C$ avec un taux d'arrivée λ et C serveurs, chacun avec un taux de service individuel μ , est stable si $\lambda < C\mu$

La file $D/D/C$ sera stable pour $\lambda = C\mu$, alors que la file stochastique $G/G/C$ sera en général instable dans ce même cas particulier.

Dans le cas d'un réseau de files d'attente, ouvert et monoclasse, la stabilité du réseau est équivalente à la stabilité de chacune des files. Si le nombre de clients dans une des files augmente indéfiniment, alors le nombre de clients dans le réseau aussi. Il s'agit donc en général de vérifier les conditions de stabilité précédentes pour chacune des files constitutives du réseau. Ceci implique en particulier de calculer au préalable les taux d'arrivées λ_i pour toutes les files $i \in \{1, \dots, M\}$ qui constituent ce réseau. Le calcul des taux λ_i , en régime stationnaire, repose sur les conditions d'équilibre en chacun des noeuds du réseau (voir section ??).

Proposition 24 (stabilité des réseaux ouverts monoclasse)

Un réseau ouvert de files d'attente monoclasse comportant M files simples (stations), chacune avec un taux de service μ_i , C_i serveurs et un taux d'arrivée λ_i est stable si $\lambda_i < C_i \mu_i$ pour tout $i \in \{1, \dots, M\}$

Remarque 14 (stabilité et systèmes limités) un système limité, que ce soit un système *fermé* ou un système à *contrainte de population*, est automatiquement *stable*. Dans le cas d'un système fermé, le nombre de clients dans le système est constant. Dans le cas d'un réseau à contrainte de population, le nombre moyen de clients entrant (autorisés à entrer) par unité de temps (débit d'entrée) est toujours inférieur ou égal au débit moyen de sortie, dès que le nombre de clients dans le système atteint la limite de population. Sur le long terme (i.e. en régime stationnaire), les débits entrant et sortant sont donc nécessairement égaux entre eux. Ils sont égaux soit au débit maximum que peut traiter le réseau, soit au débit moyen d'entrée en régime permanent, selon les situations.

ergodicité

Nous avons jusqu'ici étudié les paramètres opérationnels de réalisations particulières, sur des intervalles de temps $[0, T]$, des processus stochastiques considérés (files ou réseaux). Nous avons examiné différentes grandeurs opérationnelles dans ces réalisations et leur comportement asymptotique pour $T \rightarrow \infty$. Nous nous sommes en quelque sorte intéressé au régime permanent d'évolutions particulières du système. Une question se pose donc naturellement : toutes les réalisations particulières ont-elles *le même comportement asymptotique* ? De manière équivalente, les paramètres opérationnels de performance considérés ont-ils tous la *même limite* ? La réponse à cette question est liée à l'*ergodicité* du processus stochastique.

Pour répondre à cette question, il faut considérer le caractère aléatoire des processus et des grandeurs qui nous intéressent. Les instants d'arrivée a_k , par exemple, sont des réalisations particulières d'une variable aléatoire A_k , définie comme le temps d'arrivée du client k ⁷. De même, nous pouvons définir, pour les systèmes considérés (files simples ou réseaux de files d'attente), les variables aléatoires D_k (instant de départ du client k), $R_k := D_k - A_k$ (temps de séjour du client k), $A(T)$ (nombre de clients arrivés sur $[0, T]$), $D(T)$ (nombre de clients partis sur $[0, T]$), $N(t)$ (nombre de clients dans le système à l'instant $t > 0$) ou encore $P(n, t) := P(N(t) = n)$ (probabilité qu'il y ait n clients à l'instant t).

7. A_k est par exemple distribuée selon une loi $\Gamma(\lambda, k)$ dans le cas d'un processus stochastique d'arrivée de type Poisson, de paramètre λ

A l'aide de ces variables aléatoires, il est possible de définir ensuite les *paramètres de performance stochastiques* du système, sur un horizon de temps $[0, T]$, de la même manière que nous avons défini les paramètres de performance opérationnels :

- le *débit moyen d'entrée* $X_e(T) := \frac{A(T)}{T}$
- le *débit moyen de sortie* $X_s(T) := \frac{D(T)}{T}$
- le *nombre moyen de clients* $Q(T) := \sum_{n \geq 0} n P(n, T)$
- le *temps moyen de séjour* $R(T) := \frac{1}{A(T)} \sum_{k=1}^{A(T)} R_k$
- le *taux d'utilisation du serveur* $U(T) := \sum_{n \geq 1} P(n, T) = 1 - P(0, T)$

On s'intéressera en particulier au comportement asymptotique de ces paramètres :

$$\begin{aligned} X_e &:= \lim_{T \rightarrow \infty} X_e(T) \\ X_s &:= \lim_{T \rightarrow \infty} X_s(T) \\ Q &:= \lim_{T \rightarrow \infty} Q(T) \\ R &:= \lim_{T \rightarrow \infty} R(T) \\ U &:= \lim_{T \rightarrow \infty} U(T) \end{aligned}$$

Nous allons établir les valeurs de ces paramètres de performance stochastiques en étudiant les processus stochastiques (arrivées, services, etc.) de la file simple ou du réseau considéré. Nous pourrions par exemple représenter le système étudié par une chaîne de Markov à temps continu, puis calculer la distribution limite des probabilités d'état, à partir de laquelle nous pourrions déduire ces paramètres de performance stochastiques. Cependant, dans la plupart des situations, ce sont bien les paramètres opérationnels qui nous intéressent. Par exemple, nous nous intéresserons davantage – en général – à la proportion du temps où il n'y a aucun client dans la file (taux d'inutilisation du serveur), qu'à la probabilité qu'à un instant donné il n'y ait aucun client dans la file. Cette dernière probabilité ne reflète en effet pas fonctionnement moyen du système sur le long terme. Notons aussi que lorsque nous faisons des simulations numériques, c'est bien des réalisations particulières que nous simulons. L'ergodicité va nous permettre d'unifier les paramètres de performance asymptotiques opérationnels et stochastiques.

Définition 22 (processus ergodique) Soit $N(t)$ une chaîne stochastique à temps continu et $n(t)$ une réalisation particulière quelconque de ce processus. A cette réalisation particulière, on peut associer les moyennes temporelles

$$\overline{n^k(T)} := \frac{1}{T} \int_0^T n^k(t) dt$$

Au processus stochastique $N(t)$, on peut associer des probabilités d'état

$P(n, T) := P(N(T) = n)$ et calculer les moments d'ordre k :

$$E\left(N^k(T)\right) := \sum_{n=0}^{\infty} n^k P(n, T)$$

Le processus $N(t)$ est dit *ergodique* si ces moments sont *asymptotiquement égaux*, c'est-à-dire si

$$\lim_{T \rightarrow \infty} \sum_{n=0}^{\infty} n^k P(n, T) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T n^k(t) dt, \forall k \geq 1$$

En particulier, la limite existe et ne dépend pas de la réalisation particulière choisie

L'espérance $E(N(T))$ peut être interprétée comme une moyenne d'ensemble. La répétition, sur un même intervalle de temps $[0, T]$ de l'observation de réalisations particulières du processus et la moyenne de toutes les observations $N(T)$ réalisées, conduiront en effet à la distribution de probabilité $P(n, T)$ (par la loi des grands nombres). Pour un processus ergodique, cette espérance est donc en particulier égale à la moyenne temporelle observée pour n'importe quelle réalisation $x(t)$ du processus sur un horizon infini. Dans un système ergodique, nous pourrions donc confondre – en régime permanent – les paramètres de performances opérationnels et les paramètres de performance stochastiques.

Dans la suite de ce chapitre, tous les systèmes rencontrés seront ergodiques. Ils sont associés à des chaînes de Markov à temps continu pour lesquels la distribution stationnaire de probabilité existe et est égale à la distribution limite. Dans ce cas la définition 22 est nécessairement satisfaite. Tous les systèmes ne sont bien sûr pas ergodiques. Une chaîne non irréductible, par exemple, est un système non ergodique. Une trajectoire particulière pourra amener la système dans une sous-chaîne absorbante 1 et une autre trajectoire particulière pourra le conduire dans une autre sous-chaîne absorbante 2. Les comportements asymptotiques des deux réalisations particulières n'auront alors rien de commun. Une CMTD périodique est un autre exemple de système (à temps discret) non ergodique. Un système instable ne peut pas non plus être ergodique car les limites de la définition 22 n'existent pas ! Nous verrons que dans de nombreux exemples de files simples (représentées par des CMTC irréductibles), la condition d'ergodicité est équivalente à la condition de stabilité.

5.3.3 Loi de Little

La loi de Little est une relation très générale, qui s'applique aux systèmes *stables* et qui lie entre eux les paramètres asymptotique de performance

représentés à la figure 5.22. Aucune hypothèse spécifique sur les variables aléatoires qui caractérisent le système (temps inter-arrivées, temps de service), sur les disciplines de service ou sur la topologie et le routage (dans le cas de réseaux) n'est nécessaire pour établir la loi de Little. On suppose seulement que le système étudié – vu comme une boîte noire – est ouvert, stable et en régime stationnaire. Par la définition de la stabilité, les débit entrants et sortants sont égaux. On note

$$X := X_e = X_s \quad (5.5)$$

ce débit moyen asymptotique.

Proposition 25 (Loi de Little)

Le nombre moyen de clients Q , le temps moyen de réponse R et le débit moyen X d'un système stable en régime permanent satisfont la relation :

$$Q = RX \quad (5.6)$$

Démonstration : considérons dans un premier temps un intervalle d'observation $[0, T]$ tel que le système est vide au début $t = 0$ et à la fin $t = T$ de la période d'observation. Une telle période est représentée à la figure 5.23. Dans ce cas, on a $D(T) = A(T)$. Le nombre moyen de clients présents $Q(T)$ et le temps de séjour moyen de ces clients $R(T)$ peuvent alors être écrits :

$$\begin{aligned} Q(T) &= \frac{1}{T} \sum_{n \geq 0} nT(n, T) \\ R(T) &= \frac{1}{A(T)} \sum_{n \geq 1}^{A(T)} R_n = \frac{1}{D(T)} \sum_{n \geq 1}^{D(T)} R_n \end{aligned} \quad (5.7)$$

Or, on peut montrer que l'aire sous la courbe $(t, N(t))$ de la figure 5.23 s'écrit :

$$\sum_{n \geq 0} nT(n, T) = \sum_{n \geq 1}^{D(T)} R_n \quad (5.8)$$

En effet, si nous examinons ce qui se passe pendant l'intervalle de durée Δ entre A_5 et D_2 , à titre d'exemple :

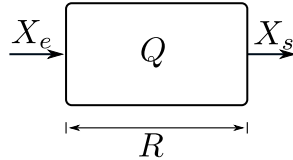
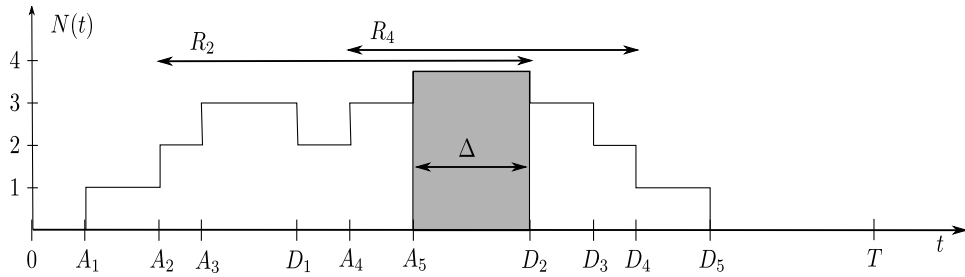


FIGURE 5.22 – Les paramètres de performance asymptotiques liés entre eux par la loi de Little. On suppose le système stable, c'est-à-dire $X_e = X_s = X$

FIGURE 5.23 – Une période d'observation $[0, T]$ avec $N(0) = 0$ et $N(T) = 0$

- les 4 clients dans la file amènent une contribution 4Δ à la somme $\sum_{n \geq 0} nT(n, T)$
- les 4 clients dans la file voient leur temps de séjour augmenter de Δ pendant cet intervalle, et donc la somme $\sum_{n \geq 1}^{D(T)} R_n$ augmenter de 4Δ

On a donc, en reportant (5.8) dans (5.7) :

$$\begin{aligned} Q(T) &= \frac{1}{T} \sum_{n \geq 1}^{D(T)} R_n \\ &= \frac{D(T)}{T} R(T) \end{aligned} \quad (5.9)$$

Or

$$X_s(T) = \frac{D(T)}{T}$$

A la limite pour $T \rightarrow \infty$, on a :

$$\begin{aligned} Q(T) &\xrightarrow{T \rightarrow \infty} Q \\ \frac{D(T)}{T} &\xrightarrow{T \rightarrow \infty} X \\ R(T) &\xrightarrow{T \rightarrow \infty} R \end{aligned} \quad (5.10)$$

et l'équation (5.9) devient

$$Q = RX$$

Nous avons montré la loi de Little pour le cas $N(0) = 0$ et $N(t) = 0$ pour $t > T$. Cependant, il convient de noter que :

- le raisonnement ci-dessus s'applique tel quel dans le cas $N(0) = N(T) \neq 0$, c'est-à-dire lorsque $D(T) = A(T)$
- la stabilité du système entraîne

$$\lim_{T \rightarrow \infty} \frac{D(T)}{A(T)} = 1$$

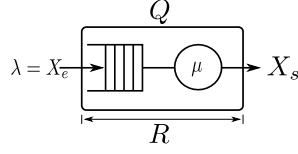


FIGURE 5.24 – Les paramètres de performance considérés dans l'application de la loi de Little à la file simple

c'est-à-dire

$$A(T) = D(T) + o(1), \text{ pour } T \rightarrow \infty$$

La différence entre $A(T)$ et $D(T)$ peut donc être négligée dans le calcul des grandeurs asymptotiques (i.e. pour $T \rightarrow \infty$).

Ceci permet d'achever la démonstration de la loi de Little

■

La loi de Little s'applique à tous les systèmes ouverts stables en régime stationnaire. Elle permet de déduire une des trois quantités Q , R ou X lorsqu'on connaît les deux autres.

Exemple 13 (Loi de Little appliquée à une file simple) Dans l'analyse d'une file simple générale (i.e. de type $G/G/1$), la loi de Little peut être appliquée séparément à l'ensemble de la file simple, ou au buffer et au serveur, pris isolément. En considérant la file simple telle que définie à la figure 5.24, on obtient par l'hypothèse de stabilité :

$$X = X_e = X_s = \lambda \quad (5.11)$$

et par la loi de Little :

$$Q = R\lambda \quad (5.12)$$

Il ne reste donc plus qu'à calculer Q ou R pour connaître les trois paramètres de performance Q , R et X . Si, d'autre part, on applique la loi de Little au buffer uniquement (voir figure 5.25), on obtient :

$$Q_W = W\lambda \quad (5.13)$$

Enfin, en appliquant la loi de Little au serveur uniquement (voir figure 5.25), on obtient :

$$Q_S = S\lambda \quad (5.14)$$

On sait par ailleurs que

$$\begin{aligned} Q_S &= \lim_{T \rightarrow \infty} \left(0 \cdot P(0, T) + 1 \cdot \sum_{n \geq 1} P(n, T) \right) \\ &= \lim_{T \rightarrow \infty} (1 - P(0, T)) \\ &= U \end{aligned} \quad (5.15)$$

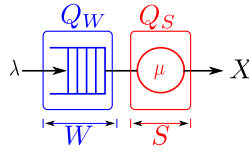


FIGURE 5.25 – Les paramètres de performance considérés dans l'application de la loi de Little au buffer et au serveur de la file simple

Le nombre moyen de client dans le serveur n'est en effet rien d'autre que le taux d'utilisation du serveur (dans le cas d'une file simple avec un seul serveur). Le temps moyen de séjour dans le serveur est le temps moyen de service de ce serveur :

$$S = \frac{1}{\mu} \quad (5.16)$$

Par ailleurs, les clients n'étant ni détruits ni créés dans la file, on peut écrire deux lois de conservation :

$$\begin{aligned} Q &= Q_W + Q_S \\ R &= W + S \end{aligned} \quad (5.17)$$

Les huit équations (5.11) à (5.17) ne sont pas indépendantes. Les lois de Little, pour la file dans son ensemble et pour le buffer et le serveur séparément, sont bien sûr compatibles avec les lois de conservation. Par exemple, à partir des lois de Little (5.13) et (5.14), pour le buffer et le serveur, et des lois de conservation (5.17), on obtient facilement la loi de Little (5.12) pour l'ensemble de la file. Nous disposons donc de sept équations indépendantes qui lient entre eux⁸ huit paramètres de performance : Q , R , X , U , Q_W , W , Q_S et S . En définitive, dans une file simple $G/G/1$, le calcul d'une seule des grandeurs Q ou R permettra de déduire toutes les autres.

Au passage, à l'aide de la loi de Little pour le serveur (5.14), ainsi que des équations générales pour les files simples mono-serveurs (5.15) et (5.16), nous avons montré que pour une file simple mono-serveur, de type $G/G/1$, le taux d'utilisation du serveur vaut :

$$U = Q_S = \frac{\lambda}{\mu} \quad (5.18)$$

8. Les taux d'arrivées λ et de service μ sont bien sûr nécessaires pour déterminer les paramètres de performance asymptotiques, mais ils sont considérés comme des paramètres exogènes, caractéristiques de la file étudiée

5.4 Analyse des files simples

Les files simples suffisent, dans de nombreux cas, à modéliser des systèmes stochastiques et permettent ainsi d'analyser leurs performances, que ce soit à des fins de compréhension, de conception, d'optimisation ou de supervision. Leurs éléments constitutifs ont été décrits dans la section 5.1.1. Dans cette section, nous allons détailler le calcul des distributions stationnaires de probabilité des modèles stochastiques de files simples, pour ensuite calculer les différents paramètres de performance qui leur sont associés. Nous examinerons en particulier le cas des files markoviennes, pour lesquelles les processus d'arrivées et de services sont sans mémoire : processus d'arrivée de Poisson et temps de service exponentiel.

5.4.1 La file M/M/1

La file simple (voir figure 5.26) est composée d'un seul serveur dont le temps de service est distribué selon une loi exponentielle $\text{Exp}(\mu)$, d'un buffer avec une discipline de service FIFO et d'un processus d'arrivée poissonnien de paramètre λ . C'est un cas particulier de *processus de naissance et de morts* (à partir d'un état $N(t) = n$, la chaîne peut seulement aller aux états $n - 1$ ou $n + 1$) dans lequel les taux de natalité, λ , et de mortalité, μ , sont constants.

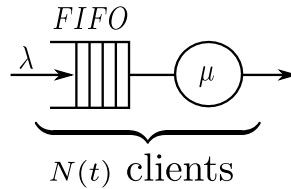


FIGURE 5.26 – Représentation schématique d'une file $M/M/1$

CMTC associée à la file $M/M/1$

Montrons dans un premier temps que les taux de transition d'un état n vers un état p sont nuls pour $p \geq n + 2$.

Une telle transition sur un intervalle de temps $[t, t + dt[$ correspond à deux arrivées au moins pendant l'intervalle de temps dt ⁹. La probabilité qu'il y ait $k \geq 2$ arrivées pendant un intervalle de temps de longueur dt , dans un processus de Poisson de paramètre λ , vaut :

$$P(N(dt) - N(0) = k) = \frac{(\lambda dt)^k}{k!} e^{-\lambda dt}$$

9. Une transition de n à $n + 2$, par exemple, pourrait également correspondre à 3 arrivées et un départ, quatre arrivées et deux départs, etc.

Ainsi, la probabilité de transition, pendant un intervalle de temps dt , entre l'état n et l'état p (avec $p \geq n + 2$) s'écrit :

$$p_{n,p}(dt) = \frac{(\lambda dt)^2}{2!} e^{-\lambda dt} \tilde{P}$$

où \tilde{P} représente la probabilité des autres évènements (en plus des deux arrivées nécessaires) pour une transition de n à p pendant l'intervalle de temps dt . On a donc pour le taux de transition correspondant :

$$\lambda_{n,p} := \lim_{dt \rightarrow 0} \frac{p_{n,p}(dt)}{dt} = 0$$

De la même manière, une transition pendant un intervalle de temps de longueur dt d'un état n ($n \geq 2$) à un autre état p avec $p \leq n - 2$ correspond à deux départs au moins pendant cet intervalle de temps. Comme les temps de service sont des exponentielles indépendantes, la probabilité d'avoir deux départs pendant l'intervalle de temps dt vaut ¹⁰ :

$$\begin{aligned} P(T_1^\mu + T_2^\mu \leq dt) &= \int_0^{dt} \mu^2 \xi e^{-\mu \xi} d\xi \\ &= -\mu dt e^{-\mu dt} + 1 - e^{-\mu dt} \\ &= -\mu dt \left(1 - (\mu dt) + \frac{(\mu dt)^2}{2!} - \dots \right) \\ &\quad + 1 - \left(1 - (\mu dt) + \frac{(\mu dt)^2}{2!} - \dots \right) \\ &= \frac{(\mu dt)^2}{2} + o(dt^2), \text{ pour } dt \rightarrow 0 \end{aligned}$$

Ainsi, la probabilité de transition, pendant un intervalle de temps dt , entre l'état n et l'état p (avec $p \leq n - 2$) s'écrit :

$$p_{n,p}(dt) = \left(\frac{(\mu dt)^2}{2} + o(dt^2) \right) \cdot \hat{P}$$

où \hat{P} représente la probabilité des autres évènements (en plus des deux départs nécessaires) pour une transition de n à p pendant l'intervalle de temps dt . On a donc pour le taux de transition correspondant :

$$\lambda_{n,p} := \lim_{dt \rightarrow 0} \frac{p_{n,p}(dt)}{dt} = 0$$

Les taux de transition d'un état n vers un état p sont donc nuls pour $p \leq n - 2$. La file simple est donc bien un exemple de processus de naissances et

10. La somme de deux exponentielles indépendantes de même paramètre μ est une loi Gamma $\Gamma(\mu, 2)$.

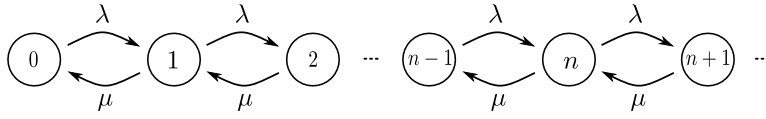


FIGURE 5.27 – La chaîne de Markov à temps continu associée à la file simple

de morts. Il nous reste, pour spécifier complètement son modèle par CMTC, à calculer les taux de transition d'un état n ($n \geq 1$) vers ces deux voisins $n-1$ et $n+1$.

Pour une transition de n vers $n+1$, il faut qu'il y ait 1 arrivée de requête et 0 départs, 2 arrivées et 1 départ, 3 arrivées et 2 départ, etc. Nous avons déjà montré que la probabilité associée à l'arrivée de deux requêtes (ou plus) pendant un intervalle de temps dt correspond à un taux de transition nul. Le seul cas à envisager, dans le calcul du taux de transition $\lambda_{n,n+1}$, est donc celui où il y a une arrivée de requête et aucun départ pendant un intervalle de temps dt . La probabilité correspondante est :

$$P(T^\mu > dt; T^\lambda \leq dt) = P(T^\mu > dt) \cdot P(T^\lambda \leq dt) = e^{-\mu dt} (1 - e^{-\lambda dt})$$

et le taux de transition $\lambda_{n,n+1}$ vaut donc :

$$\lambda_{n,n+1} := \lim_{dt \rightarrow 0} \frac{e^{-\mu dt} (1 - e^{-\lambda dt})}{dt} = \lambda$$

On trouvera de même

$$P(T^\mu \leq dt; T^\lambda > dt) = P(T^\mu \leq dt) \cdot P(T^\lambda > dt) = (1 - e^{-\mu dt}) e^{-\lambda dt}$$

et

$$\lambda_{n,n-1} := \lim_{dt \rightarrow 0} \frac{(1 - e^{-\mu dt}) e^{-\lambda dt}}{dt} = \mu$$

Nous venons d'établir que *la file simple est un processus de naissances et de morts où les taux de transition sont constants* (indépendants de l'état). La CMTC correspondante est représentée à la figure 6.20. Le *générateur infinitésimal* (infini) de la file simple s'écrit :

$$\mathbf{A} = \begin{pmatrix} -\lambda & \lambda & 0 & \dots \\ \mu & -(\lambda + \mu) & \lambda & \dots \\ 0 & \mu & -(\lambda + \mu) & \ddots \\ \vdots & \vdots & \ddots & \ddots \end{pmatrix}$$

Le temps de séjour T_n dans un état $n \in \mathbb{N}$ est une variable exponentielle

$$T_n \sim \begin{cases} \text{Exp}(\lambda) & \text{si } n = 0 \\ \text{Exp}(\lambda + \mu) & \text{si } n \geq 1 \end{cases}$$

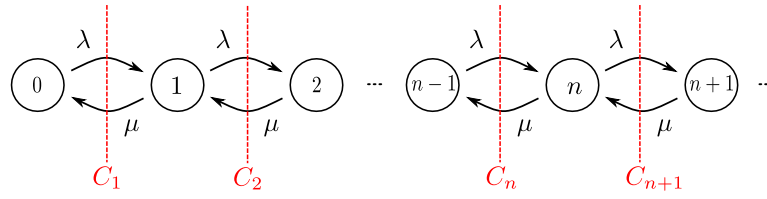


FIGURE 5.28 – Les coupes qui permettent d’obtenir la distribution stationnaire

Les probabilités de transition s’écrivent

$$p_{n,k} = \begin{cases} \frac{\mu}{\lambda + \mu} & \text{si } k = n - 1 \\ \frac{\lambda}{\lambda + \mu} & \text{si } k = n + 1 \\ 0 & \text{sinon} \end{cases}$$

pour $n \geq 2$ et

$$p_{1,2} = 1$$

dans le cas où la chaîne est dans l’état $n = 1$.

distribution stationnaire de probabilité

Pour obtenir la distribution stationnaire de probabilité, le plus simple est d’appliquer la méthode des coupes. Les équations qui correspondent aux coupes de la figure 6.21. sont :

$$\lambda \pi_{n-1} = \mu \pi_n, \quad \forall n \geq 1$$

On obtient directement :

$$\pi_n = \pi_0 \left(\frac{\lambda}{\mu} \right)^n, \quad \forall n \geq 1$$

et la condition de normalisation donne :

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

La convergence de la série au dénominateur est une condition nécessaire et suffisante d’existence de la distribution stationnaire de probabilités, c’est-à-dire également une condition *d’ergodicité*. Elle garantit de plus la stabilité de la file $M/M/1$. Cette convergence est garantie si et seulement si

$$\lambda < \mu$$

c'est-à-dire si le temps moyen entre deux arrivées est supérieur au temps moyen de service. Dans le cas où $\lambda > \mu$, le temps moyen entre deux arrivées est inférieur à la durée moyenne de service. La taille $N(t)$ de la file croît indéfiniment. Tous les états sont transitoires. La probabilité de retour dans la CMTD incluse est inférieure à 1 et

$$\lim_{t \rightarrow \infty} P(N(t) = n) = 0, \quad \forall n \geq 0$$

Les clients s'accumulent dans la file sans que le serveur puisse la résorber. Dans le cas limite $\lambda = \mu$, la méthode des coupes donne pour le régime stationnaire les équations $\pi_n = \pi_0$ pour tout $n \geq 0$. Les probabilités stationnaires π_n sont donc nécessairement toutes nulles. Tous les états sont alors récurrents nuls. La taille de la file croît indéfiniment. L'interprétation intuitive de cette instabilité a déjà été donnée pour le cas plus général de la file simple (voir section 5.3.2).

Afin d'alléger les notations, on notera dans la suite

$$\rho := \frac{\lambda}{\mu} \quad (5.19)$$

le ratio entre le taux d'arrivée et le taux de service. Pour une file $M/M/1$, la condition de stabilité (ou d'ergodicité) est $\rho < 1$. La distribution stationnaire s'écrit :

$$\pi_n = (1 - \rho)\rho^n, \quad \forall n \geq 0 \quad (5.20)$$

paramètres de performance

Dans le cas $\rho < 1$ où la file est stable, le débit moyen en régime stationnaire est

$$X = X_e = \lambda \quad (5.21)$$

C'est aussi nécessairement le débit moyen de sortie :

$$\begin{aligned} X_s &:= P(\text{file non vide}) \cdot \text{taux de service} \\ &= (1 - \pi_0)\mu \\ &= \lambda \end{aligned}$$

Comme le système est stable, on retrouve bien l'égalité – en régime stationnaire – entre débit moyen d'entrée et débit moyen de sortie. Le nombre moyen de clients dans le système est

$$\begin{aligned} Q &:= \sum_{n \geq 0} n\pi_n \\ &= (1 - \rho) \sum_{n \geq 0} n\rho^n \end{aligned}$$

Or

$$\begin{aligned}
 \sum_{n \geq 0} n \rho^n &:= \rho \sum_{n \geq 1} n \rho^{n-1} \\
 &= \rho \sum_{n \geq 1} \frac{dx^n}{dx} \Big|_{x=\rho} \\
 &= \rho \frac{d}{dx} \left[\sum_{n \geq 1} x^n \right] \Big|_{x=\rho} \\
 &= \rho \frac{d}{dx} \left[\frac{1}{1-x} \right] \Big|_{x=\rho} \\
 &= \frac{\rho}{(1-\rho)^2}
 \end{aligned}$$

Donc

$$Q = \frac{\rho}{(1-\rho)} \quad (5.22)$$

Ce nombre moyen de clients dans le système augmente quand $\rho \rightarrow 1$ (avec $\rho < 1$), c'est-à-dire quand le temps moyen entre deux arrivées $\frac{1}{\lambda}$ tend vers la durée moyenne de service $\frac{1}{\mu}$. Le temps moyen de service s'obtient par application de la loi de Little :

$$\begin{aligned}
 R &:= \frac{Q}{X} \\
 &= \frac{\rho}{\lambda(1-\rho)} \\
 &= \frac{1}{\mu(1-\rho)} = \underbrace{\frac{1}{\mu}}_S + \underbrace{\frac{\rho}{\mu(1-\rho)}}_W
 \end{aligned} \quad (5.23)$$

Il se décompose donc en deux parties : le temps moyen de service $S = \frac{1}{\mu}$ et le temps moyen de séjour dans le buffer $W = \frac{\rho}{\mu(1-\rho)}$. Le taux moyen d'utilisation du serveur est celui déjà calculé pour toute file simple mono-serveur, à savoir :

$$U = 1 - \pi_0 = \rho \quad (5.24)$$

La taille moyenne du buffer s'écrit

$$Q_W = Q - Q_S = Q - U = \frac{\rho^2}{1-\rho} \quad (5.25)$$

Les temps moyen de séjour R et nombre moyen de clients Q dans la file $M/M/1$ sont représentés à la figure 5.29. Ils tendent tous les deux vers l'infini quand le ratio ρ tend vers 1.

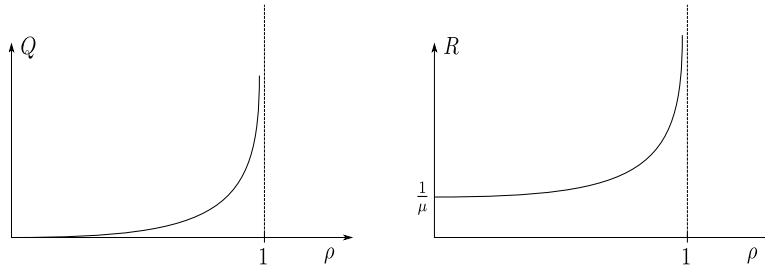


FIGURE 5.29 – nombre moyen de clients (Q) et temps moyen de séjour (R) des clients dans une file $M/M/1$, en fonction du paramètre $\rho = \lambda/\mu$

5.4.2 La file à capacité limitée $M/M/1/K$

La file $M/M/1/K$ (voir figure 5.30) est identique à la file $M/M/1$ excepté que la capacité de la file d'attente est finie : lorsqu'il y a K clients présents dans le système (soit en attente dans le buffer, soit en service), les clients supplémentaires qui arrivent sont perdus.

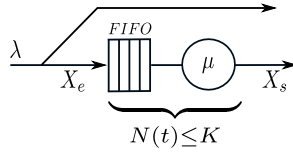


FIGURE 5.30 – Représentation schématique d'une file $M/M/1/K$

CMTC associée à la file $M/M/1/K$

Comme pour la file simple, l'état du système est décrit par le nombre de clients dans le système, $N(t)$. A la différence de la file simple par contre, l'état est fini :

$$N(t) \in E := \{1, \dots, K\}, \forall t > 0$$

Si on décrit la CMTC correspondante en terme de temps de séjour et de probabilités de transition, il faut intégrer un terme de rebouclage direct $p_{K,K}$. Ce rebouclage direct correspond à l'arrivée de clients lorsque la file est pleine. Comme nous l'avons vu dans le chapitre sur les CMTC, ce rebouclage direct n'affecte pas la représentation par taux de transitions. La CMTC qui correspond à la file $M/M/1/K$, avec les choix de notation qui précèdent, est donnée à la figure 5.31. Le *générateur infinitésimal* de la file $M/M/1/K$

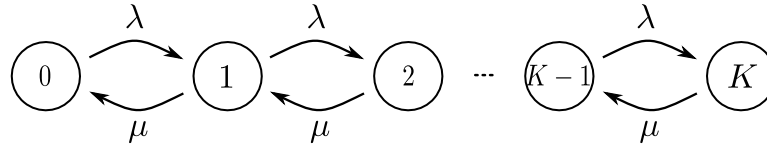


FIGURE 5.31 – La chaîne de Markov à temps continu associée à la file $M/M/1/K$

s'écrit :

$$\mathbf{A} = \begin{pmatrix} -\lambda & \lambda & 0 & \cdots & 0 \\ \mu & -(\lambda + \mu) & \lambda & \cdots & 0 \\ 0 & \mu & -(\lambda + \mu) & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \lambda \\ 0 & 0 & \cdots & \mu & -\mu \end{pmatrix}$$

distribution stationnaire de probabilité

La CMTC qui décrit la file $M/M/1/K$ est finie et irréductible. Tous les états sont donc nécessairement récurrents non nuls. La distribution stationnaire de probabilité existe, est unique et égale à la distribution limite. La file $M/M/1/K$ est donc nécessairement stable et ergodique. Comme dans le cas de la file $M/M/1$, la distribution stationnaire s'obtient facilement à l'aide de la méthode des coupes. On obtient ainsi les relations suivantes :

$$\pi_n = \pi_0 \left(\frac{\lambda}{\mu} \right)^n, \quad \forall n \in \{1, \dots, K\}$$

En posant, comme précédemment,

$$\rho := \frac{\lambda}{\mu}$$

on obtient :

$$\pi_n = \pi_0 \rho^n \tag{5.26}$$

et la condition de normalisation donne :

$$\pi_0 = \frac{1}{\sum_{n=0}^K \rho^n} = \frac{1 - \rho}{1 - \rho^{K+1}} \tag{5.27}$$

Lorsque $K \rightarrow \infty$, on retrouve le résultat obtenu pour la file $M/M/1$.

paramètres de performance

Le débit moyen en régime stationnaire n'est pas le taux d'arrivée λ puisque certains clients sont perdus (lorsque la capacité limite K est atteinte). Ce débit moyen peut être calculé de deux manières (voir figure 5.31) : soit en calculant le taux d'arrivée effectif X_e , soit en calculant le taux de sortie X_s . En régime stationnaire, les deux débits doivent bien sûr être identiques. Le débit moyen de sortie s'écrit :

$$\begin{aligned} X_s &:= P(\text{file non vide}) \cdot \text{taux de service} \\ &= (1 - \pi_0)\mu \\ &= \frac{1 - \rho^K}{1 - \rho^{K+1}}\lambda \end{aligned}$$

Lorsque $K \rightarrow \infty$, on retrouve le résultat obtenu pour la file $M/M/1$ ¹¹. On obtient le même résultat en calculant le taux effectif d'entrée :

$$\begin{aligned} X_e &:= P(\text{file non pleine}) \cdot \text{taux global d'arrivée} \\ &= (1 - \pi_K)\lambda \\ &= \frac{1 - \rho^K}{1 - \rho^{K+1}}\lambda \end{aligned}$$

On a bien le résultat attendu :

$$X = X_s = X_e = \frac{1 - \rho^K}{1 - \rho^{K+1}}\lambda \quad (5.28)$$

En ce qui concerne le taux d'utilisation du serveur, on trouve :

$$\begin{aligned} U &:= (1 - \pi_0) \\ &= \frac{\rho(1 - \rho^K)}{1 - \rho^{K+1}} \end{aligned} \quad (5.29)$$

Le taux d'utilisation d'une file à capacité limitée n'est pas égal à ρ , comme dans le cas des autres files simples. En effet, ce taux d'utilisation est aussi le nombre moyen de clients dans le serveur $U = Q_S$. En appliquant la loi de Little au serveur seul, nous obtenons dans le cas de la file $M/M/1/K$ à capacité limitée :

$$U = Q_S = SX_s = \frac{1}{\mu}X_s$$

Comme le débit moyen X_s est différent du taux d'arrivée λ , le taux d'utilisation du serveur est donc différent de $\rho := \frac{\lambda}{\mu}$. En ce qui concerne le nombre

11. dans le cas $\rho < 1$ où la file $M/M/1$ est stable

moyen de clients dans le système, on obtient :

$$\begin{aligned} Q &:= \sum_{n=0}^K n \pi_n \\ &= \frac{1-\rho}{1-\rho^{K+1}} \sum_{n=0}^K n \rho^n \end{aligned}$$

Or

$$\begin{aligned} \sum_{n=0}^K n \rho^n &:= \rho \sum_{n=0}^K n \rho^{n-1} \\ &= \rho \sum_{n=0}^K \frac{dx^n}{dx} \Big|_{x=\rho} \\ &= \rho \frac{d}{dx} \left[\sum_{n=0}^K x^n \right] \Big|_{x=\rho} \\ &= \rho \frac{d}{dx} \left[\frac{1-x^{K+1}}{1-x} \right] \Big|_{x=\rho} \\ &= \rho \frac{-(K+1)\rho^K(1-\rho) + (1-\rho^{K+1})}{(1-\rho)^2} \\ &= \rho \frac{1-(K+1)\rho^K + K\rho^{K+1}}{(1-\rho)^2} \end{aligned}$$

On trouve finalement

$$Q = \frac{\rho}{(1-\rho)} \cdot \frac{1-(K+1)\rho^K + K\rho^{K+1}}{1-\rho^{K+1}} \quad (5.30)$$

Ce nombre moyen tend vers celui d'une file $M/M/1$ lorsque $K \rightarrow \infty$ et $\rho < 1$. Le temps moyen de séjour peut être défini de deux manières distinctes : le temps moyen de séjour R d'un client effectivement admis dans la file ou le temps moyen de séjour total R_T calculé avec un temps nul pour les clients rejetés. Les deux temps moyen de séjour s'obtiennent par la loi de Little, mais appliquée à deux systèmes différents, représentés à la figure 5.32. En ne considérant que les clients effectivement admis dans la file, on obtient :

$$\begin{aligned} R &:= \frac{Q}{X} \\ &= \frac{1-(K+1)\rho^K + K\rho^{K+1}}{\mu(1-\rho)(1-\rho^K)} \end{aligned} \quad (5.31)$$

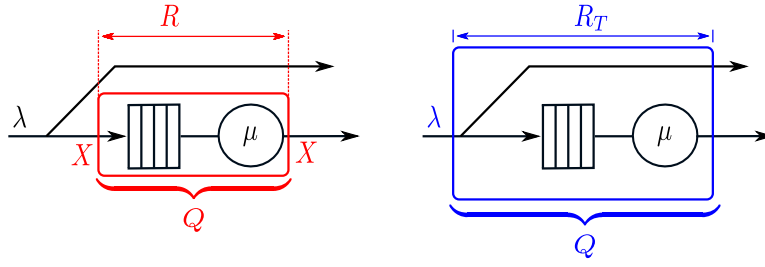


FIGURE 5.32 – Les deux systèmes distincts utilisés pour déterminer les temps de séjour R et R_T moyen, pour la file $M/M/1/K$, en utilisant la loi de Little

Par contre, en considérant pour le système tous les clients, y compris ceux rejetés, on obtient :

$$\begin{aligned} R_T &:= \frac{Q}{\lambda} \\ &= \frac{1 - (K+1)\rho^K + K\rho^{K+1}}{\mu(1-\rho)(1-\rho^{K+1})} \end{aligned} \quad (5.32)$$

On obtient ainsi la proportion α de clients admis :

$$\alpha := \frac{X}{\lambda} = \frac{R}{R_T} = \frac{1 - \rho^K}{1 - \rho^{K+1}} \quad (5.33)$$

5.4.3 La file à plusieurs serveurs $M/M/C$

La file $M/M/C$ diffère de la file $M/M/1$ parce qu'elle comporte C serveurs – en général identiques et indépendants les uns des autres – qui opèrent en parallèle. Elle est représentée à la figure 5.33. Nous considérerons ci-après

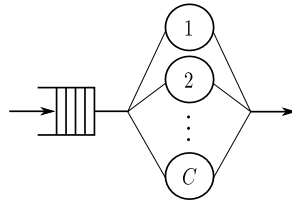


FIGURE 5.33 – Représentation schématique d'une file $M/M/C$

le cas où les temps de service sont des exponentielles indépendantes de même paramètre μ . Lorsque le nombre de clients dans le système est inférieur ou égal à C , tous les clients sont en service. Dans le cas contraire, les clients

surnuméraires sont placés dans la file et seront servi sur base d'une discipline de service *first come, first served* (FCFS). Les clients arrivent, comme précédemment, selon un processus de Poisson de paramètre λ .

CMTC associée à la file $M/M/C$

L'état du système est décrit par le nombre total de clients dans le système, $N(t) \in \mathbb{N}$. Lorsque $n \leq C$, tous les clients sont en service et chacun est susceptible de quitter son serveur avec un taux exponentiel. Le taux de transition de l'état n vers l'état $n - 1$ vaut donc $n\mu$. Lorsque $n > C$, les C serveurs sont occupés et le taux de transition vaut donc $C\mu$, quel que soit n . Le taux d'arrivée des clients, c'est-à-dire le taux de transition de n à $n + 1$ vaut lui toujours λ , quel que soit le nombre de clients dans le système. La CMTC qui correspond à la file $M/M/C$ est donnée à la figure 6.40. Le

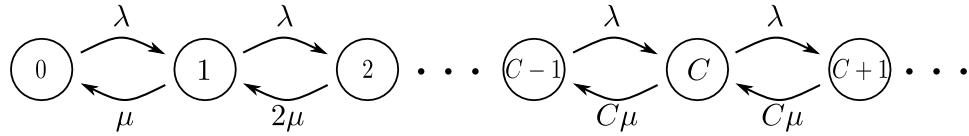


FIGURE 5.34 – La chaîne de Markov à temps continu associée à la file $M/M/C$

générateur infinitésimal (infini) de la file $M/M/C$ s'écrit :

$$\mathbf{A} = \begin{pmatrix} -\lambda & \lambda & 0 & \cdots & \cdots & \cdots & 0 & \cdots \\ \mu & -\lambda - \mu & \lambda & \ddots & \cdots & \cdots & 0 & \cdots \\ 0 & 2\mu & -\lambda - 2\mu & \ddots & \ddots & \ddots & \ddots & \ddots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \\ \vdots & \vdots & 0 & C\mu & -\lambda - C\mu & \lambda & 0 & \ddots \\ \vdots & \vdots & \ddots & 0 & C\mu & -\lambda - C\mu & \lambda & \ddots \\ \vdots & \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \end{pmatrix}$$

distribution stationnaire de probabilité

Pour obtenir la distribution stationnaire de la file $M/M/C$, nous appliquons la méthode des coupes, en choisissant les mêmes coupes que pour la file $M/M/1$. On obtient dans ce cas les relations suivantes :

$$\begin{aligned} \lambda \pi_{n-1} &= n\mu \pi_n, \quad \forall n \in \{1, \dots, C\} \\ \lambda \pi_{n-1} &= C\mu \pi_n, \quad \forall n > C \end{aligned}$$

En posant, comme précédemment,

$$\rho := \frac{\lambda}{\mu}$$

on obtient :

$$\pi_n = \begin{cases} \frac{\rho^n}{n!} \pi_0 & \text{si } n \leq C \\ \frac{\rho^n}{C! C^{n-C}} \pi_0 & \text{si } n > C \end{cases} \quad (5.34)$$

La condition de normalisation donne :

$$\begin{aligned} \sum_{n \geq 0} \pi_n &= \pi_0 \left(\sum_{n=0}^C \frac{\rho^n}{n!} + \sum_{n \geq C+1} \frac{\rho^n}{C! C^{n-C}} \right) \\ &= \pi_0 \sum_{n=0}^C \frac{\rho^n}{n!} + \pi_0 \frac{C^C}{C!} \underbrace{\sum_{n \geq C+1} \left(\frac{\rho}{C} \right)^n}_{\Sigma_\infty} \end{aligned}$$

La somme Σ_∞ converge si et seulement si $\rho < C$. Si cette condition est satisfaite, la distribution stationnaire existe et c'est également la distribution limite. Tous les états de la CMTC irréductible sont alors récurrents non nuls et la chaîne est ergodique. Les conditions de stabilité, d'ergodicité et d'existence d'une distribution stationnaire non triviale sont une seule et même condition qui s'écrit, dans le cas de la file $M/M/C$:

$$\lambda < C\mu$$

Si cette condition est satisfaite, on obtient :

$$\Sigma_\infty = \frac{\rho^{C+1}}{C^{C+1}} \frac{C}{C - \rho}$$

et

$$\begin{aligned} \pi_0 &= \frac{1}{\sum_{n=0}^C \frac{\rho^n}{n!} + \frac{\rho^{C+1}}{C! (C - \rho)}} \\ &= \frac{1}{\sum_{n=0}^{C-1} \frac{\rho^n}{n!} + \frac{\rho^C}{C!} + \frac{\rho^{C+1}}{C! (C - \rho)}} \\ &= \frac{1}{\sum_{n=0}^{C-1} \frac{\rho^n}{n!} + \frac{\rho^C}{(C-1)! (C - \rho)}} \end{aligned}$$

Lorsque $C = 1$, on retrouve les résultats obtenus pour la file $M/M/1$.

paramètres de performance

En régime stationnaire (donc lorsque la condition de stabilité $\lambda < C\mu$ est satisfaite), les débits moyens d'entrée et de sortie sont nécessairement égaux. On a alors :

$$X_s = X_e = \lambda$$

On peut d'ailleurs retrouver ce résultat par calcul. Le taux de service croît proportionnellement au nombre de serveurs en activité, jusqu'à ce que tous les serveurs soient occupés. Le débit moyen de sortie s'écrit donc dans le cas de la file $M/M/C$:

$$X_s := \sum_{n=1}^{C-1} n\mu\pi_n + \sum_{n \geq C} C\mu\pi_n$$

En substituant les valeurs obtenues dans (5.34) et (5.35) pour la distribution stationnaire, on retrouve bien par calcul $X_s = \lambda$.

Dans le cas de la file $M/M/C$, il est plus simple de calculer d'abord le temps moyen de séjour d'un client, puis d'obtenir par la loi de Little le nombre moyen de client dans le système. Le temps moyen de séjour d'un client, R , se décompose en son temps moyen d'attente dans le buffer, W , et son temps moyen de service, S :

$$R = W + S$$

Le temps moyen de service est identique pour tous les serveurs et vaut

$$S = \frac{1}{\mu}$$

Par ailleurs, en appliquant la loi de Little au buffer seulement, on obtient :

$$W = \frac{Q_W}{\lambda}$$

Il reste à calculer le nombre moyen de clients dans le buffer, Q_W . Lorsqu'il y a un client dans le buffer, il y a au moins $C + 1$ clients dans le système. Or, pour un nombre de clients supérieur à C , la distribution stationnaire de

probabilité est une suite géométrique qu'il est facile de sommer ¹² :

$$\begin{aligned}
 Q_w &= \sum_{n \geq C} (n - C) \pi_n \\
 &= \sum_{n \geq C} (n - C) \frac{\rho^n}{C! C^{n-C}} \pi_0 \\
 &= \frac{\rho^{C+1}}{C! C} \pi_0 \sum_{n \geq C} (n - C) \left(\frac{\rho}{C}\right)^{n-C-1} \\
 &= \frac{\rho^{C+1}}{C! C} \pi_0 \sum_{k \geq 0} k \left(\frac{\rho}{C}\right)^{k-1} \\
 &= \frac{\rho^{C+1}}{C! C} \pi_0 \frac{1}{\left(1 - \frac{\rho}{C}\right)^2} \\
 &= \frac{\rho^{C+1}}{(C-1)!(C-\rho)^2} \pi_0
 \end{aligned} \tag{5.35}$$

On obtient donc pour le temps moyen de séjour dans l'ensemble de la file

$$R = \frac{\rho^C}{\mu(C-1)!(C-\rho)^2} \pi_0 + \frac{1}{\mu} \tag{5.36}$$

Le nombre moyen de clients s'obtient par la loi de Little

$$Q = R\lambda = \frac{\rho^{C+1}}{(C-1)!(C-\rho)^2} \pi_0 + \rho \tag{5.37}$$

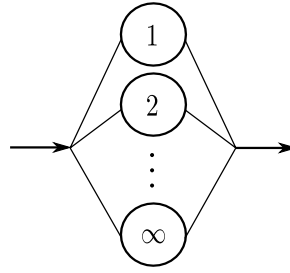
5.4.4 La file *Infinite Servers* (IS) $M/M/\infty$

La file $M/M/\infty$ est utilisée pour modéliser les systèmes où il y a au moins autant de serveurs disponibles que de clients. Il n'y a donc pas de buffer dans une file $M/M/\infty$: sitôt entrés dans la file, les clients sont immédiatement servis. La file $M/M/\infty$ représente en général un délai (lié au temps de service nécessaire à chaque client) dans un réseau de files d'attente. La représentation schématique de cette file est donnée à la figure 5.35. Nous considérerons ci-après le cas où les temps de service sont des exponentielles indépendantes de même paramètre μ et où les clients arrivent selon un processus de Poisson de paramètre λ .

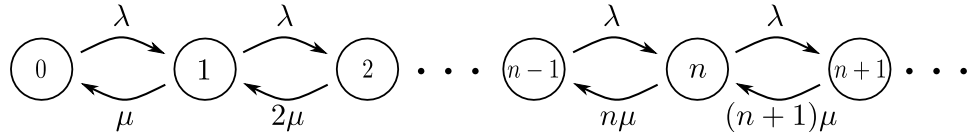
CMTC associée à la file $M/M/\infty$

L'état du système est décrit par le nombre total de clients dans le système $N(t) \in \mathbb{N}$. Lorsque n clients sont en service, chacun est susceptible de quitter

12. C'est la raison pour laquelle, pour la file $M/M/C$, on commence par calculer le paramètre Q_w avant d'en déduire tous les autres

FIGURE 5.35 – Représentation schématique d'une file $M/M/\infty$

son serveur avec un taux exponentiel. Le taux de transition de l'état n vers l'état $n - 1$ vaut donc $n\mu$. Le taux d'arrivée des clients, c'est-à-dire le taux de transition de n à $n + 1$ vaut lui toujours λ . La CMTC qui correspond à la file $M/M/\infty$ est donnée à la figure 5.36. Le *générateur infinitésimal* (infini)

FIGURE 5.36 – La chaîne de Markov à temps continu associée à la file $M/M/\infty$

de la file $M/M/\infty$ s'écrit :

$$\mathbf{A} = \begin{pmatrix} -\lambda & \lambda & 0 & 0 & \cdots \\ \mu & -\lambda - \mu & \lambda & 0 & \cdots \\ 0 & 2\mu & -\lambda - 2\mu & \lambda & \ddots \\ \vdots & \ddots & \ddots & \ddots & \ddots \end{pmatrix}$$

distribution stationnaire de probabilité

Pour obtenir la distribution stationnaire de la file $M/M/\infty$, nous appliquons la méthode des coupes, en choisissant les mêmes coupes que pour la file $M/M/1$. On obtient dans ce cas les relations suivantes :

$$\lambda\pi_{n-1} = n\mu\pi_n, \quad \forall n \geq 1$$

En posant, comme précédemment,

$$\rho := \frac{\lambda}{\mu}$$

on obtient :

$$\pi_n = \frac{\rho^n}{n!} \pi_0 \quad (5.38)$$

La condition de normalisation donne :

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

c'est-à-dire

$$\pi_0 = e^{-\rho} \quad (5.39)$$

Dans la file $M/M/\infty$, tout client qui se présente est instantément servi : la capacité de traitement est à priori illimitée. Dès lors, la stabilité de la file est garantie. Quel que soit le taux d'arrivée λ , il existe un seuil $N \in \mathbb{N}$ tel que :

$$n \geq N \Rightarrow n\mu > \lambda$$

A partir de ce seuil, le taux de service devient plus grand que le taux d'arrivée et la probabilité que la file diminue est supérieure à celle qu'elle augmente. La file $M/M/\infty$ est donc toujours stable et ergodique. Tous les états sont nécessairement récurrents non nuls.

paramètres de performance

En régime stationnaire (qui est nécessairement atteint, la file étant toujours stable et ergodique), les débits moyens d'entrée et de sortie sont égaux :

$$X_s = X_e = \lambda$$

On peut d'ailleurs retrouver le débit moyen de sortie par calcul :

$$\begin{aligned} X_s &= \sum_{n \geq 0} n\mu\pi_n \\ &= \sum_{n \geq 0} n\mu \frac{\rho^n}{n!} e^{-\rho} \\ &= \mu e^{-\rho} \rho \sum_{n \geq 1} n \frac{\rho^{n-1}}{(n-1)!} \\ &= \mu e^{-\rho} \rho e^\rho \\ &= \mu\rho = \lambda \end{aligned}$$

Le nombre moyen de clients s'écrit :

$$\begin{aligned} Q &= \sum_{n \geq 0} n\pi_n \\ &= e^{-\rho} \sum_{n \geq 1} \frac{\rho^n}{(n-1)!} \\ &= e^{-\rho} \rho e^\rho \\ &= \rho \end{aligned} \quad (5.40)$$

Le temps moyen de séjour s'obtient par la loi de Little :

$$R = \frac{Q}{\lambda} = \frac{\rho}{\lambda} = \frac{1}{\mu} \quad (5.41)$$

Ce résultat était évident dès le départ puisque les clients entrant dans la file sont servis immédiatement et que le temps de leur service est distribué selon une exponentielle de paramètre μ . Nous aurions donc pu déduire sans calcul le nombre moyen de clients dans la file, par la loi de Little, à partir de ce temps moyen de séjour. Dans ce cas particulier de la file $M/M/\infty$, le calcul préalable de la distribution stationnaire de probabilité n'était en fait pas nécessaire pour calculer les paramètres de performance.

5.4.5 La file markovienne générale

Dans cette file généralisée, les taux d'arrivées et de service dépendent de l'état $N(t)$, c'est-à-dire du nombre de clients présents dans le système à l'instant $t > 0$ considéré. La représentation schématique de cette file est donnée à la figure 5.37.

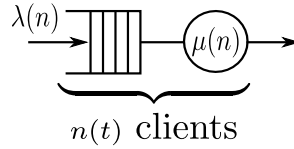


FIGURE 5.37 – Représentation schématique d'une file markovienne généralisée. Les taux d'arrivées $\lambda(n)$ et de service $\mu(n)$ dépendent du nombre de client n présents dans le système

CMTC associée à la file markovienne générale

L'état du système est le nombre $N(t)$ de clients dans le système au temps $t > 0$. Nous avons déjà rencontré la chaîne de Markov associée à cette file markovienne générale : il s'agit d'un exemple de *processus de naissance et de mort*¹³ dans lequel les taux de transition dépendent de l'état. La CMTC associée à la file markovienne générale est donnée à la figure 5.38. Le modèle de ce processus général de naissance et de mort est défini par les *taux de naissance* $\lambda(n)$ et les *taux de mortalité* $\mu(n)$. Le *générateur infinitésimal*

13. i.e. d'une chaîne de Markov sur \mathbb{N} où les transitions d'état sont de deux types seulement : les *naissances* où l'état passe de n à $n + 1$ et les *morts* où l'état passe de n à $n - 1$

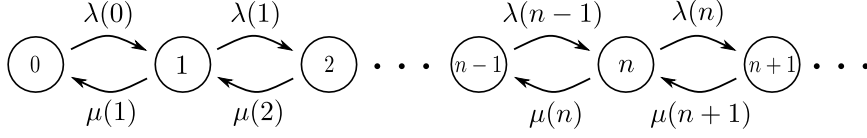


FIGURE 5.38 – La chaîne de Markov à temps continu associée à la file markovienne générale

(infini) associé s'écrit :

$$\mathbf{A} = \begin{pmatrix} -\lambda(0) & \lambda(0) & 0 & \dots \\ \mu(1) & -(\lambda(1) + \mu(1)) & \lambda(1) & \dots \\ 0 & \mu(2) & -(\lambda(2) + \mu(2)) & \ddots \\ \vdots & \vdots & \ddots & \ddots \end{pmatrix} \quad (5.42)$$

distribution stationnaire de probabilité

En utilisant les mêmes coupes que pour la file $M/M/1$, on obtient cette fois :

$$\lambda(n-1)\pi_{n-1} = \mu(n)\pi_n, \quad \forall n \geq 1 \quad (5.43)$$

c'est-à-dire

$$\pi_n = \pi_0 \prod_{k=0}^{n-1} \frac{\lambda(k)}{\mu(k+1)}, \quad \forall n \geq 1 \quad (5.44)$$

La condition de normalisation donne :

$$\pi_0 = \frac{1}{1 + \sum_{n=1}^{\infty} \prod_{k=0}^{n-1} \frac{\lambda(k)}{\mu(k+1)}} \quad (5.45)$$

La convergence de la série au dénominateur, dans l'équation (5.45), est une condition nécessaire et suffisante d'existence de la distribution stationnaire de probabilités (c'est donc aussi la condition de stabilité et d'ergodicité). Une condition nécessaire et suffisante pour la convergence de cette série est l'existence d'un indice $k_0 >$ tel que :

$$k \geq k_0 \Rightarrow \lambda(k) < \mu(k) \quad (5.46)$$

Si cette condition est satisfaite, à partir d'un certain rang le taux d'arrivée est inférieur au taux de service. La *queue* de la CMTC est alors stable. Tous les états de la chaînes sont récurrents non nuls. A contrario, si cette condition n'est pas satisfaite, la taille de la file va croître indéfiniment, sur le long terme, et tous les états seront (ensemble) soit récurrents nuls, soit transitoires.

paramètres de performance

Dans le cas d'une file markovienne générale, les paramètres de performance sont le débit moyen (calculé à partir du débit moyen de sortie)

$$X_e = X_s = \sum_{n \geq 1} \mu(n) \pi_n$$

le nombre moyen de clients

$$Q = \sum_{n \geq 0} n \pi_n$$

et le temps moyen de séjour (obtenu par la loi de Little)

$$R = \frac{Q}{X} = \frac{\sum_{n \geq 0} n \pi_n}{\sum_{n \geq 1} \mu(n) \pi_n}$$

5.5 Exercices

Exercice 28

Dans une école, se proposent deux choix : soit une installer une imprimante, soit installer 3 imprimantes indépendantes, mais 3 fois moins rapides que dans le premier cas. Pour simplifier le calcul, le processus d'arrivée des requêtes d'impression sera modélisé comme un processus de Poisson. Les temps moyen de traitement des requêtes d'impression sont supposés distribués selon des lois exponentielles indépendantes.

1. Dans un premier temps, la configuration envisagée pour la solution à trois imprimantes est de les doter chacune d'un spooler d'impression indépendant. On suppose que l'installation des trois imprimantes peut être faite de manière à ce que les requêtes d'impression soit équitablement distribuée – en régime permanent – entre les trois imprimantes. Quel est le meilleur choix à faire, en terme de temps de réponse (confort de l'utilisateur) et en terme de stockage (nombre de requêtes stockées) entre les deux configurations ?
2. La solution avec une seule imprimante étant jugée trop fragile et incommode pour les usagers, on cherche une configuration plus efficace pour la solution à trois imprimantes. On regarde alors une solution où les trois imprimantes travaillent en parallèle, mais sont alimentées par un spooler d'impression unique. Comparer cette solution avec les deux précédentes en terme de temps de réponse.

Exercice 29

Dans cet exercice, on s'intéresse en particulier à l'effet du routage dynamique dans des files d'attente aux guichets des administrations, dans les grands magasins, etc. On dispose de deux serveurs (par exemple deux employés) qui traitent des requêtes similaires avec la même efficacité. Le traitement des requêtes (de durée variée, en général) nécessite un temps de service modélisé par une exponentielle de paramètre μ .

1. Dans un premier temps, on cherche à comparer les deux organisations représentées à la figure 6.36. Dans l'organisation *à l'américaine* (à gauche), les clients attendent dans un buffer unique et la discipline de service est *first come, first served*. Dans la file *à la française* (à droite), les clients se répartissent aléatoirement dans deux files, de manière équiprobable. On posera pour simplifier les notations dans les calculs :

$$\rho := \frac{\lambda}{\mu}$$

- (a) donner les conditions de stabilité pour les deux organisations
- (b) calculer pour les deux organisations le nombre moyen de clients présents et le temps moyen de séjour

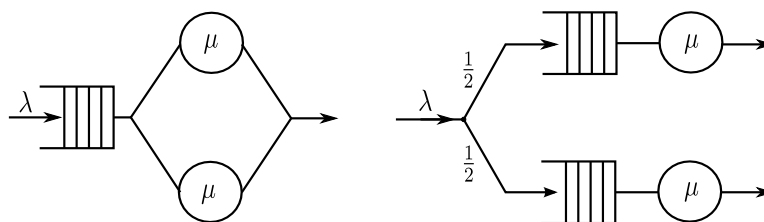


FIGURE 5.39 – Les deux organisations de files étudiées dans l'exercice 29 (premier point). Elles sont parfois appelées respectivement – improprement – *à l'américaine* (à gauche) et *à la française* (à droite)

- (c) comparer les performances de ces deux organisations en terme de nombre de clients et de temps de séjour
 - (d) calculer et comparer la proportion du temps où au moins un employé est inactif, dans les deux organisations
2. Dans un deuxième temps, on cherche à représenter plus fidèlement l'organisation *à la française*. Dans la réalité, les clients ne se répartissent pas aléatoirement mais choisissent, en arrivant, de se ranger dans la file qui contient le moins de clients¹⁴. Ce type d'organisation est représentée à la figure 6.37. Le routage devient donc dynamique : la

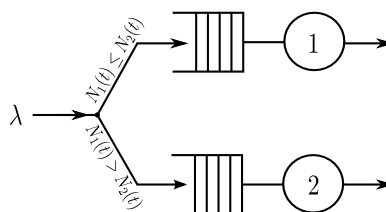


FIGURE 5.40 – Une organisations de file *à la française* avec un routage dynamique

route choisie par un client qui arrive au temps t dépend de l'état du système $(N_1(t), N_2(t))$ à cet instant.

- (a) donner une condition de stabilité pour cette organisation
- (b) donner un modèle de ce système sous la forme d'une chaîne de Markov à temps continu
- (c) écrire les équations d'équilibre qui permettent de calculer la distribution stationnaire de probabilité (sans les résoudre)

14. un observateur attentif de l'organisation des files *à la française* pourrait critiquer cette représentation simpliste et affirmer – non sans raison – que dans ce type d'organisation, les clients changent de plus de file lorsqu'ils ont l'impression que l'autre file avance plus vite ...

- (d) donner les expressions du nombre moyen de clients présents et du temps moyen de séjour des clients, en fonction des composantes (non calculées) de la distribution stationnaire de probabilité

Exercice 30 (La file M/M/C/C) On considère un système de type *commutateur téléphonique*, composé de C relais permettant de rediriger les appels vers la bonne destination. Chaque relais ne traite qu'un seul appel à la fois, avec un temps de service aléatoire de distribution exponentielle $\text{Exp}(\mu)$. Les appels arrivent selon un Processus de Poisson d'intensité λ , et sont ensuite dirigés vers un relais libre. Si un nouvel appel arrive alors que tous les relais sont occupés, il est perdu.

1. Donner l'expression (en fonction de λ , μ et n) des taux d'arrivées $\lambda(n)$ et des taux de départ $\mu(n)$ des appels dans le commutateur lorsque n appels sont en cours de traitement. Modéliser ce problème par une chaîne de Markov à temps continu (CMTC). Donner le graphe associé, les taux de transition et le générateur infinitésimal de la CMTC.
2. Quelle est la condition de stabilité de cette file ? Le système est-il ergodique ?
3. Calculer la distribution stationnaire de probabilité
4. Quel est le taux de perte (proportion d'appels perdus) ?
5. Donner l'expression du nombre moyen d'appels en cours de traitement, du temps moyen de séjour des appels dans le commutateur et du débit moyen de sortie des appels

Exercice 31 (discipline de service dynamique) Dans un magasin, une caisse fonctionne en permanence. Parfois la file d'attente à cette caisse est trop longue. Une deuxième caisse est alors ouverte (partageant la même file) si le nombre de clients attendant dans la file (i.e. le buffer) est au moins égal à K . Le premier client arrivé passe alors immédiatement à cette deuxième caisse (le buffer devient donc de taille $K - 1$ à cet instant). Quand le nombre de clients dans la file (i.e. le buffer) devient strictement inférieur à $K - 1$, la première des deux caisses qui termine le service en cours est fermée (et ne prend donc plus de client, jusqu'à ce qu'elle soit à nouveau ouverte). On suppose que le taux de service de la deuxième caisse est le même que celui de la première.

1. Modéliser cette file par une CMTC et calculer la distribution stationnaire de probabilité de cette CMTC
2. Cette file est-elle stable et ergodique ?
3. Calculer le nombre moyen de clients dans le système en régime permanent
4. Calculer, pour le cas $\lambda = \mu$, le seuil d'ouverture K de façon à ce que le nombre moyen de clients dans la file (buffer) soit – en régime stationnaire – inférieur ou égal à 4.

5. Calculer dans le cas général le taux d'utilisation¹⁵ de la caisse N°2.

Exercice 32 (File avec plantage) On considère une file d'attente $M/M/1$ (de paramètres λ, μ) *modifiée*, où le serveur peut en “plantant” détruire tous les clients de la file. Après le “plantage”, le serveur reprend son service normalement. On suppose que le temps de bon fonctionnement du serveur suit une loi $Exp(\gamma)$. On suppose $\gamma > \mu$ et $\lambda < \mu$

1. Modéliser cette file par une chaîne de Markov à temps continu (CMTC) en choisissant comme état le nombre de clients dans le système au temps $t > 0$.
2. A l'aide du théorème des coupes, calculer le nombre moyen de clients dans la file *en régime stationnaire* en fonction de π_0 (on note π le vecteur des probabilités stationnaires et π_0 sa première composante)
3. On cherche à calculer le régime stationnaire
 - (a) En utilisant les équations d'équilibre, donner une relation liant π_n, π_{n+1} et π_{n-1}
 - (b) Montrer qu'il existe $r_1 > 1$ et $r_2 < 1$ tels que $\pi_n = Ar_1^n + Br_2^n$ (on pourra introduire $\rho = \frac{\lambda}{\mu}$ et $\alpha = \frac{\gamma}{\mu}$)
 - (c) La file considérée est-elle stable? Pourquoi? Que peut-on en conclure sur la valeur des paramètres A ou B ?
 - (d) Calculer π_n ($n \geq 0$), en utilisant la condition de normalisation du vecteur des probabilités stationnaires

Exercice 33 (File avec impatience) On considère une *file avec impatience*, i.e. une file dans laquelle un client qui arrive et voit n clients (dans le système) devant lui a une probabilité p_n de s'insérer dans la file (et une probabilité $1 - p_n$ de partir). Dans cet exercice, on considèrera que les clients arrivent selon un processus de Poisson de paramètre λ (avant de décider si ils entrent dans le système). Le temps de service d'un client est une variable aléatoire avec une distribution exponentielle de paramètre μ . La probabilité d'entrée effective dans le système pour un nouveau client arrivant vaut

$$p_n := \frac{1}{1 + n}, \quad \forall n \geq 0$$

où n désigne le nombre de clients déjà dans le système lorsque le nouveau client arrive.

1. Modéliser cette file par une chaîne de Markov à temps continu
2. Calculer la distribution stationnaire de probabilité associée à cette chaîne

15. Ce taux reflète par exemple la proportion du temps où l'employé de la caisse N°2 sera disponible pour d'autres tâches dans le magasin

3. Calculer le taux moyen stationnaire d'entrée (X_e) et le taux moyen stationnaire de sortie (X_s). A partir de ces deux résultats, discuter la stabilité de la file (éventuellement, en fonction des valeurs des paramètres λ et μ)
4. calculer le nombre moyen de clients dans le système
5. calculer le temps de séjour moyen d'un client dans le système (lorsqu'il y entre effectivement)
6. la distribution stationnaire de cette file avec impatience est-elle la même que pour une file $M/M/\infty$? Les paramètres de performance qui en découlent sont-ils les mêmes que pour une file $M/M/\infty$?

Exercice 34 (File simple à priorité) On considère un *commutateur téléphonique à priorités* capable de gérer simultanément jusqu'à C appels. Les appels sont subdivisés en trois classes : les appels de type 1, 2 et 3. Les appels de type r arrivent selon un processus de Poisson de taux λ_r (pour $r \in \{1, 2, 3\}$). Lorsqu'un appel arrive alors qu'il y a (strictement) moins de C appels déjà présents dans le commutateur, il est instantanément traité. Tous les appels nécessitent un temps de traitement aléatoire – indépendant du type d'appel – distribué suivant une loi exponentielle de paramètre μ . Lorsqu'un appel arrive alors qu'il y a déjà C appels en cours de traitement dans le commutateur, cet appel est dans certains cas mis en attente dans un buffer FCFS de capacité limitée et, dans d'autres cas, rejeté. Le buffer est subdivisé en trois zones distinctes de tailles respectives k_r pour $r \in \{1, 2, 3\}$. La zone r ne peut accueillir que des appels de type $l \leq r$. La zone 2, par exemple, est de capacité k_2 et peut accueillir des appels de types 1 et 2. Lorsqu'un appel arrive et que les clients déjà présents dans le buffer remplissent toutes les places disponibles pour ce type d'appel, l'appel arrivant est rejeté. Sinon, l'appel arrivant est placé en dernière position dans le buffer.

1. Représenter schématiquement la file simple qui correspond à ce problème. Est-il nécessaire de mémoriser le type des appels une fois ceux-ci admis dans le commutateur?
2. Modéliser le commutateur comme une file markovienne monoclasse. Donner la chaîne de Markov à temps continu associée à cette file
3. La file est-elle stable et ergodique?
4. Calculer la distribution stationnaire de probabilité
5. Quelles sont les probabilités de rejet pour les appels de chacun des trois types?
6. Quels sont les nombres moyens d'appels satisfaits par unité de temps? Quel est le débit moyen de sortie tous types d'appels confondus?
7. Calculer le temps moyen d'attente (dans le buffer) pour chaque type d'appel

8. En déduire le nombre moyen d'appels de chaque type dans le buffer

Exercice 35 (file multiclasse)

On considère une file d'attente $M/M/1$ où les clients arrivés sont de 2 types, A ou B . Les arrivées sont les sommes de 2 processus de Poisson indépendants, d'intensités λ_A et λ_B . Par ailleurs, le temps de service des clients A est $Exp(\mu_A)$ et ceux des clients B est $Exp(\mu_B)$

1. Un client entre en service (dans le serveur). On note T son temps de service. Calculer la densité de T , constater que ce n'est pas une loi exponentielle.
2. Comment peut-on décrire l'état du système à un temps $t > 0$ pour que représenter le système par un modèle markovien ?
3. Donner une représentation par graphe et taux de transition d'un modèle markovien associé à ce système.

Chapitre 6

Corrections des exercices

6.1 Processus de Poisson

Exercice 1

Robert est un grand romantique. Pour conquérir le coeur de Stéphanie il a décidé de sortir le grand jeu et de l'emmener contempler les étoiles filantes. Pour préparer sa sortie, et par commodité, il décide de modéliser le processus d'apparition des étoiles filantes par un processus de Poisson, $N(t)$ (comptant le nombre d'étoiles filantes apparues entre le début de l'observation et le temps t , exprimé en minutes), de paramètre λ . Lors d'une observation préparatoire, la veille du jour j , il a compté 30 étoiles filantes entre 22h00 et 23h00. Il décide donc de tenter sa chance le lendemain, à la même heure.

1. Calculez le temps moyen que Stéphanie et Robert attendront entre 2 étoiles filantes. En déduire λ
2. Ce qui est capital se dit Robert c'est qu'il y ait tout de suite des étoiles, sinon elle va s'impatiser. Calculez la probabilité qu'il y ait une étoile dans les 5 premières minutes.
3. Calculer la probabilité que la première étoile apparaisse avant 22h06
4. Calculez la probabilité pour que la dixième étoile arrive après 23 heures
5. Stéphanie est jolie mais pas très romantique ; elle voudra partir si au bout de 10 minutes il n'y a pas d'étoiles. calculez la probabilité de ce (sinistre) évènement
6. Sachant que 20 étoiles sont apparues dans la première heure, calculez la probabilité qu'exactly 10 étoiles sont apparues dans le deuxième quart d'heure.

Correction de l'Exercice 1

1. Les temps inter-arrivées sont des variables aléatoires indépendantes de distributions exponentielles

$$T_n \sim \text{Exp}(\lambda), \forall n \geq 1$$

L'espérance d'une variable aléatoire exponentielle de paramètre λ est :

$$E(T_n) = \frac{1}{\lambda}$$

L'estimation par maximum de vraisemblance¹ du paramètre λ vaut :

$$\hat{\lambda} = \frac{N(T)}{T} = \frac{30}{60} = \frac{1}{2} \text{ [min}^{-1}\text{]}$$

La durée moyenne entre les apparitions de deux étoiles filantes successives est donc

$$\frac{1}{\hat{\lambda}} = 2 \text{ [min]}$$

ce qui était bien sûr prévisible dès l'énoncé du problème ...

2. Il est possible d'évaluer de nombreuses probabilités relatives à un processus de Poisson de deux manières différentes :
 - (i) en les formulant comme des probabilités sur les instants d'arrivées et la durée des intervalles inter-arrivées, faisant intervenir des calculs sur des variables aléatoires continues (loi Gamma, loi uniforme ou loi exponentielle)
 - (ii) en les formulant comme des probabilités sur le nombre d'arrivées sur des intervalles connus, faisant intervenir dans ce cas des calculs sur des variables aléatoires discrètes (loi de Poisson ou loi binomiale)

La première méthode consiste ici à calculer :

$$IP[T_1 \leq 5] = \int_0^5 \lambda e^{-\lambda t} dt = \left[-e^{-\lambda t} \right]_{t=0}^{t=5} = 1 - e^{-\frac{5}{2}}$$

Tandis que la deuxième méthode consiste à calculer :

$$\begin{aligned} IP[N(5) \geq 1] &= 1 - P(N(5) < 1) \\ &= 1 - \frac{(\lambda t)^0 e^{-\lambda t}}{0!} \Big|_{t=5} \\ &= 1 - e^{-\frac{5}{2}} \end{aligned}$$

Dans les deux cas, on obtient bien sûr la même réponse $1 - e^{-\frac{5}{2}} \simeq 0.918..!$

1. On peut montrer qu'il s'agit de l'estimation non biaisée de variance minimum, connaissant les instants d'arrivées t_i pour $i \in \{1, \dots, N\}$ de chacune des N étoiles filantes sur $[0, T]$

3. C'est la même question que la précédente. On a par exemple :

$$IP[T_1 \leq 6] = \int_0^6 \lambda e^{-\lambda t} dt = 1 - e^{-3} \simeq 0.95..$$

4. Avec la première méthode, on obtient

$$\begin{aligned} IP[A(10) \geq 60] &= 1 - IP[A(10) < 60] \\ &= 1 - \int_0^{60} \frac{(\lambda t)^9 \lambda e^{-\lambda t}}{9!} dt \end{aligned} \quad (6.1)$$

qui nécessite 9 intégrations par parties successives. En utilisant la deuxième méthode, on obtient :

$$\begin{aligned} IP[N(60) < 10] &= \sum_{k=0}^9 \frac{(\lambda 60)^k}{k!} e^{-\lambda 60} \\ &= \sum_{k=0}^9 \frac{30^k}{k!} e^{-30} \simeq 0.712.. 10^{-5} \end{aligned}$$

qui doit nécessairement être le même résultat que (6.1), mais nécessite beaucoup moins de calcul !

5. On a par exemple :

$$IP[N(10) = 0] = \frac{(\lambda 10)^0}{0!} e^{-\lambda 10} = e^{-5}$$

6. On peut ici utiliser le fait que le nombre total d'arrivées $N(60) = 20$ sur l'intervalle $[0, 60]$ est connu. Dans ce cas, on sait que les instants d'arrivées A_n sont des variables aléatoires indépendantes distribuées de manière uniforme sur $[0, 60]$, c'est-à-dire selon la loi uniforme :

$$f_{A_n}(t) = \begin{cases} \frac{1}{60} & \text{si } t \in [0, 60] \\ 0 & \text{sinon} \end{cases}$$

La probabilité que 10 étoiles arrivent dans l'intervalle $[15, 30]$ (et que les autres arrivent en dehors de cet intervalle) vaut donc

$$C_{20}^{10} \left(\frac{1}{4}\right)^{10} \left(\frac{3}{4}\right)^{10} \simeq .0099..$$

où

$$C_{20}^{10} = \binom{20}{10} = \frac{20!}{10!10!}$$

est le nombre de choix possibles pour les 10 étoiles (parmi 20) qui arrivent dans le deuxième quart d'heure et

$$\left(\frac{1}{4}\right)^{10} \left(\frac{3}{4}\right)^{10}$$

est la probabilité, pour l'un de ces choix possibles, que les 10 étoiles choisies arrivent pendant le deuxième quart d'heure.

Exercice 2

On suppose que des foies arrivent à l'hôpital suivant un processus de Poisson de paramètre λ . Deux malades attendent d'être greffés. Leurs durées de vie suivent des lois exponentielles indépendantes entre elles (et indépendantes du processus d'arrivée des foies), respectivement de paramètres λ_1 et λ_2 . Le premier foie va au premier patient, si celui est encore en vie ; sinon au second (s'il est encore en vie...).

1. Calculer la probabilité que le premier patient soit greffé
2. Calculer la probabilité que le second patient soit greffé

Correction de l'Exercice 2

1. Notons D_1 la variable aléatoire "durée de vie du patient 1" et T_1 l'intervalle entre l'instant 0 et l'arrivée du premier foie. Par hypothèse, D_1 et T_1 sont deux variables aléatoires indépendantes, de distributions exponentielles, respectivement de paramètres λ_1 et λ . La distribution conjointe du couple (D_1, T_1) est donc :

$$f_{(D_1, T_1)}(x, y) = f_{D_1}(x)f_{T_1}(y) = \begin{cases} \lambda_1 e^{-\lambda_1 x} \lambda e^{-\lambda y} & \text{si } x \text{ et } y \geq 0 \\ 0 & \text{sinon} \end{cases}$$

La probabilité que le premier patient soit greffé s'écrit donc :

$$\begin{aligned} IP [D_1 \geq T_1] &= \iint_{x \geq y} f_{(D_1, T_1)}(x, y) dx dy \\ &= \int_0^\infty \int_0^x \lambda_1 e^{-\lambda_1 x} \lambda e^{-\lambda y} dy dx \\ &= \int_0^\infty \lambda_1 e^{-\lambda_1 x} (1 - e^{-\lambda x}) dx \\ &= \frac{\lambda}{\lambda + \lambda_1} = \frac{1}{1 + \frac{\lambda_1}{\lambda}} \end{aligned}$$

Cette probabilité tend vers 0 lorsque l'espérance de vie du patient 1 (en l'absence de greffe) tend vers 0 (i.e. $\frac{1}{\lambda_1} \rightarrow 0$), le paramètre $\lambda > 0$ étant par ailleurs fixé. Elle tend vers 1 lorsque la durée moyenne entre deux arrivées de foies à l'hôpital tend vers 0 (i.e. $\frac{1}{\lambda} \rightarrow 0$), le paramètre $\lambda_1 > 0$ étant par ailleurs fixé.

2. Le patient 2 sera greffé dans deux cas :

- lorsque le patient 1 meurt avant l'arrivée du foie 1, alors que le patient 2 survit jusqu'à cette arrivée du foie 1
- le patient 1 est greffé avec le foie 1 et le patient 2 survit jusqu'à l'arrivée du deuxième foie

La probabilité totale de survie du patient 2 est la somme de ces deux probabilités (relatives à deux événements disjoints). La probabilité que le patient 1 meure avant l'arrivée du foie 1 et que le patient 2 survive jusqu'à cette arrivée du foie 1 s'écrit ² (voir calcul précédent) :

$$\begin{aligned}
 IP[\{D_1 < T_1\} \cap \{D_2 \geq T_1\}] &= IP[D_1 < T_1] \cdot IP[D_2 \geq T_1] \\
 &= \left(1 - \frac{\lambda}{\lambda + \lambda_1}\right) \cdot \frac{\lambda}{\lambda + \lambda_2} \\
 &= \frac{\lambda_1 \lambda}{(\lambda + \lambda_1)(\lambda + \lambda_2)} \\
 &= \frac{1}{1 + \frac{\lambda}{\lambda_1}} \cdot \frac{1}{1 + \frac{\lambda_2}{\lambda}}
 \end{aligned}$$

La probabilité que le patient 1 soit greffé avec le foie 1 et que le patient 2 survive jusqu'à l'arrivée du deuxième foie s'écrit :

$$\begin{aligned}
 IP[\{D_1 \geq T_1\} \cap \{D_2 \geq (T_1 + T_2)\}] &= IP[D_1 \geq T_1] \cdot IP[D_2 \geq (T_1 + T_2)] \\
 &= \frac{\lambda}{\lambda + \lambda_1} \cdot IP[D_2 \geq (T_1 + T_2)] \\
 &= \frac{1}{1 + \frac{\lambda_1}{\lambda}} \cdot IP[D_2 \geq (T_1 + T_2)]
 \end{aligned}$$

avec $T_1 + T_2 =: A_2 \sim \Gamma(2, \lambda)$ (somme de deux variables aléatoires exponentielles indépendantes de même paramètre λ) et

$$\begin{aligned}
 IP[D_2 \geq A_2] &= \iint_{x \geq y} f_{(D_2, A_2)}(x, y) dx dy \\
 &= \int_0^\infty \int_0^x \lambda_2 e^{-\lambda_2 x} \lambda^2 y e^{-\lambda y} dy dx \\
 &= \int_0^\infty \lambda_2 e^{-\lambda_2 x} \left(1 - e^{-\lambda x} - \lambda x e^{-\lambda x}\right) dx \\
 &= 1 - \frac{\lambda_2}{\lambda + \lambda_2} + \frac{\lambda \lambda_2}{(\lambda + \lambda_2)^2} = 1 - \left(\frac{1}{1 + \frac{\lambda}{\lambda_2}}\right)^2
 \end{aligned}$$

La probabilité totale de survie du patient 2 s'écrit alors :

$$\frac{\lambda^3 + 2\lambda^2\lambda_2 + \lambda^2\lambda_1 + \lambda\lambda_1\lambda_2}{(\lambda + \lambda_1)(\lambda + \lambda_2)^2}$$

2. Nous noterons D_2 la durée de vie du patient 2, distribuée suivant une exponentielle de paramètre λ_2

Exercice 3 Les élèves de 1^{ère} année IR&C (ils sont 50 en tout) arrivent en retard à un cours en suivant un processus de Poisson de paramètre $\lambda = 4$. On compte le temps en minutes.

1. Ce jour là le professeur est de bonne humeur, il se contente de faire une remarque chaque fois que 3 élèves sont passés. On note Z_n le temps qui sépare la $(n-1)^{ième}$ remarque de la $n^{ième}$. Calculer la loi de Z_n
2. Ce jour là le professeur est de mauvaise humeur, il décide de ne plus accepter les élèves se présentant avec plus de 5 min de retard. On note Z le nombre d'élèves refusés au cours. Calculer la loi de Z

Correction de l'Exercice 3

1. Notons T_n l'intervalle entre l'instant d'arrivée de l'élève $n-1$ et l'instant d'arrivée de l'élève n ($n \geq 1$). On a donc :

$$Z_n = T_{3n-2} + T_{3n-1} + T_{3n}, \forall n \geq 1$$

Les variables Z_n sont distribuées selon des lois $\Gamma(3, \lambda)$ (somme de trois exponentielles indépendantes de même paramètres λ). On a donc :

$$f_{Z_n}(t) = \begin{cases} \frac{\lambda^3}{2!} e^{-\lambda t} (\lambda t)^2 & \text{si } t \geq 0 \\ 0 & \text{sinon} \end{cases}$$

2. La variable Z est une variable aléatoire discrète qui peut prendre à priori toutes les valeurs entières entre 0 et 50. Si nous notons $N(t)$ le nombre d'élèves arrivés à la minute t (processus de Poisson), la distributions de Z s'écrit :

$$\begin{aligned} IP[Z = k] &= IP[N(5) = 50 - k] \\ &= \frac{(5\lambda)^{50-k} e^{-5\lambda}}{(50-k)!} \\ &= \frac{(20)^{50-k} e^{-20}}{(50-k)!}, \forall k \in \{0, \dots, 50\} \end{aligned}$$

Exercice 4 Robert a trouvé un job de standardiste, on suppose que les appels téléphoniques se succèdent à ce standard selon un processus de Poisson de paramètre λ . De 8h à 12h, il y a eu 10 appels cette matinée

1. Calculer la probabilité que 5 appels arrivent dans la première heure, par deux méthodes.
2. Robert s'est absenté de 10h à 10h 15 pour draguer sa collègue. Quelle est la probabilité pour que Robert aie raté un appel ?

Correction de l'Exercice 4

1. Nous allons calculer la probabilité demandée de deux manières. La première méthode se basera sur le fait que le nombre d'appels sur la durée totale est connu et que les instants de ces appels sont en conséquence distribués de manière uniforme sur l'intervalle $[8, 12]$. La deuxième méthode consiste à appliquer directement la définition du processus de Poisson et la calcul des probabilités conditionnelles. Nous verrons que nous obtenons le même résultat avec les deux méthodes.

— **Première méthode :** Le nombre total d'appels reçus sur l'intervalle $[8 - 12]$ étant connu (10 appels), nous savons que les instants de ces appels sont distribués selon des variables uniformes indépendantes. L'instant de l'appel k est donc une variable aléatoire continue de densité :

$$f_{A_k}(t) = \begin{cases} \frac{1}{4} & \text{si } 8 \leq t \leq 12 \\ 0 & \text{sinon} \end{cases}, \forall k \in \{1, \dots, 10\}$$

La probabilité que l'appel $k = 1$ tombe pendant la première heure est donc :

$$\mathbb{P}[8 \leq A_1 \leq 9] = \int_8^9 f_{A_k}(t) dt = \frac{1}{4}$$

La probabilité que les appels $k \in \{1, \dots, 5\}$ tombent pendant la première heure et que les appels $k \in \{6, \dots, 10\}$ tombent entre 9h et 12h vaut

$$\left(\frac{1}{4}\right)^5 \left(\frac{3}{4}\right)^5$$

car tous les instants d'appel sont distribués selon des variables aléatoires (uniformes) indépendantes les unes des autres. Pour trouver la probabilité demandée, il reste à comptabiliser le nombre de choix possibles des 5 appels qui tombent la première heure parmi les 10 appels possibles. La probabilité associée à chacun de ces choix est la même que précédemment. La probabilité totale est donc :

$$C_{10}^5 \left(\frac{1}{4}\right)^5 \left(\frac{3}{4}\right)^5 \simeq 0.058..$$

avec

$$C_{10}^5 = \binom{10}{5} = \frac{10!}{5!(10-5)!}$$

- **deuxième méthode (directe) :** on note $N(t)$ le nombre d'appels reçus au temps t où $t \in [8, 12]$ est exprimé en heures. Par hypothèse,

$N(t)$ est un processus de Poisson de paramètre λ . La meilleure estimation possible pour la valeur de λ est donnée par la valeur de l'estimateur $\frac{N(T)}{T}$, ici $\hat{\lambda} = \frac{10}{4} = 2.5h^{-1}$. La probabilité d'avoir 5 appels la première heure, sachant qu'il y a eu 10 appels pendant les quatre heures, s'écrit donc :

$$\begin{aligned} IP[N(1) = 5 | N(4) = 10] &= \frac{IP[N(1) = 5; N(4) = 10]}{IP[N(4) = 10]} \\ &= \frac{IP[N(1) = 5; N(4) - N(1) = 5]}{IP[N(4) = 10]} \\ &= \frac{IP[N(1) = 5] \cdot IP[N(4) - N(1) = 5]}{IP[N(4) = 10]} \\ &= \frac{\lambda^5 e^{-\lambda}}{5!} \cdot \frac{(3\lambda)^5 e^{-3\lambda}}{5!} \\ &= \frac{(4\lambda)^{10} e^{-4\lambda}}{10!} \\ &= \frac{10!}{5!5!} \left(\frac{1}{4}\right)^5 \left(\frac{3}{4}\right)^5 \end{aligned}$$

2. On calcule la probabilité complémentaire que Robert n'ait manqué aucun appel, c'est-à-dire que les dix appels tombent en dehors de l'intervalle $[10, 10.25]$. Cette dernière vaut :

$$\begin{aligned} IP\left[\bigcap_{k \in \{1, \dots, 10\}} A_k \notin [10, 10.25]\right] &= \prod_{k \in \{1, \dots, 10\}} IP[A_k \notin [10, 10.25]] \\ &= \left(\frac{15}{16}\right)^{10} \end{aligned}$$

La probabilité que Robert manque **au moins un appel** vaut donc :

$$1 - \left(\frac{15}{16}\right)^{10}$$

Exercice 5

Un signal $X(t)$ ne peut prendre que les valeurs 0 ou 1. Les instants auxquels il change de valeur définissent un processus de Poisson de paramètre λ . On suppose qu'à l'instant $t = 0$ le signal vaut 1. Calculer $P(X(t) = 1)$

Correction de l'Exercice 5

Le signal vaut 1 en t si le nombre $N(t)$ d'arrivées du processus de Poisson au temps t est pair. On a donc :

$$IP[X(t) = 1] = \sum_{k=0}^{\infty} IP[N(t) = 2k]$$

Comme $N(t)$ est un processus de Poisson de taux λ , on a :

$$IP[X(t) = 1] = \sum_{k=0}^{\infty} e^{-\lambda t} \frac{(\lambda t)^{2k}}{(2k)!}$$

Le développement en série de la fonction cosinus hyperbolique

$$\cosh(x) = \frac{e^x + e^{-x}}{2} = \sum_{k=0}^{\infty} \frac{x^{2k}}{(2k)!}$$

converge pour tout $x > 0$ et permet de conclure

$$IP[X(t) = 1] = e^{-\lambda t} \frac{e^{\lambda t} + e^{-\lambda t}}{2} = \frac{1 + e^{-2\lambda t}}{2}$$

On notera en particulier que lorsque t augmente, la valeur du signal au temps $t = 0$ est progressivement oubliée :

$$\lim_{t \rightarrow \infty} IP[X(t) = 1] = \frac{1}{2}$$

Exercice 6

Le nombre de personnes franchissant une barrière automatique de métro à l'heure de pointe est modélisé par un processus de Poisson. Il arrive une personne toutes les 2 secondes en moyenne. Parmi les personnes certaines fraudent en n'ayant pas de ticket. La probabilité pour qu'une personne soit un fraudeur est $p = 0.1$.

1. Modéliser cette situation
2. Quelle est la probabilité qu'au moins une personne fraude dans un intervalle d'une minute ?
3. Sachant que 200 personnes sont passées en 12 minutes, quel est le nombre moyen de fraudeurs ?
4. 60 personnes sont passées en 8 minutes. Quelle est la probabilité que 5 fraudeurs soient passés dans les 5 premières minutes
5. Chaque fraudeur fait perdre 1 euro à la compagnie de transports. Combien la compagnie perd-elle en moyenne à cette barrière en une heure ?

Correction de l'Exercice 6

1. On appelle $N(t)$ le nombre d'arrivées au temps t , mesuré en secondes $[s]$. $\{N(t)\}_{t \geq 0}$ est un processus de Poisson de paramètre

$$\lambda = \frac{1}{2} [s^{-1}]$$

On décompose ce processus en deux sous-processus indépendants. $N_1(t)$ désignera le nombre d'arrivées de fraudeurs au temps t . $N_2(t)$ désignera le nombre d'arrivées de clients en règle au temps t . $\{N_1(t)\}_{t \geq 0}$ et $\{N_2(t)\}_{t \geq 0}$ sont donc aussi deux processus de Poisson, respectivement de paramètre $\lambda_1 := \lambda p$ et $\lambda_2 := \lambda(1 - p)$.

2. Il s'agit de calculer :

$$\begin{aligned} \mathbb{P}[N_1(60) \geq 1] &= 1 - \mathbb{P}[N_1(60) = 0] \\ &= 1 - \frac{(\lambda_1 t)^0 e^{-\lambda_1 t}}{0!} \Big|_{t=60} \\ &= 1 - e^{-60\lambda p} = 1 - e^{-3} \end{aligned}$$

3. Parmi les 200 personnes arrivées, la probabilité qu'il y ait k fraudeurs est distribuée selon la loi binomiale

$$\mathbb{P}[N_1(720) = k | N(720) = 200] = C_{200}^k p^k (1 - p)^{200-k}$$

où

$$C_{200}^k = \binom{200}{k} = \frac{200!}{k!(200 - k)!}$$

est le nombre de choix possibles de k clients parmi les 200 arrivés et

$$p^k (1 - p)^{200-k}$$

est la probabilité, pour l'un de ces choix possibles, que les k choisis soient des fraudeurs et les autres non. On sait que l'espérance d'une variable binomiale $X \sim B(n, p)$ vaut $E(X) = np$. On a donc :

$$E(\mathbb{P}[N_1(720) = k | N(720) = 200]) = 200 \cdot p = 20,$$

résultat qui était évident dès le départ.

4. Attention : on connaît le nombre total de clients arrivés $N(480) = 60$, mais la question porte sur le nombre de fraudeurs sur une période différente, $N_1(300)$. Nous allons donc décomposer le calcul selon la formule des probabilités totales. En appelant :

$$\tilde{P} := \mathbb{P}[N_1(300) = 5 | N(480) = 60]$$

on trouve

$$\tilde{P} = \sum_{j=5}^{60} \mathbb{P}[N_1(300) = 5 | N(480) = 60; N(300) = j] \cdot \mathbb{P}[N(300) = j | N(480) = 60]$$

avec (distribution uniforme des instants d'arrivées sur $[0, 480]$)

$$\mathbb{P}[N(300) = j | N(480) = 60] = C_{60}^j \left(\frac{300}{480}\right)^j \left(\frac{180}{480}\right)^{60-j} = C_{60}^j \left(\frac{5}{8}\right)^j \left(\frac{3}{8}\right)^{60-j}$$

et

$$\begin{aligned} \mathbb{P}[N_1(300) = 5 | N(480) = 60; N(300) = j] &= \mathbb{P}[N_1(300) = 5 | N(300) = j] \\ &= C_j^5 \left(\frac{1}{10}\right)^5 \left(\frac{9}{10}\right)^{j-5} \end{aligned}$$

On a donc :

$$\begin{aligned} \tilde{P} &= \sum_{j=5}^{60} C_{60}^j \left(\frac{5}{8}\right)^j \left(\frac{3}{8}\right)^{60-j} C_j^5 \left(\frac{1}{10}\right)^5 \left(\frac{9}{10}\right)^{j-5} \\ &= \sum_{j=5}^{60} \frac{60!j!}{j!(60-j)!5!(j-5)!} \left(\frac{5}{8}\right)^j \left(\frac{3}{8}\right)^{60-j} \left(\frac{1}{10}\right)^5 \left(\frac{9}{10}\right)^{j-5} \\ &= \left(\frac{5}{80}\right)^5 C_{60}^5 \sum_{k=0}^{55} C_{55}^k \left(\frac{9}{16}\right)^k \left(\frac{3}{8}\right)^{55-k} \\ &= \left(\frac{1}{16}\right)^{60} C_{60}^5 \sum_{k=0}^{55} C_{55}^k 9^k 6^{55-k} \\ &= \frac{15^{55}}{16^{60}} C_{60}^5 \simeq 0.14967.. \end{aligned}$$

5. Il s'agit de déterminer l'espérance du nombre de fraudeurs en une heure (on sait à priori que la réponse est 180!). La loi du nombre de fraudeurs en une heure est la loi de Poisson

$$\mathbb{P}[N_1(3600) = k] = \left[\frac{(\lambda p t)^k}{k!} e^{-\lambda p t} \right] \Big|_{t=3600} = \frac{180^k}{k!} e^{-180}, k \geq 0$$

de paramètre $\tilde{\lambda} = 180$. Or, l'espérance d'une variable de Poisson $X \sim \text{Poi}(\tilde{\lambda})$ est $\tilde{\lambda}$. On a donc :

$$E(N_1(3600)) = 180$$

Exercice 7

On suppose que les clients d'un magasin arrivent dans un magasin selon un processus de Poisson au rythme de 3 par minute. Le temps de séjour d'un client dans le magasin suit une loi exponentielle de paramètre $\mu = 0.1$. Le temps que passe chaque client dans le magasin est indépendant du temps qu'y passent les autres clients. On note $N(T)$ le nombre de clients encore dans le magasin à l'instant $T = 30$ min. On note $N_p(T)$ le nombre de clients partis à l'instant T . Donner la loi de ces deux variables aléatoires.

Correction de l'Exercice 7

Soit $N_a(t)$ le nombre de clients arrivés dans le magasin au temps t . C'est un processus de Poisson de paramètre $\lambda = 3 \text{ min}^{-1}$. Soit $T_s \sim \exp \mu$ (avec $\mu = 0.1 \text{ min}^{-1}$) la variable aléatoire exponentielle décrivant le temps de séjour des clients dans le magasin. En supposant qu'il n'y a pas de client dans le magasin au temps $t = 0$, on a :

$$N_a(t) = N(t) + N_p(t)$$

Nous allons dans cet exercice redémontrer un résultat mentionné en cours sur la décomposition du processus de Poisson $N_a(t)$ aux deux sous processus, $N(t)$ et $N_p(t)$, qui correspondent aux deux sous-classes des clients arrivés au temps t qui seront respectivement encore dans le magasin ou partis au temps T . Dans ce cas, les probabilités d'appartenance des clients à l'une ou l'autre de ces sous-classes varient avec le temps. Appelons $p(t)$ la probabilité qu'un client arrivé au temps $t \in [0, T]$ soit encore dans le magasin au temps $T \geq t$. Cette probabilité s'écrit :

$$p(t) := \mathbb{P}[T_s \leq T - t] = \int_0^{T-t} \mu e^{-\mu x} dx = 1 - e^{-\mu(T-t)}$$

D'après le cours, nous savons que dans ce cas la loi conditionnelle de $N(T)$ (respectivement $N_p(T)$), le nombre d'arrivées $N_a(T) = n$ étant fixé, est distribuée selon une loi binomiale³ $B(n, p)$ (respectivement $B(n, 1 - p)$) de paramètre :

$$\bar{p} := \frac{1}{T} \int_0^T p(t) dt = 1 - \frac{1}{\mu T} (1 - e^{-\mu T}) \quad (6.2)$$

En utilisant la formule des probabilités totales, on a donc :

$$\begin{aligned} \mathbb{P}[N(T) = k] &= \sum_{n \geq k} \mathbb{P}[N(T) = k | N_a(T) = n] \cdot \mathbb{P}[N_a(T) = n] \\ &= \sum_{n \geq k} \binom{n}{k} \bar{p}^k (1 - \bar{p})^{n-k} \frac{(\lambda T)^n e^{-\lambda T}}{n!} \\ &= \frac{(\bar{p} \lambda T)^k e^{-\lambda T}}{k!} \sum_{n \geq k} \frac{((1 - \bar{p}) \lambda T)^{n-k}}{(n - k)!} \\ &= \frac{(\bar{p} \lambda T)^k e^{-\lambda T}}{k!} \sum_{m \geq 0} \frac{((1 - \bar{p}) \lambda T)^m}{m!} \\ &= \frac{(\bar{p} \lambda T)^k e^{-\lambda T}}{k!} e^{(1 - \bar{p}) \lambda T} \\ &= \frac{(\bar{p} \lambda T)^k}{k!} e^{-\bar{p} \lambda T} \end{aligned} \quad (6.3)$$

3. ce résultat est redémontré, pour information, dans la remarque 15 à la fin de la correction de l'exercice

La variable $N(T)$ est donc distribuée selon une loi de Poisson de paramètre $\bar{p}\lambda T$ où \bar{p} est la probabilité moyenne sur $[0, T]$ définie en (6.2). Attention, le processus $\{N(T)\}_{T>0}$ n'en est pas pour autant un processus de Poisson homogène puisque la probabilité $\bar{p} = \bar{p}(T)$ dépend explicitement du temps T . La variable $N_p(T)$ est elle aussi distribuée selon une loi de Poisson de paramètre $(1 - \bar{p})\lambda T$ (même raisonnement que pour $N(T)$).

Remarque 15 : si le nombre total d'arrivées sur $[0, T]$ est connu (notons $N_a(T) = n$), alors les instants de ces arrivées sont distribués selon des lois uniformes indépendantes (notées $A_k, \forall k \in \{1, \dots, n\}$). Les temps de séjour des n clients arrivés étant distribués selon des variables aléatoires exponentielles (indépendantes entre elles et indépendantes des instants d'arrivées), la probabilité pour le client k (n'importe lequel des clients arrivés sur $[0, T]$) d'être encore dans le magasin au temps T est $P(S \leq T - A_k)$. Or $T - A_k$ est elle aussi distribuée de manière uniforme sur l'intervalle $[0, T]$. On a donc

$$IP[T_s \leq T - A_k] = \iint_{x \leq y} \mu e^{-\mu x} \cdot \frac{1}{T} dx dy = 1 - \frac{1 - e^{-\mu T}}{\mu T}$$

C'est précisément la probabilité que nous avons notée précédemment \bar{p} . La probabilité d'avoir k clients encore présents au temps T , parmi les n clients arrivés sur $[0, T]$, s'écrit donc :

$$IP[N(T) = k | N_a(T) = n] = \binom{n}{k} \bar{p}^k (1 - \bar{p})^{n-k} \quad (6.4)$$

car il y a $\binom{n}{k}$ choix possibles des k clients qui sont encore dans le magasin au temps T parmi les n clients arrivés sur $[0, T]$ et - pour chacun de ces choix possibles - une probabilité $\bar{p}^k (1 - \bar{p})^{n-k}$.

Remarque 16 : les résultats (6.3) et (6.4) (qui utilisent la définition (6.2) de la probabilité moyenne \bar{p}) ont été vus en cours, mais les démonstrations détaillées dans cette correction avaient été admises. Elles sont développées ici pour votre information. Pour répondre à la question initiale, il aurait pu suffire de citer ces résultats sans les démontrer.

Exercice 8

On modélise le téléchargement d'un fichier par l'envoi de segments. Lorsqu'on télécharge un fichier, la transmission de ces segments peut être erronée.

La liaison téléphonique entre Robert et son fournisseur d'accès internet est mauvaise : on suppose que des erreurs de transmission se produisent selon un processus de Poisson. Le temps moyen entre deux erreurs de transmission est de 1 minute. Robert télécharge un fichier sur un site internet ; Entre le FAI de Robert et ce site les erreurs de transmission sont également modélisées par un processus de Poisson. Elles sont en moyenne espacées de 5 minutes.

1. Le téléchargement dure une demi heure. En moyenne, combien de segments sont erronés parmi ceux qu'a reçus Robert ?
2. Quelle est la probabilité que le téléchargement se déroule sans accroc ?

Correction de l'Exercice 8

Par hypothèses, nous supposons que les erreurs ne se compensent pas et que les segments erronés sont suffisamment peu nombreux pour négliger la probabilité de deux erreurs consécutives sur le même segment. Ainsi, le nombre total d'erreurs dans les segments reçus par Robert est la somme des erreurs de transmission entre le site distant et le FAI, d'une part, et entre le FAI et Robert, d'autre part. Soit $N_1(t)$ le nombre de segments erronés parmi ceux transmis par le site distant au FAI, au temps t et $N_2(t)$ le nombre de segments erronés parmi ceux transmis par le FAI à Robert, toujours au temps t . On a donc :

$$N(t) = N_1(t) + N_2(t)$$

où $N(t)$ désigne le nombre total de segments erronés transmis à Robert au temps t . $N(t)$ est donc un processus de Poisson de paramètre

$$\lambda = \lambda_1 + \lambda_2 = \lambda_1 = \frac{6}{5} \text{ min}^{-1}$$

où $\lambda_1 = \frac{1}{5} \text{ min}^{-1}$ et $\lambda_2 = 1 \text{ min}^{-1}$ désignent respectivement les paramètres des processus de Poisson $N_1(t)$ et $N_2(t)$ estimés avec le maximum de vraisemblance.

1. La distribution de probabilité du nombre total de segments erronés reçu en 30 min (distribuée selon une loi de Poisson de paramètre $\lambda t = \frac{6}{5}30 = 36$) :

$$\mathbb{P}[N(t) = k] = \frac{(\lambda t)^k e^{-\lambda t}}{k!} = \frac{36^k e^{-36}}{k!}$$

L'espérance du nombre de segments erronés sur ces 30 min vaut donc ⁴

$$E(N(30) = k) = \sum_{k \geq 0} k \mathbb{P}[N(t) = k] = 36$$

2. La probabilité que le téléchargement se déroule sans accroc vaut

$$\mathbb{P}[N(30) = 0] = e^{-\frac{6}{5}30} = e^{-36}$$

4. L'espérance d'une variable de Poisson de paramètre $\tilde{\lambda}$ est $\tilde{\lambda}$

Exercice 9

On suppose que les arrivées d'un bus à un arrêt suivent un processus de Poisson de paramètre λ .

Soit $t > 0$. Montrer que le temps moyen séparant le dernier autobus arrivé avant t et le premier après t est strictement plus grand que le temps moyen séparant 2 autobus ; et que si t est grand, il vaut approximativement le double.

Correction de l'Exercice 9

Il s'agit d'un nouvel exemple du paradoxe de l'inspection (voir section 2.3). Au temps $t > 0$, l'instant d'arrivée du dernier autobus est la variable aléatoire $A_{N(t)}$, tandis que l'instant d'arrivée du prochain autobus est $A_{N(t)+1}$. L'intervalle de temps séparant ces deux instants d'arrivées est donc :

$$T_{N(t)+1} = A_{N(t)+1} - A_{N(t)}$$

qui peut également s'écrire

$$T_{N(t)+1} = t - A_{N(t)} + A_{N(t)+1} - t = T_a(t) + T_r(t)$$

où les deux variables aléatoires $T_a(t)$ et $T_r(t)$ sont deux variables aléatoires indépendantes que nous avons appelées *âge* et *temps de vie résiduel* dans la section 2.3. On a en particulier, en ce qui concerne l'espérance :

$$E(T_{N(t)+1}) = E(T_a(t)) + E(T_r(t))$$

Le *temps de vie résiduel* T_r a une distribution exponentielle. On a donc :

$$E(T_r(t)) = \frac{1}{\lambda}$$

La distribution de la variable *âge* est donnée à la proposition 7. On a en particulier :

$$E(T_a(t)) = \int_0^t e^{-\lambda\xi} d\xi = \frac{1 - e^{-\lambda t}}{\lambda}$$

Au final, l'espérance du temps séparant l'arrivée du dernier bus de celle du prochain (inspection au temps t) vaut donc

$$E(T_{N(t)+1}) = \frac{1}{\lambda} + \frac{1 - e^{-\lambda t}}{\lambda} = \frac{2}{\lambda} - \frac{e^{-\lambda t}}{\lambda}$$

La durée de l'intervalle entre deux arrivées d'autobus inspectée à l'instant t est donc strictement supérieure à la durée moyenne de l'intervalle entre deux arrivées de bus en général. En effet, les bus arrivant selon un processus de Poisson de paramètre λ , le temps entre deux arrivées successives quelconques T_n est distribué selon une exponentielle de paramètre λ et son espérance vaut

$$E(T_n) = \frac{1}{\lambda} < \frac{2}{\lambda} - \frac{e^{-\lambda t}}{\lambda}$$

Exercice 10

Les employés d'une entreprise arrivent au travail selon un processus de Poisson $N(t)$ d'intensité λ . On note A_k l'instant d'arrivée du k^{ieme} employé.

1. Exprimer le nombre d'heures totales $X(t)$ de travail effectuées dans l'entreprise jusqu'à la date t en fonction de $N(t)$, et des A_k
2. On cherche à calculer $E(X(t))$
 - (a) Montrer que $E(X(t)) = \sum_{k=0}^{+\infty} E(X(t)|N(t) = n)IP[N(t) = n]$
 - (b) En déduire que $E(X(t)) = \frac{\lambda t^2}{2}$

Correction de l'Exercice 10

1. Un employé qui arrivé à l'instant A_k aura travaillé à l'instant $t > A_k$ une durée totale ⁵ $t - A_k$. Le nombre d'employés arrivés à l'instant t est $N(t)$. La durée totale travaillée par l'ensemble des employés arrivés au temps t s'écrit donc :

$$X(t) = \sum_{k=1}^{N(t)} (t - A_k)$$

La durée totale $X(t)$ est donc une fonction de la variable aléatoire discrète $N(t)$ et des variables aléatoires continues A_k

2. Le calcul direct de l'espérance $E(X(t))$ nécessite le calcul de l'espérance de la somme des variables aléatoires A_k qui ne sont pas, dans le cas général, indépendantes. Ce calcul peut toutefois se décomposer en fonction du nombre total d'arrivées sur $[0, t]$. En effet, si le nombre total d'arrivées sur $[0, t]$ est fixé (connu), alors les instant d'arrivées sont des variables aléatoires indépendantes de distribution uniforme, ce qui simplifie considérablement les calculs.
 - (a) La formule proposée n'est rien d'autre que le théorème de l'espérance conditionnelle

$$E(E(X|N)) = E(X) \quad (6.5)$$

appliqué à un couple de variables aléatoire (X, N) où X est une variable aléatoire continue et N une variable aléatoire discrète. Dans cette dernière équation, l'espérance conditionnelle $E(X|N)$ est la variable aléatoire discrète

$$E(X|N) = \varphi(N)$$

5. La durée $t - A_k$ ne désigne pas l'âge du client k car celui-ci n'est pas nécessairement le dernier arrivé immédiatement avant le temps t , mais peut être n'importe lequel des clients déjà arrivés au temps t

où la fonction $\varphi(n)$ est définie par

$$\varphi : \mathbb{N} \rightarrow \mathbb{R} : n \mapsto \varphi(n) := E^{\{N=n\}}(X) := \int_{\mathbb{R}} \xi f_X^{\{N=n\}}(\xi) d\xi$$

où $f_X^{\{N=n\}}(\xi)$ désigne la densité de probabilité conditionnelle

$$f_X^{\{N=n\}}(\xi) := \frac{f_{(X,N)}(\xi, n)}{IP[N = n]}$$

et où $f_{(X,N)}(\xi, n)$ désigne la loi du couple de variables aléatoires (X, N) . La démonstration du théorème de l'espérance conditionnelle (équation (6.5)) s'obtient par calcul direct. On a, par définition de l'espérance d'une variable aléatoire discrète :

$$\begin{aligned} E(E(X|N)) &= \sum_{n \geq 0} IP[N = n] E(X|N = n) \\ &= \sum_{n \geq 0} IP[N = n] \int_{\mathbb{R}} \xi f_X^{\{N=n\}}(\xi) d\xi \\ &= \sum_{n \geq 0} IP[N = n] \int_{\mathbb{R}} \xi \frac{f_{(X,N)}(\xi, n)}{IP[N = n]} d\xi \\ &= \int_{\mathbb{R}} \xi \left(\sum_{n \geq 0} f_{(X,N)}(\xi, n) \right) d\xi \\ &= \int_{\mathbb{R}} \xi f_X(\xi) d\xi \\ &= E(X) \end{aligned}$$

Pour tout $t > 0$, $X(t)$ est une variable aléatoire continue et $N(t)$ une variable aléatoire discrète (de Poisson). On a donc

$$E(X(t)) = \sum_{n \geq 0} E(X(t)|N(t) = n) IP[N(t) = n]$$

qui correspond exactement à la formule proposée dans l'énoncé

(b) On a donc :

$$\begin{aligned} E(X(t)) &= E(E(X(t)|N(t))) \\ &= \sum_{n \geq 0} IP[N(t) = n] E(X(t)|N(t) = n) \\ &= \sum_{n \geq 0} \frac{(\lambda t)^n e^{-\lambda t}}{n!} E\left(\sum_{k=1}^n (t - A_k)\right) \\ &= \sum_{n \geq 0} \frac{(\lambda t)^n e^{-\lambda t}}{n!} \sum_{k=1}^n E(t - A_k) \end{aligned}$$

Or, n étant fixé, les instants d'arrivées A_k sont des variables aléatoires indépendantes distribuées selon la loi uniforme sur $[0, t]$ et :

$$\begin{aligned} E(t - A_k) &= \int_{\mathbb{R}} f_{U([0,t])}(\xi)(t - \xi) d\xi \\ &= \int_0^t \frac{1}{t}(t - \xi) d\xi \\ &= \frac{t}{2} \end{aligned}$$

On obtient finalement

$$\begin{aligned} E(X(t)) &= \sum_{n \geq 0} \frac{(\lambda t)^n e^{-\lambda t}}{n!} \frac{nt}{2} \\ &= \frac{t}{2} \sum_{n \geq 0} n \frac{(\lambda t)^n e^{-\lambda t}}{n!} \\ &= \frac{\lambda t^2}{2} \end{aligned}$$

6.2 Chaînes de Markov à temps discret

Exercice 11

Les études dans une école d'ingénieurs durent 3 ans⁶. A l'issue de chaque année, un étudiant peut redoubler avec la probabilité r , passer dans la classe suivante avec la probabilité p , ou être exclu avec la probabilité e . On supposera dans un premier temps que ces probabilités ne dépendent ni de l'année d'étude de l'étudiant, ni du nombre de redoublements éventuels de celui-ci ... Ces hypothèses pourront être levées ensuite pour proposer un modèle plus réaliste !

1. Modéliser le cursus d'un élève par une chaîne de Markov à cinq états.
2. Quatre ans viennent de s'écouler depuis que Robert est entré à l'école. Que peut on savoir sur l'année d'étude dans laquelle se trouve Robert ?
3. Calculer la probabilité qu'un élève, venant de rentrer dans l'école, arrive à avoir son diplôme. Même question avec un élève de deuxième, puis de troisième année.
4. Calculer la durée moyenne que passe un étudiant à l'école, avant l'obtention du diplôme ou son renvoi, en supposant qu'il entre en première année.

6. Toute ressemblance avec des établissements actuels ou ayant existé serait purement fortuite et ne pourrait être que le fruit d'une pure coïncidence...

5. Proposez une modélisation plus réaliste du cursus “ingénieur” à L’ESI-SAR dans laquelle un seul redoublement est permis.

Correction de l’Exercice 11

1. Soit $X_n \in E := \{1, 2, 3, d, e\}$, l’année d’étude dans laquelle se trouve un étudiant au cours de sa n -ième année passée dans l’école. L’état d correspond à l’état *diplômé* et est absorbant, de même que l’état e qui correspond à l’état *exclu*. Le parcours d’un étudiant dans cette école est modélisé comme une CMTD représentée à la figure 6.1. Les probabilités p , r et e satisfont les contraintes

$$r, p, e \in [0, 1]$$

$$p + r + e = 1$$

La matrice de passage associée à la chaîne de Markov représentée à la figure 6.1 est

$$\mathbf{P} = \begin{pmatrix} r & p & 0 & 0 & e \\ 0 & r & p & 0 & e \\ 0 & 0 & r & p & e \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

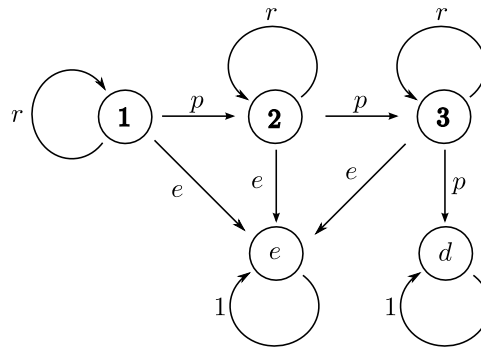


FIGURE 6.1 – Le graphe associé à la CMTD qui modélise le parcours d’un élève dans une école d’ingénieur, dans l’exercice 11, parties 1 à 4

2. En utilisant l’approche algébrique, on obtient :

$$P(X_4 = j | X_0 = 1) = (\mathbf{P}^4)_{1,j}, \forall j \in E := \{1, 2, 3, d, e\}$$

Le calcul de \mathbf{P}^4 étant un peu fastidieux, on peut aussi utiliser le graphe de la CMTD et la formule de probabilité d’un chemin. On a ainsi, par

exemple

$$P(X_4 = 1 | X_0 = 1) = P(1 \rightarrow 1 \rightarrow 1 \rightarrow 1 \rightarrow 1) = r^4$$

car il n'y a qu'un seul chemin possible, le chemin $1 \rightarrow 1 \rightarrow 1 \rightarrow 1 \rightarrow 1$, pour partir de l'année 1 et après 4 ans écoulés, y être toujours. Par contre, il y a quatre chemins différents pour aller de l'année 1 à l'année 2 en quatre ans : les chemins $1 \rightarrow 1 \rightarrow 1 \rightarrow 1 \rightarrow 2$, $1 \rightarrow 1 \rightarrow 1 \rightarrow 2 \rightarrow 2$, $1 \rightarrow 1 \rightarrow 2 \rightarrow 2 \rightarrow 2$ et $1 \rightarrow 2 \rightarrow 2 \rightarrow 2 \rightarrow 2$. Chacun de ces chemins correspond à un passage et trois redoublements. On a donc :

$$P(X_4 = 2 | X_0 = 1) = 4pr^3$$

Pour aller en quatre ans de l'année 1 à l'année 3, il faut deux passage et deux redoublements. On a donc :

$$P(X_4 = 3 | X_0 = 1) = C_4^2 p^2 r^2 = 6p^2 r^2$$

où C_4^2 est le nombre de choix possibles des deux années réussies parmi les 4. On obtient par ailleurs pour la probabilité d'être diplômé après quatre ans :

$$P(X_4 = d | X_0 = 1) = p^3 + 3rp^3$$

qui correspond soit à la réussite des trois premières années sans redoublement, soit à un éventuel unique redoublement de la première, de la deuxième ou de la troisième année. Enfin, la probabilité de finir exclu est la probabilité complémentaire :

$$P(X_4 = e | X_0 = 1) = 1 - r^4 - 4pr^3 - 6p^2 r^2 - p^3 - 3rp^3 = (1 - e)^4 + ep^3$$

3. On demande ici les probabilités $f_{1,d}$, $f_{2,d}$ et $f_{3,d}$. En utilisant les formules du cours, on obtient :

$$\begin{aligned} f_{1,d} &= p_{1,d} + p_{1,1}f_{1,d} + p_{1,2}f_{2,d} + p_{1,3}f_{3,d} + p_{1,e}f_{e,d} \\ &= p_{1,1}f_{1,d} + p_{1,2}f_{2,d} \\ &= rf_{1,d} + pf_{2,d} \end{aligned}$$

De même, on trouve

$$f_{2,d} = rf_{2,d} + pf_{3,d}$$

et

$$f_{3,d} = p + rf_{3,d}$$

En résolvant ces trois dernière équations, on obtient :

$$\begin{aligned} f_{3,d} &= \frac{p}{1-r} \\ f_{2,d} &= \frac{p^2}{(1-r)^2} \\ f_{1,d} &= \frac{p^3}{(1-r)^3} \end{aligned}$$

4. On demande ici les temps moyens d'entrée dans la sous-chaîne $F := \{e, d\}$, en partant de l'état $i \in \{1, 2, 3\}$, que nous noterons \mathbf{T}_i . En appliquant la formule vue en cours, on trouve :

$$\begin{pmatrix} \mathbf{T}_1 \\ \mathbf{T}_2 \\ \mathbf{T}_3 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + \begin{pmatrix} r & p & 0 \\ 0 & r & p \\ 0 & 0 & r \end{pmatrix} \begin{pmatrix} \mathbf{T}_1 \\ \mathbf{T}_2 \\ \mathbf{T}_3 \end{pmatrix}$$

On trouve pour la solution :

$$\begin{aligned} \mathbf{T}_3 &= \frac{1}{1-r} \\ \mathbf{T}_2 &= \frac{1}{1-r} \left(1 + \frac{p}{1-r} \right) \\ \mathbf{T}_1 &= \frac{1}{1-r} \left(1 + \frac{p}{1-r} \left(1 + \frac{p}{1-r} \right) \right) \end{aligned}$$

5. Les probabilités de transition dépendent dans ce cas, non seulement de l'année d'étude dans laquelle se trouve l'élève, mais également du fait qu'il ait (ou non) déjà redoublé. Ainsi, si nous voulons maintenir la propriété markovienne, nous devons dédoubler les états $\{1, 2, 3\}$, pour distinguer les cas des élèves redoublants des autres. Le nouvel espace d'état est donc $E := \{1, 1_r, 2, 2_r, 3, 3_r, d, e\}$. Avec cet espace d'état étendu pour la chaîne, les probabilités de transition d'un état $i \in E$ vers un état $j \in E$ ne dépendent plus que de l'état initial et de l'état final. La propriété markovienne est donc restaurée. Le parcours d'un étudiant dans cette école est modélisé par la CMTD représentée à la figure 6.2. Les probabilités de transition des états $i \in \{1, 2, 3\}$ vers l'état e valent $e + r$, la probabilité d'être exclu ou de redoubler (ce qui pour un redoublant équivaut à une exclusion).

Exercice 12

On dispose de 2 machines qui fonctionnent indépendamment l'une de l'autre. Au cours d'une journée chaque machine peut tomber en panne avec la probabilité $1 - p$.

1. On suppose que l'on peut réparer la nuit une seule machine qui marchera ainsi le lendemain. Modéliser le système par une chaîne de Markov (donner son graphe, sa matrice)
2. Même question si la réparation ne se fait pas pendant la nuit mais le lendemain.
3. Calculer dans le deuxième cas (réparations en journée) la distribution stationnaire de probabilités. Quelle est la probabilité pour un ouvrier qui arrive le matin de se retrouver en chômage technique dans le deuxième cas ?

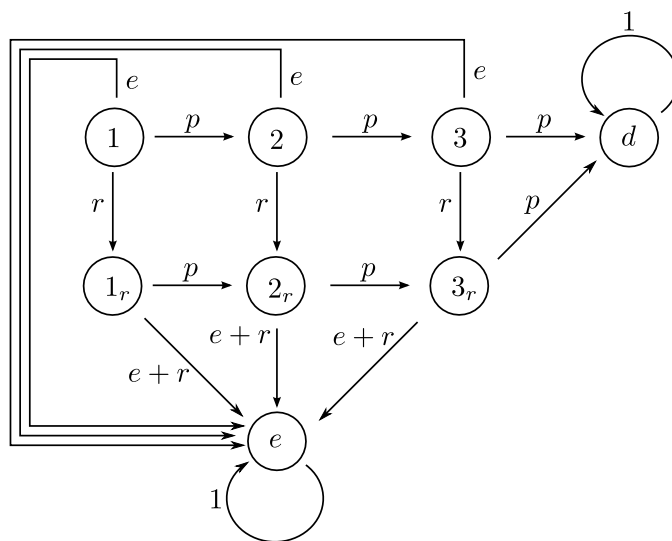


FIGURE 6.2 – Le graphe associé à la CMTD qui modélise le parcours d'un élève dans une école d'ingénieur, dans l'exercice 11, partie 5

Correction de l'Exercice 12

1. Soit $X_n \in E := \{1, 2\}$, le nombre de machine(s) qui fonctionne(nt) le matin du jour n . Le graphe associé à la CMTD qui modélise le fonctionnement de l'atelier est représenté à la figure 6.3 (avec $p \in [0, 1]$). La transition de 1 vers 1 suppose que la machine fonctionnant le jour n tombe en panne, tandis que l'autre est en réparation. Il n'y aura donc de nouveau qu'une seule machine qui fonctionne le jour $n+1$. La transition de 2 vers 1 suppose que les 2 machines qui fonctionnent le jour n tombent en panne. Une seule pourra être réparée pendant la nuit. La transition de 1 vers 2 suppose que la machine qui fonctionne le jour n ne tombe pas en panne. La probabilité de transition de 2 vers 2 est calculée comme la complémentaire des autres. Elle correspond soit au cas où aucune des deux machines ne tombe en panne, soit au cas où une seule tombe en panne.

La matrice de passage associée à cette chaîne de Markov est

$$\mathbf{P} = \begin{pmatrix} 1-p & p \\ (1-p)^2 & p(2-p) \end{pmatrix}$$

2. Si les deux machines sont en panne à la fin du jour $n-1$, l'une d'elle sera réparée le lendemain. Au matin du jour n , il n'y aura donc aucune

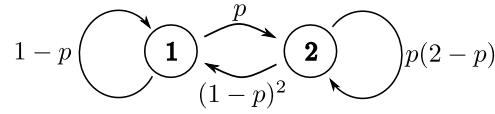


FIGURE 6.3 – Le graphe associé à la CMTD qui modélise le fonctionnement de l’atelier, dans l’exercice 12. Premier cas : une machine au plus est réparée pendant la nuit

machine opérationnelle dans l’atelier. On a donc dans ce scénario trois états possibles. Soit $X_n \in E := \{0, 1, 2\}$, le nombre de machine(s) qui fonctionne(nt) le matin du jour n . Le graphe associé à la CMTD qui modélise le fonctionnement de l’atelier est représenté à la figure 6.4 (avec $p \in [0, 1]$). Par rapport à la situation précédente, il y a un nouvel état et la plupart des probabilités de transition changent. La transition de 0 vers 1 est certaine (une machine sera nécessairement réparée le jour n et opérationnelle le jour $n + 1$). La transition de 1 vers 0 est impossible (la machine en panne est nécessairement réparée pendant le jour n). La transition de l’état 2 vers l’état 0 signifie que deux machines tombent en panne le jour n . Au contraire, la transition de 2 vers 2 signifie qu’aucune des deux machines qui fonctionnent le jour n ne tombe en panne. La transition de 1 vers 2 signifie que la machine qui fonctionne le jour n ne tombe pas en panne, pendant que l’autre est réparée. Au contraire, la transition de 1 vers 1 signifie que la machine qui fonctionne le jour n tombe en panne au cours de cette journée. Enfin, la probabilité de transition de 2 vers 1 est calculée comme la probabilité complémentaire des probabilités de transition de 2 vers 0 ou 1.

La matrice de passage associée à cette chaîne de Markov est

$$\mathbf{P} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 1-p & p \\ (1-p)^2 & 2p(1-p) & p^2 \end{pmatrix}$$

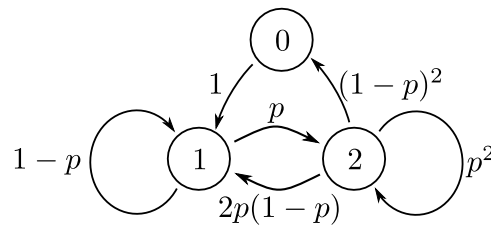


FIGURE 6.4 – Le graphe associé à la CMTD qui modélise le fonctionnement de l’atelier, dans l’exercice 12.

3. Il s'agit ici de calculer la distribution stationnaire de probabilités, en particulier la probabilité d'être d'en l'état 0 en régime stationnaire (qui est aussi la proportion du temps que la chaîne passe dans l'état 0, sur le long terme). La distribution stationnaire de probabilité satisfait les équations :

$$\begin{pmatrix} \pi_0 & \pi_1 & \pi_2 \end{pmatrix} = \begin{pmatrix} \pi_0 & \pi_1 & \pi_2 \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 \\ 0 & 1-p & p \\ (1-p)^2 & 2p(1-p) & p^2 \end{pmatrix}$$

et

$$\pi_0 + \pi_1 + \pi_2 = 1$$

En résolvant ces équations, on obtient :

$$\begin{aligned} \pi_0 &= \frac{(1-p)^2 p}{p^3 - 3p^2 + 2p + 1} \\ \pi_1 &= \frac{1-p^2}{p^3 - 3p^2 + 2p + 1} \\ \pi_2 &= \frac{p}{p^3 - 3p^2 + 2p + 1} \end{aligned}$$

La probabilité d'avoir les 2 machines en panne en arrivant le matin est π_0 .

Exercice 13

Dans une maison constituée de 4 pièces se déroule un cruel épisode de la vie domestique. Une souris se promène pour trouver à manger. Un chat ⁷ décide d'attraper la souris.

La souris démarre sa recherche au hasard dans la maison (mais pas par la cuisine, où la nourriture se trouve), elle reste 10 minutes dans chaque pièce, puis ayant oublié où elle a déjà cherché, change d'endroit pour une des 3 autres pièces. Le chat, en bonne larve qu'il est, ne veut pas courir. Il se place dans la cuisine et attend la souris.

1. Pourquoi peut-on modéliser cette situation par une chaîne de Markov ?
2. Donner le graphe et la matrice de transition de cette chaîne.
3. Classifier les états de la chaîne.
4. Quelle est l'espérance de vie de la souris ?
5. Déterminer le régime stationnaire de la chaîne (i.e. donner la distribution stationnaire). Etait-il possible de donner à priori, sans calcul, cette distribution stationnaire de probabilités ?

7. qui lui n'a pas ce genre de soucis, c'est une larve gavée par ses maîtres

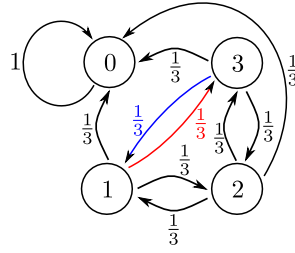


FIGURE 6.5 – Le graphe associé à la CMTD qui modélise la parcours aléatoire de la souris dans la maison (exercice 13).

Correction de l'Exercice 13

1. Oui. Comme la souris n'a pas de mémoire, la probabilité que l'état au temps $n + 1$ soit une certaine valeur $x_{n+1} \in \{0, 1, 2, 3\}$ (c'est-à-dire que la souris décide d'aller dans la pièce x_{n+1}) ne dépend que de l'état au temps n (la pièce x_n où elle se trouve) et pas des pièces qu'elle a visitées avant (les pièces x_k pour $k < n$). Cette absence de mémoire entraîne donc bien le caractère markovien de la chaîne.
2. Soit $X_n \in E := \{0, 1, 2, 3\}$, la pièce où se trouve la souris au temps n . On supposera dans la suite que la cuisine est la pièce 0. Le graphe associé à la CMTD qui modélise ce problème est donnée à la figure 6.5. La matrice de passage associée à cette chaîne de Markov est

$$\mathbf{P} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ \frac{1}{3} & 0 & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & 0 & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 \end{pmatrix}$$

3. L'état 1 est récurrent et absorbant. Une fois que la souris entre dans la cuisine, elle y meurt ! Les états 1, 2 et 3 communiquent entre eux et forment donc une seule classe d'équivalence. Ils sont tous transitoires.
4. On demande l'espérance du temps d'entrée la classe $F := \{0\}$, ou encore l'espérance du temps de séjour dans la classe $\bar{F} := \{1, 2, 3\}$. On sait (voir cours) que cette espérance vaut :

$$E(T_F) = \sum_{i \in \bar{F}} \pi_i^{(0)} \cdot E(T_F | X_0 = i) = \boldsymbol{\pi}^{(0)} \mathbf{T}$$

où

$$\mathbf{T}_i := E(T_F | X_0 = i)$$

désigne l'espérance du temps d'entrée la classe $F := \{0\}$ partant de l'état $X_0 = i \in \bar{F}$, et où

$$\pi_i^{(0)} := P(X_0 = i), \forall i \in \{1, 2, 3\}$$

est la distribution initiale de probabilité. On sait par ailleurs que le vecteur \mathbf{T} satisfait l'équation

$$\begin{aligned}\mathbf{T}_1 &= 1 + \frac{1}{3}(\mathbf{T}_2 + \mathbf{T}_3) \\ \mathbf{T}_2 &= 1 + \frac{1}{3}(\mathbf{T}_1 + \mathbf{T}_3) \\ \mathbf{T}_3 &= 1 + \frac{1}{3}(\mathbf{T}_1 + \mathbf{T}_2)\end{aligned}$$

dont la solution est $\mathbf{T}_1 = \mathbf{T}_2 = \mathbf{T}_3 = 3$. On trouve donc pour l'espérance de vie de la souris (mesurée en nombre de pas de temps)

$$E(T_F) = \sum_{i=1}^3 \pi_i^{(0)} \mathbf{T}_i = 3 \sum_{i=1}^3 \pi_i^{(0)} = 3$$

c'est-à-dire 30 minutes.

5. La distribution stationnaire de cette CMTD satisfait les équations

$$\begin{pmatrix} \pi_0 & \pi_1 & \pi_2 & \pi_3 \end{pmatrix} = \begin{pmatrix} \pi_0 & \pi_1 & \pi_2 & \pi_3 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ \frac{1}{3} & 0 & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & 0 & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 \end{pmatrix}$$

avec la condition de normalisation

$$\pi_0 + \pi_1 + \pi_2 + \pi_3 = 1$$

On obtient par calcul $\pi_0 = 1$ et $\pi_1 = \pi_2 = \pi_3 = 0$. Cette solution était évidente à priori car les états 1, 2 et 3 sont transitoires. Leur probabilités limites valent donc nécessairement

$$\pi_1^{(\infty)} = \pi_2^{(\infty)} = \pi_3^{(\infty)} = 0$$

et sont égales aux probabilités stationnaires correspondantes.

Exercice 14

Une entreprise utilise des machines dont l'usure est caractérisée par quatre valeurs, notées $\{1, 2, 3, 4\}$, qui correspondent respectivement aux états des machines *neuves*, *usées*, *très usées* et *inutilisables*. Chaque jour, une machine dans un de ces états rapporte à l'entreprise respectivement 1000 €, 1000 €, 800 € et 0 €. On modélise le processus de vieillissement de la machine par une chaîne de Markov à temps discret (l'unité est le jour) de matrice de transition :

$$\mathbf{P} = \begin{pmatrix} 0.6 & 0.2 & 0.1 & 0.1 \\ 0 & 0.7 & 0.2 & 0.1 \\ 0 & 0 & 0.8 & 0.2 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

1. Calculer la durée de vie moyenne d'une machine neuve. Même question avec une machine très usée.
2. Le remplacement de la machine coûte $K \text{ €}$ et demande une journée (pendant laquelle la machine ne produit rien). Choisir parmi ces deux politiques laquelle est la plus profitable à long terme pour l'entreprise :
 P1 : remplacer la machine dès qu'elle est *inutilisable*
 P2 : remplacer la machine dès qu'elle est *très usée* ou *inutilisable*

Correction de l'Exercice 14

1. Le graphe associée à la chaîne de Markov qui représente l'usure des machines est donné à la figure 6.6. La durée de vie moyenne d'une

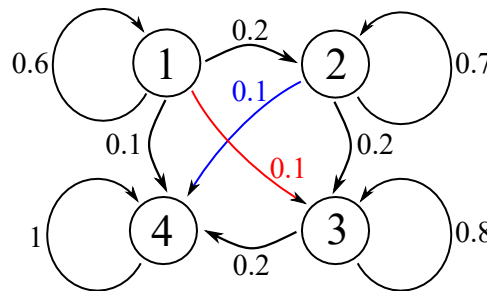


FIGURE 6.6 – Le graphe associé à la CMTD qui modélise l'usure des machines dans l'exercice 14. Les états $\{1, 2, 3, 4\}$ correspondent respectivement aux états des machines *neuves*, *usées*, *très usées* et *inutilisables*.

machine *neuve* est l'espérance du temps d'entrée la classe $F := \{4\}$ (ou encore l'espérance du temps de séjour dans la classe $\bar{F} := \{1, 2, 3\}$) partant de l'état 1, tandis que la durée de vie moyenne d'une machine *très usée* est la même espérance, mais partant de l'état 3. En notant T_F le temps d'entrée dans la classe $F := \{4\}$ (i.e. la durée de vie) et

$$\mathbf{T}_i := E(T_F | X_0 = i)$$

l'espérance du temps d'entrée la classe $F := \{4\}$ partant de l'état $X_0 = i \in \bar{F}$, on a (voir cours)

$$\begin{aligned} \mathbf{T}_1 &= 1 + 0.6 \cdot \mathbf{T}_1 + 0.2 \cdot \mathbf{T}_2 + 0.1 \cdot \mathbf{T}_3 \\ \mathbf{T}_2 &= 1 + 0.7 \cdot \mathbf{T}_2 + 0.2 \cdot \mathbf{T}_3 \\ \mathbf{T}_3 &= 1 + 0.8 \cdot \mathbf{T}_3 \end{aligned}$$

dont les solutions sont

$$\begin{aligned} \mathbf{T}_1 &= \frac{85}{12} \simeq 7.0833.. \\ \mathbf{T}_2 &= \frac{20}{3} \simeq 6.6666.. \\ \mathbf{T}_3 &= 5 \end{aligned}$$

exprimées en unité de temps (ici, le jour).

- En appliquant le politique P1 (remplacer immédiatement les machines inutilisables et seulement celles-là) on modifie la CMTD comme illustré à la figure 6.7. Cette modification change radicalement le comportement de la chaîne qui devient irréductible. Il n'y a plus qu'une seule classe d'équivalence et tous les états deviennent récurrents non nuls (alors qu'auparavant l'état 4 était absorbant et les autres états étaient transitoires).

La matrice de transition de cette nouvelle CMTD s'écrit :

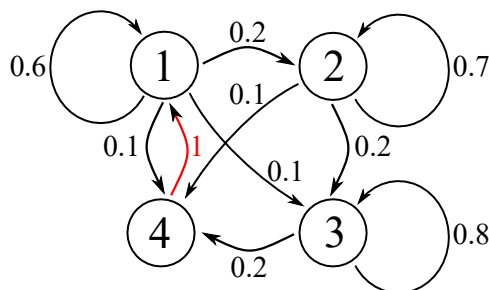


FIGURE 6.7 – Le graphe associé à la CMTD qui modélise l'usure des machines dans l'exercice 14 lorsqu'on ne remplace que les machines inutilisables. La modification correspond à l'arrêt *rouge*

$$\mathbf{P}_1 = \begin{pmatrix} 0.6 & 0.2 & 0.1 & 0.1 \\ 0 & 0.7 & 0.2 & 0.1 \\ 0 & 0 & 0.8 & 0.2 \\ 1 & 0 & 0 & 0 \end{pmatrix}$$

On s'intéresse au comportement de la chaîne sur le long terme, c'est-à-dire à la distribution stationnaire de probabilités, qui dans ce cas est aussi la distribution limite. Celle-ci satisfait les équations :

$$\begin{pmatrix} \pi_1 & \pi_2 & \pi_3 & \pi_4 \end{pmatrix} = \begin{pmatrix} \pi_1 & \pi_2 & \pi_3 & \pi_4 \end{pmatrix} \begin{pmatrix} 0.6 & 0.2 & 0.1 & 0.1 \\ 0 & 0.7 & 0.2 & 0.1 \\ 0 & 0 & 0.8 & 0.2 \\ 1 & 0 & 0 & 0 \end{pmatrix}$$

avec la condition de normalisation

$$\pi_1 + \pi_2 + \pi_3 + \pi_4 = 1$$

On obtient par calcul

$$\begin{pmatrix} \pi_1 & \pi_2 & \pi_3 & \pi_4 \end{pmatrix} = \begin{pmatrix} \frac{30}{97} & \frac{20}{97} & \frac{35}{97} & \frac{12}{97} \end{pmatrix}$$

L'espérance de gain quotidien par machine associé à la politique P1 est donc

$$E_1 := \pi_1 \cdot 1000 + \pi_2 \cdot 1000 + \pi_3 \cdot 800 - \pi_4 \cdot K = \frac{78000 - 12K}{97}$$

euros par jour et par machine, où K représente le coût de remplacement d'une machine.

En appliquant la politique P2 (remplacer immédiatement les machines dès qu'elles sont inutilisables ou très usées) on modifie la CMTD comme illustré à la figure 6.8. Il n'y a plus qu'une seule classe d'équivalence et tous les états deviennent récurrents non nuls.

La matrice de transition de cette nouvelle CMTD s'écrit :

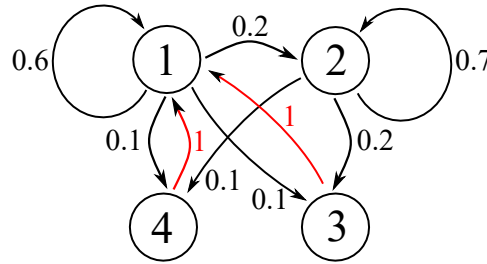


FIGURE 6.8 – Le graphe associé à la CMTD qui modélise l'usure des machines dans l'exercice 14 lorsqu'on ne remplace que les machines inutilisables. Les modifications correspondent aux arrêtes rouges

$$\mathbf{P}_2 = \begin{pmatrix} 0.6 & 0.2 & 0.1 & 0.1 \\ 0 & 0.7 & 0.2 & 0.1 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}$$

La distribution stationnaire de probabilités satisfait les équations :

$$\begin{pmatrix} \pi_1 & \pi_2 & \pi_3 & \pi_4 \end{pmatrix} = \begin{pmatrix} \pi_1 & \pi_2 & \pi_3 & \pi_4 \end{pmatrix} \begin{pmatrix} 0.6 & 0.2 & 0.1 & 0.1 \\ 0 & 0.7 & 0.2 & 0.1 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}$$

avec la condition de normalisation

$$\pi_1 + \pi_2 + \pi_3 + \pi_4 = 1$$

On obtient par calcul

$$\begin{pmatrix} \pi_1 & \pi_2 & \pi_3 & \pi_4 \end{pmatrix} = \begin{pmatrix} \frac{30}{62} & \frac{20}{62} & \frac{7}{62} & \frac{5}{62} \end{pmatrix}$$

L'espérance de gain quotidien par machine est alors

$$E_2 := \pi_1 \cdot 1000 + \pi_2 \cdot 1000 - (\pi_3 + \pi_4) \cdot K = \frac{50000 - 12K}{62}$$

euros par jour et par machine, où K représente le coût de remplacement d'une machine. Dans ce cas particulier où une machine très usée continue à rapporter 800€ par jour, la politique P2 est avantageuse si $E_2 \geq E_1$, c'est-à-dire si le coût de remplacement d'une machine satisfait

$$K \leq \frac{100}{3} \simeq 33.33.. \text{€}$$

Si la productivité des machines très usées diminue, ce seuil va s'élever.

Exercice 15

Une information (de type binaire 0 ou 1) est transmise d'un émetteur R_0 , jusqu'à un récepteur R_n à travers une suite de n relais intermédiaires désignés par R_1, R_2, \dots, R_{n-1} ($n \geq 2$). Chaque relais R_k transmet l'information reçue de R_{k-1} au suivant R_{k+1} avec la probabilité p ($0 < p < 1$). Il transmet donc l'information contraire à celle reçue (erreur) avec la probabilité $1 - p$. Ces probabilités sont indépendantes des informations transmises par les autres relais, en amont ou en aval de la chaîne.

1. Modéliser ce problème par une chaîne de Markov.
2. Calculer p_n la probabilité pour l'information initiale soit correctement transmise de R_0 à R_n .
3. Calculer la limite de p_n , pour $n \rightarrow \infty$

Correction de l'Exercice 15

1. Une des solutions les plus simples consiste à représenter cette chaîne de transmission comme une CMTD avec un espace d'état binaire $E := \{0, 1\}$. On notera alors X_n ($n \geq 1$) la valeur de l'information transmise par le relais R_{n-1} au relais R_n (X_0 représentera l'information initiale à transmettre). Les probabilités d'état "au temps" n sont

$$\pi_i^{(n)} := P(X_n = i), i \in \{0, 1\}$$

La position n dans la chaîne est donc représentée par le “temps” ou encore le nombre de transitions. Le graphe associé à la chaîne de Markov qui modélise cette transmission d’information est donné à la figure 6.9. La matrice des probabilités de transition associée est

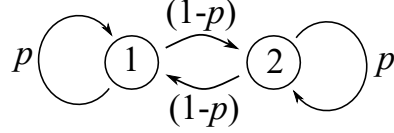


FIGURE 6.9 – Le graphe associé à la CMTD qui modélise la transmission d’une information binaire à travers n relais, avec une probabilité d’erreur $(1 - p)$ à chaque relais (exercice 15).

$$\mathbf{P} = \begin{pmatrix} p & 1-p \\ 1-p & p \end{pmatrix}$$

Dans la suite, on s’intéressera au cas non trivial où $0 < p < 1$. Les résultats s’obtiennent très facilement pour les deux chaînes (déterministes) $p = 0$ et $p = 1$.

2. On s’intéresse aux probabilités d’état après n transitions, données par la formule de Chapman-Kolmogorov :

$$\boldsymbol{\pi}^{(n)} = \begin{pmatrix} \pi_0^{(0)} & \pi_1^{(0)} \end{pmatrix} \mathbf{P}^n$$

La matrice \mathbf{P} se diagonalise sous la forme :

$$\mathbf{P} = \mathbf{U}^T \mathbf{D} \mathbf{U} = \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 2p-1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$$

avec

$$\mathbf{U} := \frac{\sqrt{2}}{2} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \text{ et } \mathbf{D} := \begin{pmatrix} 1 & 0 \\ 0 & 2p-1 \end{pmatrix}$$

où est unitaire, c’est-à-dire

$$\mathbf{U}^T \mathbf{U} = \mathbf{U} \mathbf{U}^T = \mathbf{I}_2$$

On obtient donc

$$\mathbf{P}^n = \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & (2p-1)^n \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$$

La probabilité demandée s’écrit, dans le cas où $X_0 = 1$:

$$\begin{aligned} p_n &= P(X_n = 1 | X_0 = 1) \\ &= (\mathbf{P}^n)_{2,2} \\ &= \frac{1 + (2p-1)^n}{2} \end{aligned} \tag{6.6}$$

On trouve la même valeur dans le cas où $X_0 = 0$:

$$\begin{aligned} p_n &= P(X_n = 0 | X_0 = 0) \\ &= (\mathbf{P}^n)_{1,1} \\ &= \frac{1 + (2p - 1)^n}{2} \end{aligned} \quad (6.7)$$

Donc, dans tous les cas :

$$p_n = \frac{1 + (2p - 1)^n}{2}$$

3. Pour $0 < p < 1$, on a $-1 < (2p - 1) < +1$ et donc $(2p - 1)^n \rightarrow 0$. Ainsi

$$\lim_{n \rightarrow \infty} p_n = \frac{1}{2}$$

C'est assez logique. La probabilité p_n diminue lorsque le nombre de relais n augmente. Elle tend vers $\frac{1}{2}$ qui est la valeur obtenue dans le cas où la valeur de sortie est choisie totalement au hasard (avec une distribution uniforme de probabilité).

Exercice 16

On considère une suite aléatoire de bits indépendants b_n . Chacun de ces bits est distribué selon la même loi :

$$\begin{aligned} P(b_i = 1) &= p \\ P(b_i = 0) &= 1 - p \end{aligned}$$

pour tout $i \geq 1$. Soit X_n = longueur du dernier bloc de 1 touchant le n^e bit (cela implique notamment que le n^e bit lui-même est un 1). Par exemple, pour la suite

$$0101110 \underbrace{11111}_{\text{bloc de 1}} 100110...$$

on a $X_{12} = 5$, $X_{13} = 6$ et $X_{14} = 0$, alors que pour la suite

$$01011100000 \underbrace{0}_{\text{bloc de 0}} 110010...$$

on a $X_{12} = 0$, $X_{13} = 1$ et $X_{14} = 2$

1. Montrer que X_n est une chaîne de Markov dont on calculera les probabilités de transition.
2. Montrer que cette chaîne est irréductible et apériodique
3. Calculer la distribution limite de probabilités.

Correction de l'Exercice 16

1. On pose, par convention $X_0 = 0$. On remarque que, si $b_{n+1} = 1$, alors $X_{n+1} = X_n + 1$ et que, si $b_{n+1} = 0$, alors $X_{n+1} = 0$. La valeur de X_{n+1} ne dépend que de la valeur X_n (état précédent) et de la valeur $\mathbb{P}[b_{n+1} = 1] = p$, indépendante de n . La chaîne $\{X_n\}_{n \geq 0}$ est donc une chaîne de Markov homogène définie sur l'espace d'état $E = \mathbb{N}$. Les probabilités de transition sont définies par :

$$p_{i,j} = \begin{cases} p & \text{si } j = i + 1 \\ (1 - p) & \text{si } j = 0 \\ 0 & \text{sinon} \end{cases}$$

Le graphe associé à cette CMTD est représenté à la figure

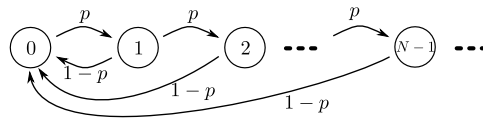


FIGURE 6.10 – Le graphe de la CMTD considérée à l'exercice 16. L'état X_n représente la longueur du bloc de 1 qui touchent le n^{e} bit

2. Quelques soient les états $i, j \in \mathbb{N}$, $i < j$, il existe un chemin direct de i à j (de longueur $j - i$). Par ailleurs, pour aller de j à i , il suffit d'aller de j à 0, puis de 0 à i (le chemin le plus court est donc dans ce cas de longueur $i + 1$). Tous les états communiquent donc et la chaîne est irréductible. Par ailleurs, il existe un rebouclage direct de l'état 0 vers lui-même. La chaîne est donc nécessairement apériodique.
3. La matrice de transition associée à la CMTD s'écrit :

$$\mathbf{P} = \begin{pmatrix} 1-p & p & 0 & 0 & 0 & \cdots & 0 \\ 1-p & 0 & p & 0 & 0 & \cdots & 0 \\ 1-p & 0 & 0 & p & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \end{pmatrix}$$

La distribution stationnaire est aussi la distribution limite et satisfait l'équation

$$\pi = \pi \mathbf{P}$$

qui s'écrit en composantes

$$\pi_0 = (1 - p) \sum_{k=0}^{\infty} \pi_k = (1 - p)$$

et

$$\pi_{n+1} = p\pi_n, \quad \forall n \geq 0$$

c'est-à-dire

$$\pi_n = p^n(1-p), \quad \forall n \geq 0$$

Exercice 17 (Les urnes d'Ehrenfest)

On considère deux urnes A et B contenant initialement b boules chacune. A chaque étape, on choisit une boule parmi les $2b$ et on la change d'urne. On note X_n le nombre de boules dans l'urne A à l'étape n .

1. Montrer que $(X_n)_{n \in \mathbb{N}}$ une chaîne de Markov, et déterminer les probabilités de transition de cette chaîne.
2. On prend $b = 5$. Calculer la distribution stationnaire de la chaîne. Est-elle la loi limite ?
3. Si $b = 5$, déterminer le nombre moyen de transitions avant que l'une des deux urnes ne se vide.

Correction de l'Exercice 17

1. Soit $X_n \in E := \{0, 1, \dots, 2b\}$ le nombre de boules contenues dans l'urne A à l'étape n (après le n^{e} tirage). Les états de E et les probabilités de transition sont représentées à la figure 6.11. Les probabilités de

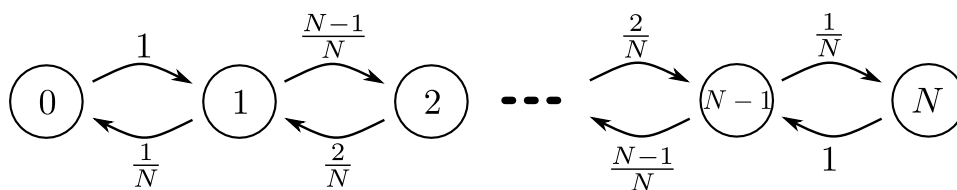


FIGURE 6.11 – Le graphe associé à la CMTD qui modélise les tirages aléatoires successifs dans le problème des urnes d'Ehrenfest avec $N := 2b$ boules (exercice 17).

transitions au n^{e} tirage ne dépendent que du nombre de boules dans chacune des urnes au moment du tirage, c'est-à-dire de l'état de la chaîne au temps $n - 1$. Il s'agit donc bien d'un processus markovien.

La matrice de transition associée est :

$$\mathbf{P} = \begin{pmatrix} 0 & 1 & 0 & \cdots & \cdots & \cdots & 0 \\ \frac{1}{2b} & 0 & \frac{2b-1}{2b} & \ddots & & & \vdots \\ 0 & \frac{2}{2b} & 0 & \ddots & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 & \frac{2}{2b} & 0 \\ \vdots & & & \ddots & \frac{2b-1}{2b} & 0 & \frac{1}{2b} \\ 0 & \cdots & \cdots & \cdots & 0 & 1 & 0 \end{pmatrix}$$

2. La CMTD de la figure 6.11 est finie et irréductible (une seule classe d'équivalence). Tous les états sont récurrents non nuls. Cependant, elle est aussi périodique, de période 2. La distribution limite n'existe donc pas. La distribution stationnaire π est solution des équations :

$$\begin{cases} \pi_0 = \frac{1}{2b} \pi_1 \\ \vdots \\ \pi_k = \frac{2b-(k-1)}{2b} \pi_{k-1} + \frac{k+1}{2b} \pi_{k+1}, \text{ pour } k \in \{1, \dots, 2b-1\} \\ \vdots \\ \pi_{2b} = \frac{1}{2b} \pi_{2b-1} \end{cases} \quad (6.8)$$

avec la condition de normalisation

$$\sum_{i=0}^{2b} \pi_i = 1 \quad (6.9)$$

En résolvant les équations (6.8), on trouve

$$\pi_k = \frac{k+1}{2b-k} \pi_{k+1}, \forall k \in \{0, \dots, 2b-1\}$$

En substituant ces valeurs dans l'équation de normalisation (6.9), on obtient :

$$\begin{aligned} \sum_{k=0}^{2b} \pi_k &= \left[\sum_{k=0}^{2b} \frac{(2b)!}{k!(2b-k)!} \right] \pi_{2b} \\ &= 2^{2b} \pi_{2b} \\ &= 1 \end{aligned}$$

On obtient donc

$$\pi_{2b} = \frac{1}{2^{2b}}$$

et

$$\pi_{2b-k} = \frac{(2b)!}{k!(2b-k)!} \frac{1}{2^{2b}}, \forall k \in \{1, \dots, 2b\} \quad (6.10)$$

En particulier, dans le cas $b = 5$, on trouve :

$$\pi = \frac{1}{1024} \begin{pmatrix} 1 & 10 & 45 & 120 & 210 & 252 & 210 & 120 & 45 & 10 & 1 \end{pmatrix}$$

D'une manière générale, la distribution stationnaire (6.10) est une loi binomiale $B(2b, \frac{1}{2})$. Dans ce modèle, en régime stationnaire, les configurations proches de l'équilibre, où le nombre de boules dans chacune des deux urnes est sensiblement le même, sont donc plus probables que les configurations déséquilibrées.

3. On cherche à déterminer \mathbf{T}_i , les temps moyens de séjour dans la classe $\bar{F} := \{1, \dots, 2b-1\}$, sachant que $X_0 = i$ (pour $i \in \bar{F}$). On sait que (voir cours) :

$$\mathbf{T} = \mathbf{1} + \mathbf{P}|_{\bar{F}} \mathbf{T}$$

c'est-à-dire

$$\mathbf{T} = (\mathbf{I}_{2b-1} - \mathbf{P}|_{\bar{F}})^{-1} \mathbf{1}$$

avec

$$\mathbf{P}|_{\bar{F}} = \begin{pmatrix} 0 & \frac{2b-1}{2b} & & & \\ \frac{2}{2b} & 0 & \ddots & & \\ & \ddots & \ddots & \ddots & \\ & & \ddots & 0 & \frac{2}{2b} \\ & & & \frac{2b-1}{2b} & 0 \end{pmatrix}$$

et où \mathbf{I}_{2b-1} désigne la matrice identité de taille $(2b-1) \times (2b-1)$. Dans le cas particulier où $b = 5$, on trouve

$$\mathbf{T}_b = \mathbf{T}_5 = 584.33..$$

Exercice 18 (Pagerank)

On examine dans cet exercice le fonctionnement de l'algorithme *Pagerank* de Google. On note r_i le nombre de liens existant sur la page i .

On modélise l'attitude d'un surfer de la façon suivante : une fois visitée la page i , soit il choisit une adresse Web au hasard (avec la probabilité p), soit il décide de visiter une des pages référencées sur la page i (avec la probabilité $1 - p$). Dans le cas où il choisit une adresse au hasard, on considère que chaque page à la même probabilité $\frac{1}{N}$ d'être choisie. Dans le cas où le surfer choisit un des liens de la page i , il le fait également avec une distribution uniforme. Dans ce cas, chaque page a donc une probabilité $\frac{1}{r_i}$.

On considère dans cet exercice un Web miniature⁸ constitué de 5 pages et représenté à la figure 6.12.

8. Dans le cas du Web mondial, les principes de modélisation et de définition de la

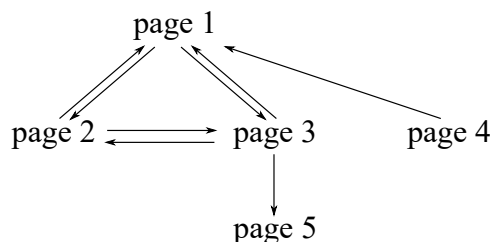


FIGURE 6.12 – Le Web simplifié considéré à l'exercice 18

1. Modéliser la promenade d'un surfer sur le Web de la figure 6.12 par une chaîne de Markov dont on donnera le graphe et la matrice de transition.
2. Pourquoi existe-t-il nécessairement une loi limite $\pi^{(\infty)}$? Comment peut-on la calculer ?
3. Soit $\pi = (\pi_1, \pi_2, \dots, \pi_N)$ la distribution stationnaire de probabilité. La mesure *Pagerank* de la page i est alors simplement définie par

$$\text{Pagerank}(i) := \pi_i$$

Calculer le *Pagerank* pour toutes les pages du Web de la figure 6.12, successivement pour $p = 0.15$, $p = 10^{-2}$ et $p = 0.99$.

4. Imaginer une méthode de calcul itérative de la mesure *Pagerank* qui soit adaptée à un Web de très grande taille mais de structure creuse (la matrice d'incidence du graphe associé est creuse ou, de manière équivalente, une page ne pointe que sur quelques autres pages)

Correction de l'Exercice 18

- Soit $X_n \in E := \{1, 2, 3, 4, 5\}$, la n^{e} page où se trouve le surfer au cours de sa visite sur le Web de la figure 6.12. Comme il choisira la page suivante avec une probabilité que ne dépend que de la page où il se trouve et de la page suivante, on peut modéliser cette visite par une chaîne de Markov dont le graphe est représenté à la figure 6.13. La matrice de transition associée à cette CMTD s'écrit :

mesure *pagerank* restent les mêmes. Les méthodes de modélisation, d'identification et de calcul, par contre, sont adaptées à des CMTD de très grande taille. On pourra pour une introduction rapide au sujet se reporter par exemple à la page Wikipedia en anglais sur le Pagerank.

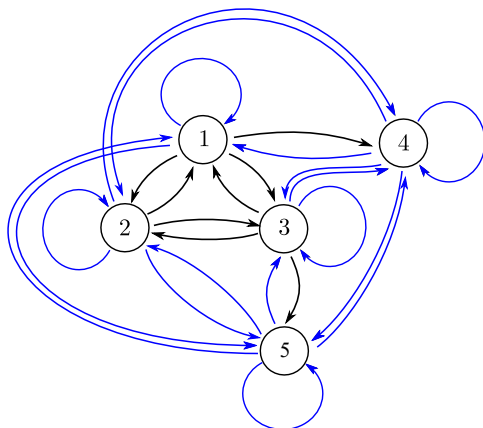


FIGURE 6.13 – La chaîne de Markov associée au Web de la figure 6.12. Les arcs en bleu ont été ajoutés pour représenter les transitions possibles en dehors des liens hypertexte présents dans le Web initial. L'ensemble des probabilités de transition (y compris celles qui correspondent aux arcs en noir) sont modifiées par la possibilité d'aller vers n'importe quel lien du Web, au hasard, en dehors des liens hypertexte.

$$\mathbf{P} = (1 - p) \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \\ \frac{1}{3} & \frac{1}{3} & 0 & 0 & \frac{1}{3} \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} + \frac{p}{5} \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix}$$

- La CMTD de la figure 6.13 est nécessairement apériodique et irréductible, grâce à l'ajout de la possibilité de transition vers n'importe quelle page du Web. Dès lors, comme il s'agit d'une chaîne finie, tous les états sont récurrents non nuls. Dans ce cas, la distribution limite de probabilités $\pi^{(\infty)}$ existe toujours et est nécessairement égale à la distribution stationnaire π solution de

$$\pi = \pi \mathbf{P}$$

avec la condition de normalisation

$$\sum_{i=1}^N \pi_i = 1$$

Il est intéressant de noter que sans la possibilité de transition aléatoire vers une page quelconque du Web, l'état 5 serait absorbant. Dans ce cas la distribution limite serait

$$\pi_i^{(\infty)} = \begin{cases} 0 & \text{si } i \neq 5 \\ 1 & \text{si } i = 5 \end{cases}$$

Il suffirait alors de créer une page sans lien vers d'autres pages pour augmenter arbitrairement son *Pagerank*. Avec une probabilité p de s'échapper d'un tel piège, on retrouvera pour chaque état une distribution limite proche du nombre de liens qui pointent vers la page correspondante.

- En résolvant le système ci-dessus, pour $p = 0.15$ ⁹, on obtient :

$$\pi = (0.1623.. \quad 0.1444.. \quad 0.1604.. \quad 0.0300.. \quad 0.5029..)$$

Cette valeur de p ne compense qu'imparfaitement la sur-représentation de la page 5. Cet effet est du à la taille très faible du Web considéré. Une augmentation de la valeur de p augmentera la probabilité de sauter vers les autres pages sans suivre les liens hypertexte. A la limite, on se rapprochera d'une distribution uniforme. Par exemple, pour $p = 0.99$, on obtient :

$$\pi = (0.2016.... \quad 0.1997.. \quad 0.2000.. \quad 0.1980.. \quad 0.2007..)$$

A contrario, pour p proche de zéro, nous ne compensons plus l'effet "cul de sac" de la page 5. Par exemple, pour $p = 0.01$, on obtient :

$$\pi = (0.0211.... \quad 0.0198.. \quad 0.0223.. \quad 0.0020.. \quad 0.9348..)$$

Si nous modifions le Web de la figure 6.12, en ajoutant un lien de la page 5 à la page 4, on obtient par contre (toujours avec la valeur $p = 0.01$) :

$$\pi = (0.3028.... \quad 0.2413.. \quad 0.2711.. \quad 0.0931.. \quad 0.0917..)$$

qui correspond à l'intuition que l'on peut avoir de l'importance relative des pages de ce "nouveau" Web (en gros, le nombre de liens qui pointent vers chaque page)

- L'application de la méthode précédente au calcul du *Pagerank* pour un Web qui comprend plusieurs milliards de pages est bien sûr exclue. On pourra utiliser pour de très grands réseaux la récurrence

$$\pi^{(n+1)} = \mathbf{P}\pi^{(n)}$$

qui converge vers la distribution stationnaire de probabilités¹⁰. En effet, d'une part la matrice d'incidence des très grands réseaux est en général creuse et, d'autre part, les renouvellements de ces réseaux sont généralement progressifs et ne concernent que peu de noeuds

9. C'est la valeur généralement utilisée en pratique, qui correspond à un changement aléatoire toutes les six pages environs

10. C'est le point fixe de cette récurrence dont on peut montrer qu'elle est toujours contractante

(et d'arrêtes). Or la récurrence ci-dessus se calcule aisément pour des matrices \mathbf{P} creuses (on n'effectue pas les multiplications avec les éléments nuls dans les rangées de \mathbf{P}). De plus, si on démarre cette récurrence avec une distribution initiale proche de la distribution stationnaire avant les derniers changements (e.g. l'ajout de quelques pages), il faudra peu d'itérations pour converger vers le nouveau point fixe qui est proche de l'ancien.

Exercice 19

Soit $n \in \mathbb{N}^*$. On pose $E = \{0, 1, \dots, n\}$.

On considère la chaîne de Markov (X_n) sur E de matrice de transition P donnée, pour $0 \leq i \leq n-1$ par

$$P(i, j) = \begin{cases} p & \text{si } j = (i + 1) \\ q & \text{si } j = 0 \\ 0 & \text{sinon} \end{cases}$$

avec $p, q \in]0, 1[$ et $p + q = 1$. L'état n est absorbant.

1. Faire le graphe de (X_n) et classer les états de la chaîne.
2. On note T le temps d'entrée en n . Que vaut $E_x(T)$, pour $x \in E$?
3. Une pièce donne pile avec la probabilité p . On la lance successivement plusieurs fois. Soit $n \in \mathbb{N}^*$ fixé.
 - (a) Montrer que $P(\text{obtenir } n \text{ piles consécutifs en un nombre fini de lancers}) = 1$.
 - (b) Combien de lancers sont nécessaires en moyenne?

Correction de l'Exercice 19

1. Le graphe de la CMTD (X_n) est donné à la figure 6.14. Tous les états sont transitoires (dans la cas $0 < p < 1$), sauf l'état n qui est récurrent (absorbant).

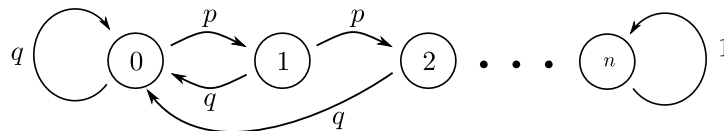


FIGURE 6.14 – Le graphe de la CMTD considérée à l'exercice 19

2. Suivant les notations introduites à la section 3.5, nous noterons $F := \{n\}$ la sous-chaîne cible, T_F le temps d'entrée dans F défini par

$$T_F := \inf \{n \geq 0 \mid X_n \in F\}$$

et

$$\mathbf{T}_i := E(T_F \mid X_0 = i)$$

l'espérance du temps d'entrée dans F en partant de l'état initial $X_0 = i$. On a alors

$$\mathbf{T} = \mathbf{1} + \mathbf{P}|_{\bar{F}} \mathbf{T}$$

avec

$$\mathbf{P} = \begin{pmatrix} q & p & 0 & 0 & 0 & \cdots & 0 & 0 \\ q & 0 & p & 0 & 0 & \cdots & 0 & 0 \\ q & 0 & 0 & p & 0 & \cdots & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots & \vdots \\ q & 0 & 0 & 0 & 0 & \cdots & p & 0 \\ q & 0 & 0 & 0 & 0 & \cdots & 0 & p \\ 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 1 \end{pmatrix}$$

et

$$\mathbf{P}|_{\bar{F}} = \begin{pmatrix} q & p & 0 & 0 & 0 & \cdots & 0 \\ q & 0 & p & 0 & 0 & \cdots & 0 \\ q & 0 & 0 & p & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ q & 0 & 0 & 0 & 0 & \cdots & p \\ q & 0 & 0 & 0 & 0 & \cdots & 0 \end{pmatrix}$$

On obtient ainsi un système d'équations en \mathbf{T}_i (notés $E_i(T)$ dans l'énoncé) :

$$\begin{aligned} \mathbf{T}_0 &= 1 + q\mathbf{T}_0 + p\mathbf{T}_1 \\ \mathbf{T}_1 &= 1 + q\mathbf{T}_0 + p\mathbf{T}_2 \\ &\vdots \\ \mathbf{T}_{n-2} &= 1 + q\mathbf{T}_0 + p\mathbf{T}_{n-1} \\ \mathbf{T}_{n-1} &= 1 + q\mathbf{T}_0 \end{aligned}$$

qui peuvent être écrites :

$$\begin{aligned}
\mathbf{T}_{n-1} &= 1 + q\mathbf{T}_0 \\
\mathbf{T}_{n-2} &= 1 + q\mathbf{T}_0 + p\mathbf{T}_{n-1} \\
&= (1 + p) + q(1 + p)\mathbf{T}_0 \\
\mathbf{T}_{n-3} &= 1 + q\mathbf{T}_0 + p\mathbf{T}_{n-2} \\
&= (1 + p + p^2) + q(1 + p + p^2)\mathbf{T}_0 \\
&\vdots \\
\mathbf{T}_1 &= 1 + q\mathbf{T}_0 + p\mathbf{T}_2 \\
&= (1 + p + p^2 + \dots + p^{n-2}) + q(1 + p + p^2 + \dots + p^{n-2})\mathbf{T}_0 \\
\mathbf{T}_0 &= 1 + q\mathbf{T}_0 + p\mathbf{T}_1 \\
&= (1 + p + p^2 + \dots + p^{n-1}) + q(1 + p + p^2 + \dots + p^{n-1})\mathbf{T}_0 \\
&= \frac{1 - p^n}{1 - p} + (1 - p^n)\mathbf{T}_0
\end{aligned}$$

La dernière équation donne

$$\mathbf{T}_0 = \frac{1 - p^n}{(1 - p)p^n} = \frac{1 - p^n}{qp^n}$$

On obtient alors de manière itérative :

$$\begin{aligned}
\mathbf{T}_1 &= \frac{1 - p^{n-1}}{1 - p} + (1 - p^{n-1})\mathbf{T}_0 = \frac{1 - p^{n-1}}{qp^n} \\
\mathbf{T}_2 &= \frac{1 - p^{n-2}}{1 - p} + (1 - p^{n-2})\mathbf{T}_0 = \frac{1 - p^{n-2}}{qp^n} \\
&\vdots \\
\mathbf{T}_{n-2} &= \frac{1 - p^2}{1 - p} + (1 - p^2)\mathbf{T}_0 = \frac{1 - p^2}{qp^n} \\
\mathbf{T}_{n-1} &= \frac{1 - p}{1 - p} + (1 - p)\mathbf{T}_0 = \frac{1 - p}{qp^n}
\end{aligned}$$

c'est-à-dire

$$\mathbf{T}_k = \frac{1 - p^{n-k}}{qp^n}, \quad \forall k \in \{0, \dots, n-1\}$$

3. (a) L'état n étant le seul état récurrent de la CMTD, il est clair qu'on y arrivera, en moyenne, en un nombre fini d'étapes, puis qu'on y restera indéfiniment. On peut calculer la probabilité d'entrer dans la classe $F := \{n\}$, partant de l'état initial $X_0 = 0$, notée $\mathbf{V}^{(0)}$ dans la section 3.5. Nous noterons \mathbf{V} le vecteur colonne des probabilités $\mathbf{V}^{(i)}$, d'entrer dans la classe $F := \{n\}$, partant de

l'état initial $X_0 = i$. On a alors $\mathbf{V} = \mathbf{P}\mathbf{V}$ et $\mathbf{V}^{(n)} = 1$, c'est-à-dire :

$$\begin{aligned}\mathbf{V}^{(0)} &= q\mathbf{V}^{(0)} + p\mathbf{V}^{(1)} \\ \mathbf{V}^{(1)} &= q\mathbf{V}^{(0)} + p\mathbf{V}^{(2)} \\ &\vdots \\ \mathbf{V}^{(n-2)} &= q\mathbf{V}^{(0)} + p\mathbf{V}^{(n-1)} \\ \mathbf{V}^{(n-1)} &= q\mathbf{V}^{(0)} + p\mathbf{V}^{(n)} \\ \mathbf{V}^{(n)} &= 1\end{aligned}$$

qui donne de manière itérative

$$\begin{aligned}\mathbf{V}^{(n)} &= 1 \\ \mathbf{V}^{(n-1)} &= q\mathbf{V}^{(0)} + p \\ \mathbf{V}^{(n-2)} &= q\mathbf{V}^{(0)}(1 + p) + p^2 \\ &\vdots \\ \mathbf{V}^{(1)} &= q\mathbf{V}^{(0)}(1 + p + \dots + p^{n-2}) + p^{n-1} \\ \mathbf{V}^{(0)} &= q\mathbf{V}^{(0)}(1 + p + \dots + p^{n-1}) + p^n\end{aligned}$$

et en particulier

$$\mathbf{V}^{(0)} = q\mathbf{V}^{(0)} \frac{1 - p^n}{1 - p} + p^n = \mathbf{V}^{(0)}(1 - p^n) + p^n$$

c'est-à-dire

$$\mathbf{V}^{(0)} = 1$$

conformément au résultat attendu. On a de même, $\mathbf{V}^{(j)} = 1$ quel que soit $j \in \{0, \dots, n-1\}$. Dans cet exemple, la sous-chaîne cible, $F := \{n\}$, est un singleton. Il est donc possible, de manière alternative, de calculer la probabilité $f_{0,n}$ d'aller de l'état 0 à l'état n , en un nombre quelconque d'étapes, et de montrer $f_{0,n} = 1$. Pour conclure, il faut alors montrer que la probabilité qu'un nombre infini de lancers soit nécessaire pour atteindre l'état n est nulle.

- (b) Il s'agit de l'espérance du temps d'entrée dans la classe $F := \{n\}$, partant de l'état initial $X_0 = 0$, déjà calculée précédemment :

$$\mathbf{T}_0 = \frac{1 - p^n}{qp^n}$$

Exercice 20

On considère la chaîne de Markov (X_n) définie sur l'espace d'état $E = \{1, 2, \dots, 5\}$, dont la matrice de transition est donnée par

$$\mathbf{P} = \begin{pmatrix} \frac{1}{2} & 0 & \frac{1}{12} & \frac{1}{4} & \frac{1}{6} \\ 0 & \frac{1}{4} & 0 & 0 & \frac{3}{4} \\ \frac{1}{3} & \frac{1}{12} & \frac{1}{2} & 0 & \frac{1}{12} \\ 0 & \frac{1}{3} & \frac{1}{4} & \frac{1}{4} & \frac{1}{6} \\ 0 & \frac{1}{3} & 0 & 0 & \frac{2}{3} \end{pmatrix}$$

1. Classer les états de la chaîne. On notera dans la suite C , la classe formée par les états récurrents.
2. On note $T = \inf\{n \geq 0, X_n \in C\}$ le temps d'entrée dans C . Calculer $E_x(T)$, l'espérance de T partant de l'état x , pour tout $x \in \bar{C} = E \setminus C$.
3. Calculer la loi d'entrée dans C , c'est-à-dire les probabilités $P_x(X_T = y)$ si d'entrer dans la classe C par l'état $y \in C$, partant de $x \in \bar{C}$.

Correction de l'Exercice 20

1. La chaîne de Markov associée à la matrice \mathbf{P} est représentée à la figure 6.15. Elle peut être partitionnée en deux classes d'équivalences : la classe $\{1, 3, 4\}$, composée d'états transitoires, et la classe $\{2, 5\}$, composée d'états récurrents non nuls. Aucun état de la chaîne n'est périodique.
2. Définissons, en adoptant les notations du cours, le vecteur des espérances des temps d'entrée dans la classe C :

$$\mathbf{T} = \begin{pmatrix} \mathbf{T}_1 \\ \mathbf{T}_3 \\ \mathbf{T}_4 \end{pmatrix} := \begin{pmatrix} E(T_F | X_0 = 1) \\ E(T_F | X_0 = 3) \\ E(T_F | X_0 = 4) \end{pmatrix}$$

Le vecteur \mathbf{T} est solution de :

$$\begin{pmatrix} \mathbf{T}_1 \\ \mathbf{T}_3 \\ \mathbf{T}_4 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + \begin{pmatrix} \frac{1}{2} & \frac{1}{12} & \frac{1}{4} \\ \frac{1}{3} & \frac{1}{2} & 0 \\ 0 & \frac{1}{3} & \frac{1}{4} \end{pmatrix} \begin{pmatrix} \mathbf{T}_1 \\ \mathbf{T}_3 \\ \mathbf{T}_4 \end{pmatrix}$$

On trouve

$$\begin{pmatrix} \mathbf{T}_1 \\ \mathbf{T}_3 \\ \mathbf{T}_4 \end{pmatrix} = \begin{pmatrix} \frac{30}{7} \\ \frac{34}{7} \\ \frac{62}{21} \end{pmatrix}$$

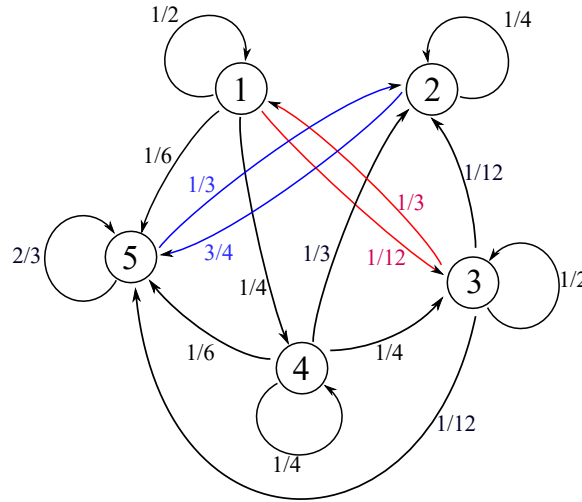


FIGURE 6.15 – Le graphe associé à la CMTD de l'exercice 20.

3. En suivant les notations du cours, nous noterons

$$\mathbf{W}_2^{(i)} = \begin{pmatrix} \mathbf{W}_2^{(1)} \\ \mathbf{W}_2^{(2)} \\ \mathbf{W}_2^{(3)} \\ \mathbf{W}_2^{(4)} \\ \mathbf{W}_2^{(5)} \end{pmatrix} := \begin{pmatrix} P(X_T = 2 | X_0 = 1) \\ P(X_T = 2 | X_0 = 2) \\ P(X_T = 2 | X_0 = 3) \\ P(X_T = 2 | X_0 = 4) \\ P(X_T = 2 | X_0 = 5) \end{pmatrix}$$

le vecteur des probabilités d'entrer dans C par l'état 2, en partant de l'état i pour $i \in \{1, 2, 3, 4, 5\}$. Comme $C = \{2, 5\}$, on a nécessairement $\mathbf{W}_2^{(2)} = 1$ et $\mathbf{W}_2^{(5)} = 0$. Il reste à résoudre :

$$\begin{pmatrix} \mathbf{W}_2^{(1)} \\ \mathbf{W}_2^{(3)} \\ \mathbf{W}_2^{(4)} \end{pmatrix} = \begin{pmatrix} \frac{1}{2} & \frac{1}{12} & \frac{1}{4} \\ \frac{1}{3} & \frac{1}{2} & 0 \\ 0 & \frac{1}{4} & \frac{1}{4} \end{pmatrix} \begin{pmatrix} \mathbf{W}_2^{(1)} \\ \mathbf{W}_2^{(3)} \\ \mathbf{W}_2^{(4)} \end{pmatrix} + \begin{pmatrix} 0 \\ \frac{1}{12} \\ \frac{1}{3} \end{pmatrix} \mathbf{W}_2^{(2)} + \begin{pmatrix} \frac{1}{6} \\ \frac{1}{12} \\ \frac{1}{6} \end{pmatrix} \mathbf{W}_2^{(5)}$$

On trouve

$$\begin{pmatrix} \mathbf{W}_2^{(1)} \\ \mathbf{W}_2^{(3)} \\ \mathbf{W}_2^{(4)} \end{pmatrix} = \begin{pmatrix} \frac{45}{122} \\ \frac{161}{427} \\ \frac{784}{1281} \end{pmatrix} \simeq \begin{pmatrix} 0.368.. \\ 0.307.. \\ 0.612.. \end{pmatrix}$$

Comme les probabilités d'entrer dans la classe C par l'état 5 sont complémentaires, on trouve également :

$$\begin{pmatrix} \mathbf{W}_5^{(1)} \\ \mathbf{W}_5^{(3)} \\ \mathbf{W}_5^{(4)} \end{pmatrix} = \begin{pmatrix} \frac{77}{122} \\ \frac{266}{427} \\ \frac{497}{1281} \end{pmatrix}$$

avec $\mathbf{W}_5^{(2)} = 0$ et $\mathbf{W}_5^{(5)} = 1$.

Exercice 21

On considère un codage bipolaire, qui associe à la suite de symboles a_k la suite b_k de la façon suivante :

$$\begin{aligned} a_k = 0 &\Rightarrow b_k = 0 \\ a_k = 1 &\Rightarrow b_k = 1 \text{ ou } -1 \text{ alternativement} \end{aligned}$$

On suppose que la suite (a_k) est une suite de symboles indépendants, avec $P(a_k = 1) = p$ et $P(a_k = 0) = 1 - p$

1. Montrer que (b_k) n'est pas une chaîne de Markov.
2. Proposer un modèle markovien du système en 4 états. Tracer son graphe
3. Calculer la loi limite de b_k quand $k \rightarrow +\infty$

Correction de l'Exercice 21

1. Il suffit de remarquer que la valeur de b_{k+1} ne dépend pas seulement de la valeur de b_k , mais également des états visités avant l'instant k . On a en effet :

$$b_{k+1} = \begin{cases} 0 & \text{si } a_k = 0 \\ +1 & \text{si } a_k = 1 \text{ et le dernier } b_n \text{ } (n \leq k) \text{ non nul vaut } -1 \\ -1 & \text{si } a_k = 1 \text{ et le dernier } b_n \text{ } (n \leq k) \text{ non nul vaut } +1 \end{cases}$$

En particulier, si $b_k = 0$, il y a une probabilité $1 - q$ de transition vers la sous-chaîne $\{-1, +1\}$ (qui correspond à un nouveau bit $a_{k+1} = 1$). Dans ce cas, le choix de la valeur $b_{k+1} = +1$ ou de la valeur $b_{k+1} = -1$ est alternatif. Il dépend donc de la dernière valeur non nulle de b_k ou $-$ ce qui revient au même $-$ du nombre de valeurs $a_n = 1$ pour $n \leq k$. Il n'est pas possible de calculer les probabilités de transition en connaissant seulement la valeur de b_k . Les états et probabilités de transition de cette chaîne b_k sont représentés à la figure 6.16

2. Pour obtenir la propriété markovienne, on peut séparer l'état 0 de la chaîne en deux états distincts, notés ci-après 0_{-1} et 0_{+1} , qui correspondent respectivement à l'état 0 *en provenance* de l'état -1 et à l'état 0 *en provenance* de l'état $+1$. Les états et probabilités de transition de cette nouvelle chaîne sont représentés à la figure 6.17. Il s'agit bien cette fois d'une chaîne de Markov. Les probabilités de transition entre états sont toutes définies et ne dépendent que de l'état d'origine et de

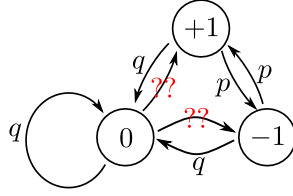


FIGURE 6.16 – Le chaîne à trois états de l'exercice 21 n'est pas une chaîne de Markov

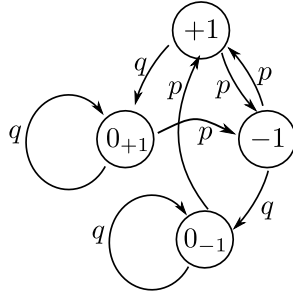


FIGURE 6.17 – Le chaîne de Markov à quatre états utilisée pour résoudre l'exercice 21

l'état d'arrivée. La matrice de transition qui correspond à cette CMTD s'écrit :

$$\mathbf{P} = \begin{pmatrix} q & 0 & 0 & p \\ 0 & q & p & 0 \\ q & 0 & 0 & p \\ 0 & q & p & 0 \end{pmatrix}$$

si les états sont rangés dans l'ordre $\{0_{+1}, 0_{-1}, +1, -1\}$

3. La chaîne de Markov à quatre états est irréductible et apériodique. La distribution limite de probabilité existe donc et est égale à la distribution stationnaire π solution de

$$\pi = \pi \mathbf{P}$$

avec la condition de normalisation

$$\sum_{i=1}^N \pi_i = 1$$

On peut résoudre facilement ce système d'équations en observant que, par symétrie, on a nécessairement $\pi_{0_{+1}} = \pi_{0_{-1}}$, d'une part, et $\pi_{+1} = \pi_{-1}$, d'autre part. La première équation

$$p\pi_{0_{+1}} = q\pi_{+1}$$

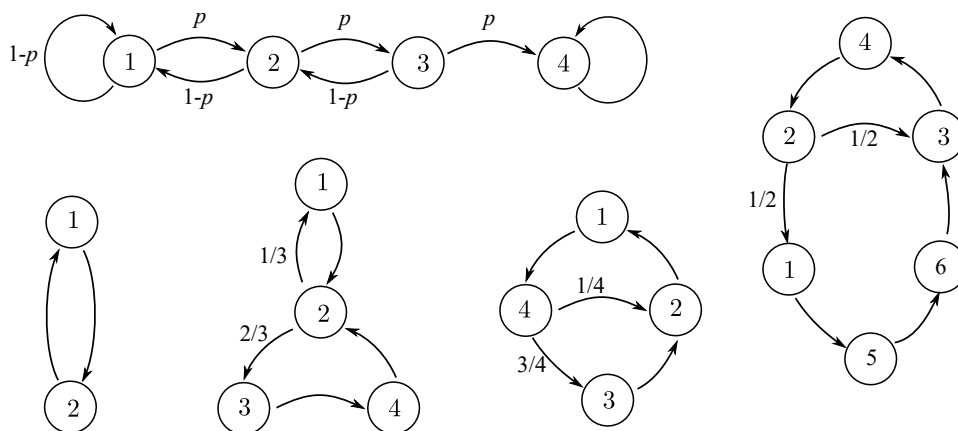


FIGURE 6.18 – Les chaînes de Markov pour l'exercice 22

et la condition de normalisation suffisent pour trouver :

$$\pi_{0+1} + \pi_{0-1} + \pi_{+1} + \pi_{-1} = \frac{2\pi_{0+1}}{q} = 1$$

c'est-à-dire

$$\pi_{0+1} = \pi_{0-1} = \frac{q}{2}$$

et

$$\pi_{+1} = \pi_{-1} = \frac{p}{2}$$

La loi limite demandée est donc :

$$\begin{aligned} \lim_{n \rightarrow \infty} IP[b_n = 0] &= \pi_{0+1} + \pi_{0-1} = q \\ \lim_{n \rightarrow \infty} IP[b_n = +1] &= \pi_{+1} = \frac{p}{2} \\ \lim_{n \rightarrow \infty} IP[b_n = -1] &= \pi_{-1} = \frac{p}{2} \end{aligned}$$

Exercice 22

Pour les cinq chaînes de Markov dont les graphes sont représentés à la figure 6.18, déterminer la matrice de transitions \mathbf{P} , les classes d'équivalence, la nature des états de ces classes (récurrents ou transitoires), ainsi que les éventuels états absorbants ou périodiques (dans le cas d'états périodiques, donner leur période).

Correction de l'Exercice 22

- Dans le cas de la CMTD représentée en haut à gauche, sur la figure 6.18, la matrice de transition s'écrit :

$$\mathbf{P} = \begin{pmatrix} 1-p & p & 0 & 0 \\ 1-p & 0 & p & 0 \\ 0 & 1-p & 0 & p \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

avec $0 < p < 1$. Les classes sont $\{1, 2, 3\}$ et $\{4\}$. L'état 4 est récurrent et absorbant. Les états 1, 2 et 3 sont transitoires. Aucun état n'est périodique.

- Dans le cas de la CMTD représentée en bas à gauche, la matrice de transition s'écrit :

$$\mathbf{P} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

Il s'agit d'une CMTD irréductible (la seule classe est $\{1, 2\} = E$). Les états 1 et 2 sont récurrents et périodiques, de période 2. La CMTD est donc 2-périodique.

- Dans le cas de la deuxième CMTD sur la rangée du bas, dans la figure 6.18, la matrice de transition s'écrit :

$$\mathbf{P} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ \frac{1}{3} & 0 & \frac{2}{3} & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{pmatrix}$$

Il s'agit d'une CMTD irréductible (la seule classe est $\{1, 2, 3, 4\} = E$). Tous les états sont récurrents. Il est possible de partir de l'état 1 et d'y revenir en 2, 5, 7, ... étapes. L'état 1 n'est donc pas périodique. En conséquence, aucun état de cette CMTD irréductible ne peut l'être. La CMTD n'est donc pas périodique.

- Dans le cas de la troisième CMTD sur la rangée du bas, dans la figure 6.18, la matrice de transition s'écrit :

$$\mathbf{P} = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & \frac{1}{4} & \frac{3}{4} & 0 \end{pmatrix}$$

Il s'agit d'une CMTD irréductible (la seule classe est $\{1, 2, 3, 4\} = E$). Tous les états sont récurrents. Il est possible de partir de l'état 1 et d'y revenir en 3, 4, 6, 7, ... étapes. L'état 1 n'est donc pas périodique. En conséquence, aucun état de cette CMTD irréductible ne peut l'être. La CMTD n'est donc pas périodique.

- Dans le cas de la CMTD à droite dans la figure 6.18, la matrice de transition s'écrit :

$$\mathbf{P} = \begin{pmatrix} 0 & 0 & 0 & 0 & 1 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{pmatrix}$$

Il s'agit d'une CMTD irréductible (la seule classe est $\{1, 2, 3, 4, 5, 6\}$). Tous les états sont récurrents. Il est possible de partir de l'état 4 et d'y revenir en 3, 6, 9, ... étapes. Tous les états de cette chaîne sont donc périodiques, de période 3.

6.3 Chaînes de Markov à temps continu

Exercice 23 (la file simple)

On considère un processeur auquel arrivent des requêtes, ces arrivées étant modélisées par un processus de Poisson de paramètre λ . Le temps que met le processeur à traiter chaque requête est modélisé par une variable aléatoire de distribution exponentielle $Exp(\mu)$, de paramètre μ . Le processeur traite la première requête qui lui arrive. Pendant ce temps, les autres requêtes qui arrivent sont rangées dans *une file d'attente*. Lorsque le processeur a terminé le traitement de la requête en cours, il commence le traitement de la requête la plus ancienne dans la file d'attente (discipline de file “*first in, first out*” (FIFO), ou *premier arrivé, premier servi*).

1. Modéliser ce problème par une chaîne de Markov à temps continu (CMTC) : donner le graphe associé avec les taux de transitions, écrire le générateur infinitésimal \mathbf{A} et donner la caractérisation terme de temps de séjour et de probabilités de transition.
2. Déterminer le régime stationnaire.
3. Quel sont - en régime stationnaire - le nombre moyen de requêtes dans le système (buffer et processeur ensemble), la taille moyenne de la file (i.e. le nombre de requêtes en attente dans le buffer) et le taux d'utilisation du processeur (en pourcentage du temps total) ?

Correction de l'Exercice 23

1. On pose $X(t) \in E := \mathbb{N}$, le nombre de requêtes dans le système (i.e. traitées par le processeur ou en attente dans la file) au temps $t > 0$. Le

processus stochastique considéré est appelé *file simple* ou encore *file M/M/1*. Il est caractérisé par des arrivées poissonniennes, des temps de service exponentiels, la présence d'un seul serveur et une gestion FIFO de la file d'attente. Il est représenté schématiquement à la figure 6.19. Nous allons montrer qu'il s'agit d'un cas particulier de processus de naissances et de morts (avec des taux de naissance et de mort constants), c'est-à-dire que les taux de transition à partir d'un état n ne sont non nuls que pour les transitions vers les états $n - 1$ (mort) ou $n + 1$ (naissance).

Montrons dans un premier temps que les taux de transition d'un état n vers un état p sont nuls pour $p \geq n + 2$.

Une telle transition sur un intervalle de temps $[t, t + dt[$ correspond à deux arrivées au moins pendant l'intervalle de temps dt ¹¹. La probabilité qu'il y ait $k \geq 2$ arrivées pendant un intervalle de temps de longueur dt , dans un processus de Poisson de paramètre λ , vaut :

$$P(N(dt) - N(0)) = k = \frac{(\lambda dt)^k}{k!} e^{-\lambda dt}$$

Ainsi, la probabilité de transition, pendant un intervalle de temps dt , entre l'état n et l'état p (avec $p \geq n + 2$) s'écrit :

$$p_{n,p}(dt) = \frac{(\lambda dt)^2}{2!} e^{-\lambda dt} \tilde{P}$$

où \tilde{P} représente la probabilité des autres événements (en plus des deux arrivées nécessaires) pour une transition de n à p pendant l'intervalle de temps dt . On a donc pour le taux de transition correspondant :

$$\lambda_{n,p} := \lim_{dt \rightarrow 0} \frac{p_{n,p}(dt)}{dt} = 0$$

De la même manière, une transition pendant un intervalle de temps de longueur dt d'un état n ($n \geq 2$) à un autre état p avec $p \leq n - 2$

11. Une transition de n à $n + 2$, par exemple, pourrait également correspondre à 3 arrivées et un départ, quatre arrivées et deux départs, etc.

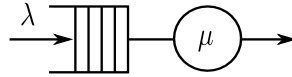


FIGURE 6.19 – Représentation schématique d'une file simple (M/M/1) : les arrivées se font selon un processus de Poisson de paramètre λ , le temps de service d'un client est distribué selon une exponentielle de paramètre μ , la file est sans limite et la discipline de service est *premier arrivé, premier servi*.

correspond à deux départs au moins pendant cet intervalle de temps. Comme les temps de service sont des exponentielles indépendantes, la probabilité d'avoir deux départs pendant l'intervalle de temps dt vaut ¹² :

$$\begin{aligned}
 P(T_1^\mu + T_2^\mu \leq dt) &= \int_0^{dt} \mu^2 \xi e^{-\mu \xi} d\xi \\
 &= -\mu dt e^{-\mu dt} + 1 - e^{-\mu dt} \\
 &= -\mu dt \left(1 - (\mu dt) + \frac{(\mu dt)^2}{2!} - \dots \right) \\
 &\quad + 1 - \left(1 - (\mu dt) + \frac{(\mu dt)^2}{2!} - \dots \right) \\
 &= \frac{(\mu dt)^2}{2} + o(dt^2), \text{ pour } dt \rightarrow 0
 \end{aligned}$$

Ainsi, la probabilité de transition, pendant un intervalle de temps dt , entre l'état n et l'état p (avec $p \leq n - 2$) s'écrit :

$$p_{n,p}(dt) = \left(\frac{(\mu dt)^2}{2} + o(dt^2) \right) \cdot \hat{P}$$

où \hat{P} représente la probabilité des autres événements (en plus des deux départs nécessaires) pour une transition de n à p pendant l'intervalle de temps dt . On a donc pour le taux de transition correspondant :

$$\lambda_{n,p} := \lim_{dt \rightarrow 0} \frac{p_{n,p}(dt)}{dt} = 0$$

Les taux de transition d'un état n vers un état p sont donc nuls pour $p \leq n - 2$. La file simple que nous considérons dans cet exercice est donc bien un exemple de *processus de naissances et de morts*. Il nous reste, pour spécifier complètement son modèle par CMTC, à calculer les taux de transition d'un état n ($n \geq 1$) vers ces deux voisins $n - 1$ et $n + 1$.

Pour une transition de n vers $n + 1$, il faut qu'il y ait 1 arrivée de requête et 0 départs, 2 arrivées et 1 départ, 3 arrivées et 2 départs, etc. Nous avons déjà montré que la probabilité associée à l'arrivée de deux requêtes (ou plus) pendant un intervalle de temps dt correspond à un taux de transition nul. Le seul cas à envisager, dans le calcul du taux de transition $\lambda_{n,n+1}$, est donc celui où il y a une arrivée de requête et aucun départ pendant un intervalle de temps dt . La probabilité correspondante est :

$$P(T^\mu > dt; T^\lambda \leq dt) = P(T^\mu > dt) \cdot P(T^\lambda \leq dt) = e^{-\mu dt} (1 - e^{-\lambda dt})$$

12. La somme de deux exponentielles indépendantes de même paramètre μ est une loi Gamma $\Gamma(\mu, 2)$.

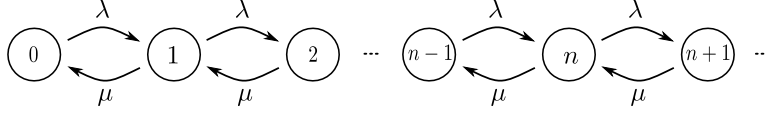


FIGURE 6.20 – La chaîne de Markov à temps continu associée à la file simple

et le taux de transition $\lambda_{n,n+1}$ vaut donc :

$$\lambda_{n,n+1} := \lim_{dt \rightarrow 0} \frac{e^{-\mu dt} (1 - e^{-\lambda dt})}{dt} = \lambda$$

On trouvera de même

$$P(T^\mu \leq dt; T^\lambda > dt) = P(T^\mu \leq dt) \cdot P(T^\lambda > dt) = (1 - e^{-\mu dt}) e^{-\lambda dt}$$

et

$$\lambda_{n,n-1} := \lim_{dt \rightarrow 0} \frac{(1 - e^{-\mu dt}) e^{-\lambda dt}}{dt} = \mu$$

Nous venons d'établir que la file simple est un processus de naissances et de morts où les taux de transition sont constants (indépendants de l'état). La CMTC correspondante est représentée à la figure 6.20. Le *générateur infinitésimal* (infini) de la file simple s'écrit donc :

$$\mathbf{A} = \begin{pmatrix} -\lambda & \lambda & 0 & \dots \\ \mu & -(\lambda + \mu) & \lambda & \dots \\ 0 & \mu & -(\lambda + \mu) & \ddots \\ \vdots & \vdots & \ddots & \ddots \end{pmatrix}$$

Le temps de séjour T_n dans un état $n \in \mathbb{N}$ est une variable exponentielle

$$T_n \sim \begin{cases} \text{Exp}(\lambda) & \text{si } n = 0 \\ \text{Exp}(\lambda + \mu) & \text{si } n \geq 1 \end{cases}$$

Les probabilités de transition s'écrivent

$$p_{n,k} = \begin{cases} \frac{\mu}{\lambda + \mu} & \text{si } k = n - 1 \\ \frac{\lambda}{\lambda + \mu} & \text{si } k = n + 1 \\ 0 & \text{sinon} \end{cases}$$

pour $n \geq 2$ et

$$p_{1,2} = 1$$

dans le cas où la chaîne est dans l'état $n = 1$.

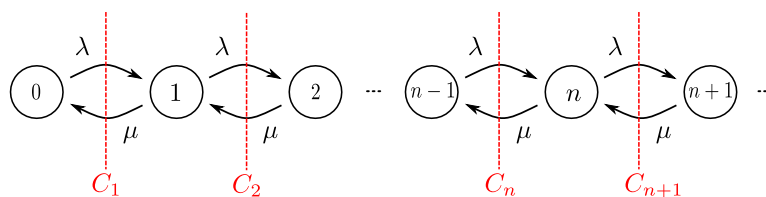


FIGURE 6.21 – Les coupes qui permettent d’obtenir la distribution stationnaire

2. Pour obtenir la distribution stationnaire de probabilité, le plus simple est d’appliquer la méthode des coupes. Les équations qui correspondent aux coupes de la figure 6.21. sont :

$$\lambda \pi_{n-1} = \mu \pi_n, \quad \forall n \geq 1$$

On obtient directement :

$$\pi_n = \pi_0 \left(\frac{\lambda}{\mu} \right)^n, \quad \forall n \geq 1$$

et la condition de normalisation donne :

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

La convergence de la série au dénominateur est donc une condition nécessaire et suffisante d’existence de la distribution stationnaire de probabilités (aussi appelée *condition d’ergodicité* de la CMTC). Elle est garantie si et seulement si

$$\lambda < \mu$$

c’est-à-dire si le temps moyen entre deux arrivées est supérieur au temps moyen mis par le processeur pour satisfaire une requête.

3. Le nombre moyen de requête dans le systèmes est

$$\begin{aligned} Q &:= \sum_{n \geq 0} n \pi_n \\ &= \pi_0 \sum_{n \geq 0} n \left(\frac{\lambda}{\mu} \right)^n \\ &= (1 - \rho) \sum_{n \geq 0} n \rho^n \end{aligned}$$

avec $\rho := \frac{\lambda}{\mu} < 1$. Or :

$$\begin{aligned}
 \sum_{n \geq 0} n \rho^n &:= \rho \sum_{n \geq 1} n \rho^{n-1} \\
 &= \rho \sum_{n \geq 1} \frac{dx^n}{dx} \Big|_{x=\rho} \\
 &= \rho \frac{d}{dx} \left[\sum_{n \geq 1} x^n \right] \Big|_{x=\rho} \\
 &= \rho \frac{d}{dx} \left[\frac{1}{1-x} \right] \Big|_{x=\rho} \\
 &= \frac{\rho}{(1-\rho)^2}
 \end{aligned}$$

Donc

$$Q = \frac{\rho}{(1-\rho)}$$

Ce nombre moyen de requêtes dans le système augmente quand $\rho \rightarrow 1$ (avec $\rho < 1$), c'est-à-dire quand le temps moyen entre deux arrivées $\frac{1}{\lambda}$ tend vers la durée moyenne de service $\frac{1}{\mu}$. La taille moyenne du buffer s'écrit

$$\begin{aligned}
 Q_f &:= \sum_{n \geq 2} (n-1) \pi_n \\
 &= \sum_{n \geq 2} n \pi_n - \sum_{n \geq 2} \pi_n \\
 &= \sum_{n \geq 0} n \pi_n - \pi_1 - (1 - \pi_0 - \pi_1) \\
 &= Q - (1 - \pi_0) \\
 &= Q - \rho \\
 &= \frac{\rho^2}{1-\rho}
 \end{aligned}$$

Le taux moyen d'utilisation du processeur est

$$U = 1 - \pi_0 = \rho$$

Exercice 24 (maintenance)

Un parking souterrain est éclairé par K ampoules allumées 24 heures sur 24, 7 jours sur 7. Dans cet exercice, on s'interroge sur la qualité des ampoules à acheter (investissement) et sur la fréquence des visites de à effectuer (maintenance). La durée de vie des ampoules est une variable aléatoire exponentielle de paramètre λ (plus λ est faible, plus les ampoules coûtent cher!).

Le concierge du parking visite de temps en temps le parking, les dates de ses visites étant espacées par un temps de loi $Exp(\mu)$. A chacune de ses visites, le concierge remplace toutes les ampoules en panne par des ampoules neuves. Il ne touche pas aux autres ¹³.

1. Modéliser l'évolution temporelle des conditions d'éclairage du parking (nombre d'ampoules qui fonctionnent) par une CMTC. Donner le graphe associé, préciser les taux de transition et écrire le générateur infinitésimal.
2. Quelle est la probabilité qu'un usager du parking se retrouve dans le noir complet quand il va chercher sa voiture? On supposera que l'installation fonctionne depuis suffisamment longtemps pour que sa configuration initiale n'ait plus d'effet sur son fonctionnement actuel (i.e. on suppose que le régime stationnaire est établi).
3. A l'aide de la méthode des coupes, calculer le nombre moyen d'ampoules qui fonctionnent (en régime stationnaire)

Correction de l'Exercice 24

1. On pose $X(t) \in E := \{0, \dots, K\}$, le nombre d'ampoules qui fonctionnent au temps $t > 0$.
Dans un premier temps, nous allons examiner les transitions associées aux visites du gardien, de l'état n (avec $n \in \{0, \dots, K-1\}$) vers l'état K . La probabilité qu'une telle transition ait lieu sur un intervalle de temps dt s'écrit :

$$p_{n,K}(dt) = P\left(T^\mu \leq dt; T_1^\lambda > dt - T^\mu; \dots; T_n^\lambda > dt - T^\mu\right)$$

où T^μ représente la durée entre deux visites du gardien et T_k^λ la durée de vie de l'ampoule k . Elle correspond à la réalisation d'événements indépendants : le gardien effectue une visite et les différentes ampoules ne tombent pas en panne après la visite du gardien. Cette probabilité

13. une action de maintenance préventive ne peut pas être optimisée à l'aide d'un modèle sous forme de CMTC où l'usure progressive des ampoules n'est pas prise en compte (absence de mémoire)

vaut donc :

$$\begin{aligned}
p_{n,K}(dt) &= P(T^\mu \leq dt) \prod_{k=1}^K P(T_k^\lambda > dt - T^\mu) \\
&= \int_0^{dt} \mu e^{-\mu\xi} \left(\prod_{k=1}^K \int_{dt-\xi}^{+\infty} \lambda e^{-\lambda\nu_k} d\nu_k \right) d\xi \\
&= \int_0^{dt} \mu e^{-\mu\xi} \left(e^{-\lambda(dt-\xi)} \right)^K d\xi \\
&= \mu e^{-\lambda K dt} \int_0^{dt} e^{\xi(K\lambda-\mu)} d\xi \\
&= \frac{\mu}{K\lambda - \mu} e^{-\lambda K dt} \left[e^{dt(K\lambda-\mu)} - 1 \right] \\
&= \frac{\mu}{K\lambda - \mu} \left[e^{-\mu dt} - e^{-\lambda K dt} \right] \\
&= \frac{\mu}{K\lambda - \mu} [1 - \mu dt + o(dt) - 1 + \lambda K dt - o(dt)] \\
&= \mu dt + o(dt), \text{ pour } dt \rightarrow 0
\end{aligned}$$

On a donc pour le taux de transition correspondant :

$$\lambda_{n,K} := \lim_{dt \rightarrow 0} \frac{p_{n,K}(dt)}{dt} = \mu, \quad \forall n \in \{0, \dots, K-1\}$$

Il est possible que, sur un intervalle de temps $[t, t + dt]$, le gardien effectue une visite, puis que des ampoules tombent en panne. Toutefois, le taux de transition correspondant est nul. Imaginons en effet, à titre d'exemple, le cas où le gardien effectue une visite et une ampoule tombe en panne après la visite. La probabilité correspondante s'écrit

$$\begin{aligned}
p_{n,K-1}(dt) &= \sum_{k=1}^K P(T^\mu \leq dt) P(T_k^\lambda \leq dt - T^\mu) \prod_{j \neq k} P(T_j^\lambda > dt - T^\mu) \\
&= K \int_0^{dt} \mu e^{-\mu\xi} \int_0^{dt-\xi} \lambda e^{-\lambda\nu} d\nu \left(\prod_{k=1}^{K-1} \int_{dt-\xi}^{+\infty} \lambda e^{-\lambda\nu_k} d\nu_k \right) d\xi \\
&= K \int_0^{dt} \mu e^{-\mu\xi} \left(e^{-\lambda(dt-\xi)} - 1 \right) \left(e^{-\lambda(dt-\xi)} \right)^{K-1} d\xi \\
&= \mu K \left[\frac{e^{-\mu dt} - e^{-\lambda K dt}}{\lambda K - \mu} - \frac{e^{-\mu dt} - e^{-\lambda(K-1)dt}}{\lambda(K-1) - \mu} \right] \\
&= \mu K \left[\frac{-\mu dt + \lambda K dt + o(dt)}{\lambda K - \mu} - \frac{-\mu dt + \lambda(K-1)dt + o(dt)}{\lambda(K-1) - \mu} \right] \\
&= \mu K o(dt), \text{ pour } dt \rightarrow 0
\end{aligned}$$

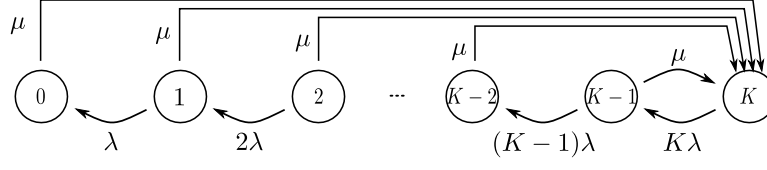


FIGURE 6.22 – La chaîne de Markov à temps continu qui modélise le nombre d'ampoules fonctionnant au temps $t > 0$

Les taux de transition correspondants sont donc

$$\lambda_{n,K-1} := \lim_{dt \rightarrow 0} \frac{p_{n,K-1}(dt)}{dt} = 0$$

pour $n \in \{0, \dots, K-2\}$. Il en va de même pour tous les taux $\lambda_{n,m}$ avec $m \in \{n+1, \dots, K-1\}$. La probabilité que deux événements ou plus¹⁴ se réalisent pendant un intervalle de temps dt sera toujours négligeable devant dt et correspondra donc à un taux de transition nul. Par contre, la probabilité de passer de l'état $n \in \{1, \dots, K\}$ à l'état $n-1$ pendant un intervalle de temps dt s'écrit :

$$\begin{aligned} p_{n,n-1}(dt) &= \sum_{k=1}^n P(T^\mu > dt) P(T_k^\lambda \leq dt) \prod_{j \neq k} P(T_j^\lambda > dt) \\ &= n \left(1 - e^{-\lambda dt}\right) e^{-\mu dt} e^{-(n-1)\lambda dt} \\ &= n\lambda dt (1 + O(dt)), \text{ pour } dt \rightarrow 0 \end{aligned}$$

où T_k^λ désigne la durée de vie d'une des n ampoules en fonctionnement. Ces probabilités correspondent à des taux :

$$\lambda_{n,n-1} := \lim_{dt \rightarrow 0} \frac{p_{n,n-1}(dt)}{dt} = n\lambda$$

Nous venons d'établir, pour le processus stochastique considéré, le modèle par CMTC représenté à la figure 6.22. Le *générateur infinitésimal* correspondant s'écrit :

$$\mathbf{A} = \begin{pmatrix} -\mu & 0 & 0 & \dots & 0 & \mu \\ \lambda & -(\lambda + \mu) & 0 & \dots & 0 & \mu \\ 0 & 2\lambda & -(2\lambda + \mu) & & \vdots & \vdots \\ \vdots & 0 & \ddots & \ddots & 0 & \mu \\ \vdots & \vdots & & (K-1)\lambda & -((K-1)\lambda + \mu) & \mu \\ 0 & 0 & \dots & 0 & K\lambda & -K\lambda \end{pmatrix}$$

14. par exemple deux ampoules qui tombent en panne, une visite du gardien et une ampoule qui tombe en panne, trois ampoules qui tombent en panne, etc.

2. La CMTC est finie et irréductible (une seule classe d'équivalence). Tous les états sont donc récurrents non nuls. La distribution limite existe et est égale à π , la distribution stationnaire de probabilités. Pour répondre à la question, il faut trouver :

$$\pi_0 := \lim_{t \rightarrow \infty} P(X(t) = 0)$$

Dans ce cas, pour obtenir la distribution stationnaire de probabilité, le plus simple est d'écrire les équations d'équilibre en chacun des noeuds du graphe (i.e. en chacun des états). On obtient pour le bilan en 0 :

$$\lambda \pi_1 = \mu \pi_0$$

Pour le noeud n avec $n \in \{1, \dots, K-1\}$, l'équilibre s'écrit

$$(n+1)\lambda \pi_{n+1} = (n\lambda + \mu) \pi_n$$

et l'équilibre au noeud K s'écrit :

$$\mu \sum_{j=1}^{K-1} \pi_j = K\lambda \pi_K$$

On obtient ainsi

$$\begin{aligned} \pi_1 &= \frac{\mu}{\lambda} \pi_0 \\ \pi_2 &= \frac{(\lambda + \mu)\mu}{2\lambda^2} \pi_0 \\ &\vdots \\ \pi_n &= \frac{(\lambda + \mu)^{n-1} \mu}{n! \lambda^n} \pi_0 \\ &\vdots \\ \pi_K &= \frac{(\lambda + \mu)^{K-1} \mu}{K! \lambda^K} \pi_0 \end{aligned}$$

et la condition de normalisation donne :

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

On a donc :

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

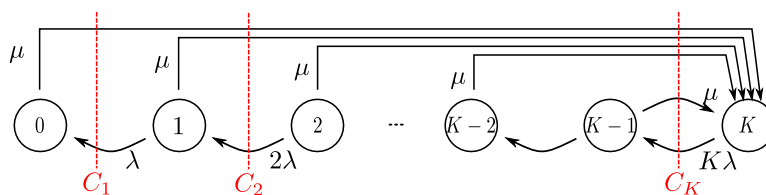


FIGURE 6.23 – Les coupes qui permettent d’obtenir plus facilement, dans l’exemple des ampoules, la valeur moyenne de la distribution stationnaire

3. Nous allons calculer les équations issues de la méthode des coupes, avec les coupes illustrées à la figure 6.23. Les équations qui correspondent aux coupes C_n , pour $n \in \{1, \dots, K\}$ s’écrivent :

$$n\lambda\pi_n = \mu \sum_{k=0}^{n-1} \pi_k$$

En additionnant ces équations, et en posant

$$\bar{N} = E(X_\infty) := \sum_{n=0}^K n\pi_n = \sum_{n=1}^K n\pi_n$$

on trouve

$$\begin{aligned} \lambda\bar{N} &= \sum_{n=0}^K \lambda n\pi_n \\ &= \mu \sum_{n=0}^K (K-n)\pi_n \\ &= \mu \left(K \sum_{n=0}^K \pi_n - \sum_{n=0}^K n\pi_n \right) \\ &= \mu(K - \bar{N}) \end{aligned}$$

c’est-à-dire

$$\bar{N} = \frac{\mu K}{\lambda + \mu} = \frac{1}{1 + \frac{\lambda}{\mu}} \cdot K$$

Exercice 25

Une salle est équipée de n ordinateurs. La durée de fonctionnement de chacun des ordinateurs suit une loi $Exp(\mu)$. La durée de remise en état de chaque ordinateur suit une loi $Exp(\lambda)$. Le réseau est maintenu en service par n techniciens. Toutes les durées sont indépendantes. On note $X(t)$ le nombre d’ordinateurs qui fonctionnent à l’instant t .

1. Donner le graphe et les taux de transition pour $X(t)$
2. Donner la distribution stationnaire de probabilités
3. Donner, en régime stationnaire, le nombre moyen d'ordinateurs en fonctionnement.
4. Donner le graphe si la salle ne possède plus qu'un technicien

Correction de l'Exercice 25

1. On pose $X(t) \in E := \{0, \dots, n\}$, le nombre d'ordinateurs qui fonctionnent au temps $t > 0$. Il est facile de calculer, en reproduisant les mêmes raisonnements que dans les deux exercices précédents, les taux de transition entre les états de la CMTC qui modélise ce processus stochastique. Les taux qui correspondent aux transitions d'un état $k \in \{0, \dots, n-2\}$ vers un état $l \geq k+2$ sont tous nuls. De même, Les taux qui correspondent aux transitions d'un état $k \in \{2, \dots, n\}$ vers un état $l \leq k-2$ sont tous nuls également. La probabilité d'une transition entre l'état k et l'état $k+1$ (pour $k \in \{0, \dots, n-1\}$), pendant un intervalle de temps dt , correspond à la fin de la réparation d'un des $n-k$ ordinateurs en cours de réparation pendant cet intervalle de temps. Cette probabilité s'écrit :

$$\begin{aligned}
 p_{k,k+1}(dt) &= \sum_{p=1}^{n-k} P\left(T_p^\lambda \leq dt\right) \prod_{\substack{j=1 \\ j \neq p}}^{n-k} P\left(T_j^\lambda > dt\right) \prod_{l=1}^k P\left(T_l^\mu > dt\right) \\
 &= (n-k)(1 - e^{-\lambda dt}) \prod_{\substack{j=1 \\ j \neq p}}^{n-k} e^{-\lambda dt} \prod_{l=1}^k e^{-\mu dt} \\
 &= (n-k)\lambda dt (1 + O(dt)), \text{ pour } dt \rightarrow 0
 \end{aligned}$$

où T_l^μ représente la durée de fonctionnement de l'ordinateur l (parmi les k en fonctionnement) et T_p^λ représente la durée de réparation de l'ordinateur p (parmi les $n-k$ en cours de réparation). Le taux de transition correspondant vaut :

$$\lambda_{k,k+1} := \lim_{dt \rightarrow 0} \frac{p_{k,k+1}(dt)}{dt} = (n-k)\lambda, \quad \forall k \in \{0, \dots, n-1\} \quad (6.11)$$

De la même manière, la probabilité d'une transition entre l'état k et l'état $k-1$ (pour $k \in \{1, \dots, n\}$), pendant un intervalle de temps

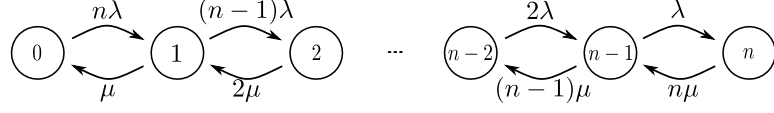


FIGURE 6.24 – La chaîne de Markov à temps continu qui modélise le nombre d'ordinateurs fonctionnant au temps $t > 0$

dt , correspond à la panne d'un des k ordinateurs en fonctionnement pendant cet intervalle de temps. Cette probabilité s'écrit :

$$\begin{aligned}
 p_{k,k-1}(dt) &= \sum_{p=1}^k P(T_p^\mu \leq dt) \prod_{j=1}^{n-k} P(T_j^\lambda > dt) \prod_{\substack{l=1 \\ l \neq p}}^k P(T_l^\mu > dt) \\
 &= k(1 - e^{-\mu dt}) \prod_{j=1}^{n-k} e^{-\lambda dt} \prod_{\substack{l=1 \\ l \neq p}}^k e^{-\mu dt} \\
 &= k\mu dt (1 + O(dt)), \text{ pour } dt \rightarrow 0
 \end{aligned}$$

Le taux de transition correspondant vaut :

$$\lambda_{k,k-1} := \lim_{dt \rightarrow 0} \frac{p_{k,k-1}(dt)}{dt} = k\mu, \quad \forall k \in \{1, \dots, n\}$$

Nous venons d'établir, pour le processus stochastique considéré, le modèle par CMTC représenté à la figure 6.24. Le *générateur infinitésimal* correspondant s'écrit :

$$\begin{pmatrix}
 -n\lambda & n\lambda & 0 & 0 & \dots & 0 \\
 \mu & -\mu - (n-1)\lambda & (n-1)\lambda & 0 & \dots & 0 \\
 0 & 2\mu & -2\mu - (n-2)\lambda & (n-2)\lambda & \ddots & \vdots \\
 \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\
 \vdots & \ddots & \ddots & (n-1)\mu & -(n-1)\mu - \lambda & \lambda \\
 0 & 0 & \dots & 0 & n\mu & -n\mu
 \end{pmatrix}$$

- les équations d'équilibre en chaque noeud sont des récurrences à trois termes. Il sera plus facile ici d'écrire plutôt les équations issues de la méthode des coupes, avec les coupes illustrées à la figure 6.25. Les équations qui correspondent aux coupes C_k , pour $k \in \{1, \dots, n\}$ s'écrivent :

$$(n - (k - 1)) \lambda \pi_{k-1} = k\mu \pi_k \quad (6.12)$$

On obtient directement :

$$\pi_k = \frac{n!}{k!(n-k)!} \left(\frac{\lambda}{\mu}\right)^k \pi_0$$

Et la condition de normalisation s'écrit

$$\sum_{k=0}^n \pi_k = \pi_0 \sum_{k=0}^n \frac{n!}{k!(n-k)!} \left(\frac{\lambda}{\mu}\right)^k = \pi_0 \left(1 + \frac{\lambda}{\mu}\right)^n = 1$$

c'est-à-dire

$$\pi_0 = \frac{\mu^n}{(\mu + \lambda)^n}$$

3. On définit le nombre moyen d'ordinateurs qui fonctionnent par

$$\bar{N} = E(X_\infty) := \sum_{n=0}^K n\pi_n = \sum_{n=1}^K n\pi_n$$

La somme des équations (6.12) donne

$$\lambda \sum_{k=1}^n (n - (k-1)) \pi_{k-1} = \mu \sum_{k=1}^n k \pi_k$$

c'est-à-dire

$$\lambda [n(1 - \pi_n) - (\bar{N} - n\pi_n)] = \mu \bar{N}$$

On trouve ainsi

$$\bar{N} = \frac{\lambda}{\lambda + \mu} n$$

4. Dans le cas où il n'y a qu'un seul technicien, un seul ordinateur peut être réparé à la fois. Il faut donc remplacer les taux de transition de l'équation (6.11) par les taux

$$\lambda_{k,k+1} := \lambda, \forall k \in \{0, \dots, n-1\}$$

On obtient alors la CMTC illustrée à la figure 6.26.

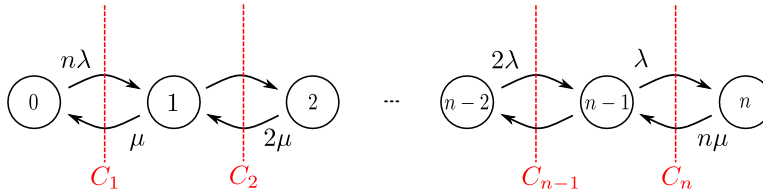


FIGURE 6.25 – Les coupes qui permettent d'obtenir plus facilement, dans l'exemple de la salle d'ordinateurs, la distribution stationnaire de probabilités

Exercice 26 (croissance logistique)

Nous nous intéressons dans cet exercice à l'évolution dynamique de la taille d'une population. Des modèles par chaînes de Markov sont fréquemment utilisés pour établir ces bilans de population, que ce soit en biologie (e.g. croissance bactérienne, modèles proies-prédateurs, etc.), en démographie humaine ou en physique (e.g. désintégration de particules radioactives). Dans notre cas, nous allons nous intéresser à un processus de naissances et de morts, pour lequel le générateur infinitésimal est tridiagonal : à partir d'une population de taille k , seules les transitions vers les tailles $k - 1$ et $k + 1$ sont possibles. Nous allons en particulier nous intéresser au *modèle logistique*. Dans celui-ci, la taille de la population peut évoluer entre une taille minimale \underline{X} et une taille maximale \bar{X} (avec $0 \leq \underline{X} < \bar{X}$). Dans le modèle logistique, le taux de naissance¹⁵ d'une population de taille X est donné par :

$$\lambda = \alpha (\bar{X} - X)$$

où α est un paramètre réel positif qui représente la capacité de reproduction de l'espèce pour une population de taille donnée. Le taux de natalité λ diminue donc au fur et à mesure que la population croît en taille, jusqu'à atteindre une valeur nulle pour $X = \bar{X}$ (valeur de saturation). C'est une manière de représenter la raréfaction des ressources nécessaires à la survie de la population quand la taille de celle-ci augmente. De même, le taux de mortalité d'une population de taille X , dans le modèle logistique, est donné par :

$$\mu = \beta (X - \underline{X})$$

où β est un paramètre réel positif. Le taux de mortalité diminue au fur et à mesure que la population diminue en taille, jusqu'à atteindre une valeur nulle pour $X = \underline{X}$ (valeur de saturation). C'est une manière de représenter l'abondance de ressources et la plus grande facilité à survivre quand la taille de la population est faible.

1. Modéliser par une CMTC l'évolution dans le temps de la taille de la population du modèle logistique. Donner le graphe associé et les taux de transition.
2. Calculer la distribution de probabilité limite de cette population

15. Le taux de naissance se définit comme la variation par unité de temps de la taille de la population rapportée à la taille de la population



FIGURE 6.26 – LLa CMTC qui modélise la salle d'ordinateurs avec un seul technicien

3. On considère le cas particulier où $\underline{N} = 0$ et où lorsque la taille de la population devient nulle, aucune reproduction n'est plus possible. Déterminer quelle est la probabilité P_k d'extinction de l'espèce si à l'instant initial la population est de taille k ¹⁶. Calculer ensuite le temps moyen T_k d'extinction de l'espèce en partant d'une population initiale de taille k .

Correction de l'Exercice 26

1. La taille de la population est comprise entre \underline{X} et \overline{X} . Nous choisirons l'écart par rapport à la population minimale, $X(t) := N(t) - \underline{X}$, comme état du système au temps t ($N(t)$ désigne la population totale au temps t , en valeur absolue). On a donc :

$$X(t) \in E := \{0, \dots, N\}$$

où $N := \overline{X} - \underline{X}$. Les taux de transition de l'état n à l'état $n + 1$ s'écrivent, sous les hypothèses d'une croissance logistique :

$$\lambda_{n,n+1} := \alpha(\underline{X} + n)(N - n), \forall n \in \{0, \dots, N - 1\}$$

De même, sous les hypothèses de décroissance logistique, les taux de transition de l'état n à l'état $n - 1$ s'écrivent :

$$\lambda_{n,n-1} := \beta(\underline{X} + n)n, \forall n \in \{1, \dots, N\}$$

La chaîne de Markov qui correspond au problème est représentée à la figure 6.27.

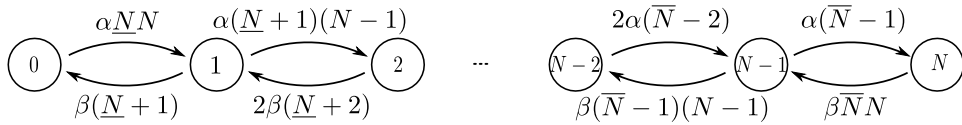


FIGURE 6.27 – La chaîne de Markov associée au modèle logistique, avec une population totale comprise entre \underline{X} et \overline{X} . La variable d'état est $X(t) := N(t) - \underline{X}$ où $N(t)$ désigne la quantité totale de population

2. Il s'agit d'une CMTC finie et irréductible. Tous les états sont récurrents non nuls. La distribution limite existe donc nécessairement. Elle est égale à la distribution stationnaire π . Les équations d'équilibre en

16. Passer par la caractérisation de la CMTC en termes de temps de séjour et de probabilités de transition. Résoudre le problème directement sur la CMTD induite par la CMTC

chaque noeud sont des récurrences à trois termes. Il sera plus facile ici d'écrire plutôt les équations issues de la méthode des coupes. Les équations qui correspondent aux coupes C_{k+1} , pour $k \in \{0, \dots, N-1\}$ s'écrivent :

$$\alpha(\underline{N} + k)(N - k)\pi_k = (k + 1)\beta(\underline{N} + k + 1)\pi_{k+1}$$

On obtient directement :

$$\pi_k = \frac{N!}{k!(N - k)!} \frac{\underline{N}}{\underline{N} + k} \left(\frac{\alpha}{\beta}\right)^k \pi_0 = C_N^k \frac{\underline{N}}{\underline{N} + k} \left(\frac{\alpha}{\beta}\right)^k \pi_0$$

Et la condition de normalisation s'écrit

$$\sum_{k=0}^N \pi_k = \pi_0 \sum_{k=0}^N C_N^k \frac{\underline{N}}{\underline{N} + k} \left(\frac{\alpha}{\beta}\right)^k = 1$$

c'est-à-dire

$$\pi_0 = \frac{1}{\sum_{k=0}^N C_N^k \frac{\underline{N}}{\underline{N} + k} \left(\frac{\alpha}{\beta}\right)^k}$$

3. Dans le cas où $\underline{N} = 0$, la CMTC se réduit à celle représentée à la figure 6.28. L'état 0 devient un état absorbant. Il correspond à l'extinction de l'espèce. On peut calculer la représentation équivalente de la CMTC

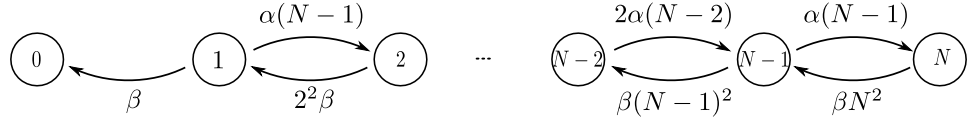


FIGURE 6.28 – La chaîne de Markov associée au modèle logistique, avec une population totale comprise entre $\underline{X} = 0$ et $\bar{X} = N$.

de la figure 6.28 en termes de probabilités de transition et de temps de séjour. Le temps de séjour dans l'état k est distribué selon une variable exponentielle $\text{Exp}(\mu_k)$ de paramètre :

$$\mu_k = \lambda_{k,k+1} + \lambda_{k,k-1} = k(\alpha(N - k) + \beta k)$$

Les probabilités de transition d'un état k vers ses voisins $k - 1$ et $k + 1$ s'écrivent respectivement :

$$\begin{aligned} p_{k,k-1} &= \frac{\beta k}{\alpha(N - k) + \beta k} \\ p_{k,k+1} &= \frac{\alpha(N - k)k}{\alpha(N - k) + \beta k} \end{aligned}$$

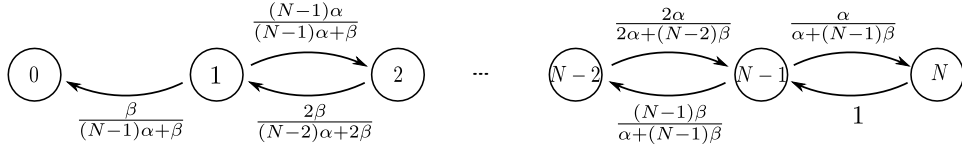


FIGURE 6.29 – La chaîne de Markov à temps discret induite par la CMTC du modèle logistique pour le problème de l’extinction de l’espèce

On obtient ainsi la figure 6.29 où sont représentées les probabilités de transition entre états (et donc la CMTD induite de la CMTC). L’état 0 devient un état absorbant. Il correspond à l’extinction de l’espèce. On sait donc par avance que $P_k = 1$ (pour tout $k \in \{1, \dots, N\}$). On peut d’ailleurs facilement retrouver ce résultat en appliquant la formule de calcul des probabilités $f_{k,0} = P_k$ vue en cours. L’application de cette formule donne ici

$$\begin{aligned}
 ((N-1)\alpha + \beta)P_1 &= \beta + (N-1)\alpha P_2 \\
 &\vdots \\
 ((N-k)\alpha + k\beta)P_k &= k\beta P_{k-1} + (N-k)\alpha P_{k+1} \\
 &\vdots \\
 (\alpha + (N-1)\beta)P_{N-1} &= (N-1)\beta P_{N-2} + \alpha P_N \\
 P_N &= P_{N-1}
 \end{aligned}$$

dont la seule solution est bien

$$P_k = 1, \forall k \in \{1, \dots, N\}$$

Le calcul du temps moyen T_k avant l’extinction, partant de l’état k suit le même raisonnement mais est un peu plus compliqué car il faut intégrer les temps moyen de séjour dans les états visités. On obtient ainsi :

$$\begin{aligned}
 T_1 &= \frac{\beta}{(N-1)\alpha + \beta} \cdot \frac{1}{\mu_1} + \frac{(N-1)\alpha}{(N-1)\alpha + \beta} \cdot \left(\frac{1}{\mu_1} + T_2 \right) \\
 &\vdots \\
 T_k &= \frac{k\beta}{(N-k)\alpha + k\beta} \cdot \left(\frac{1}{\mu_k} + T_{k-1} \right) + \frac{(N-k)\alpha}{(N-k)\alpha + k\beta} \cdot \left(\frac{1}{\mu_k} + T_{k+1} \right) \\
 &\vdots \\
 T_{N-1} &= \frac{(N-1)\beta}{\alpha + (N-1)\beta} \cdot \left(\frac{1}{\mu_{N-1}} + T_{N-2} \right) + \frac{\alpha}{\alpha + (N-1)\beta} \cdot \left(\frac{1}{\mu_{N-1}} + T_N \right) \\
 T_N &= \frac{1}{\mu_N} + T_{N-1}
 \end{aligned}$$

qui s'écrit, après simplifications :

$$\begin{aligned}
 T_1 &= \frac{1}{\mu_1} + \frac{(N-1)\alpha}{(N-1)\alpha + \beta} T_2 \\
 &\vdots \\
 T_k &= \frac{1}{\mu_k} + \frac{k\beta}{(N-k)\alpha + k\beta} T_{k-1} + \frac{(N-k)\alpha}{(N-k)\alpha + k\beta} T_{k+1} \\
 &\vdots \\
 T_N &= \frac{1}{\mu_N} + T_{N-1}
 \end{aligned}$$

En posant $W_k := T_k - T_{k+1}$, pour $k \in \{0, \dots, N-1\}$ et $T_0 = 0$, on obtient

$$W_{N-1} = -\frac{1}{\mu_N}$$

et

$$W_{N-(k+1)} = \frac{k\alpha}{(N-k)\beta} \left(W_{N-k} - \frac{1}{\mu_{N-k}} \right) - \frac{1}{\mu_{N-k}}$$

qui permet de calculer à rebours toutes les grandeurs W_k jusqu'à $W_0 = T_1$. On retrouve ensuite les temps moyen d'extinction demandé en calculant successivement

$$T_k = T_{k-1} - W_{k-1}$$

pour k allant de 2 à N .

Exercice 27

On considère le réseau de transmission par paquets, représenté à la figure 6.30, constitué d'un émetteur E , de deux noeuds relais en série notés respectivement 1, 2 et d'un récepteur R . Les noeuds 1 et 2 ont une capacité

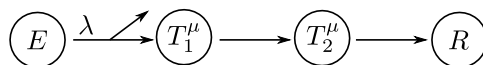


FIGURE 6.30 – réseau de transmission avec deux noeuds relais en série

limitée à un seul paquet (pas de buffer). Les paquets sont émis par E selon un processus de Poisson de paramètre λ . Lorsqu'un paquet arrive au noeud 1 alors que ce dernier est déjà occupé, le paquet est perdu. Un paquet ne peut être émis du noeud 1 vers le noeud 2 que lorsque le noeud 2 est vide. Le temps d'émission des noeuds 1 et 2 sont distribués selon des variables exponentielles de paramètre μ . Tant qu'un paquet est en cours d'émission, on considère qu'il occupe une place sur le noeud de départ. Le noeud 1 est

informé instantanément de l'état du noeud 2 (pour décider de l'émission éventuelle d'un paquet). Lorsque R reçoit un paquet, il le traite instantanément et peut immédiatement recevoir un autre paquet éventuellement envoyé par le noeud 2 (le noeud R représente la *sortie* du réseau). Le temps de propagation sur les liaisons est négligeable devant le temps d'émission.

1. Modéliser l'état du réseau de transmission par une chaîne de Markov à temps continu à quatre états (définir précisément la signification de chacun des quatre états). Donner le graphe et le générateur infinitésimal (matrice A des taux de transition).
2. En utilisant les relations d'équilibre en chaque noeud, déterminer le profil stationnaire de probabilités. S'agit-il d'une distribution limite? Pourquoi?
3. Donner la probabilité de rejet des paquets
4. Représenter le graphe de la CMTC associée au réseau de la figure 6.31 et indiquer les taux associés à chacune des transitions. Si les noeuds 2 et 3 sont vides, un noeud qui quitte le noeud 1 est routé de manière équiprobable vers les noeuds 2 ou 3. Si un seul des deux noeuds 2 ou 3 est vide, alors un paquet quittant le noeud 1 est routé vers le paquet libre. Si les deux noeuds 2 et 3 sont occupés, les paquets sont bloqués en amont au noeud 1.

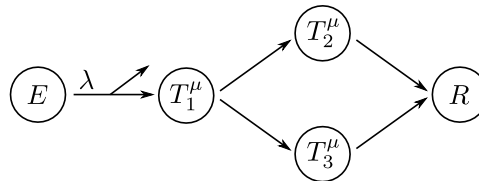


FIGURE 6.31 – réseau de transmission avec un noeud relais en série avec deux autres noeuds relais en parallèle

Correction de l'Exercice 27

1. L'état du système est déterminé par les états des noeuds 1 et 2 (pouvant être chacun vide ou occupé) qui conditionnent la transmission ou non des paquets à travers le réseau. On pose

$$X(t) \in E := \{00, 01, 10, 11\}$$

L'état 00 correspond à l'état du système où les noeuds 1 et 2 sont vides.
 L'état 01 correspond au cas où le noeud 1 est vide et le noeud 2 occupé.
 L'état 10 correspond au cas où le noeud 1 est occupé et le noeud 2

vide. Enfin, l'état 11 correspond au cas où les noeuds 1 et 2 sont occupés tous les 2. La chaîne de Markov qui correspond au problème est représentée à la figure 6.32. En rangeant les états dans l'ordre

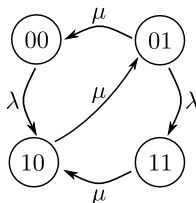


FIGURE 6.32 – La chaîne de Markov à temps continu qui modélise le système de transmission par paquets avec deux relais (quatre états)

“naturel”¹⁷ (00, 01, 10, 11), le *générateur infinitésimal* correspondant s’écrit :

$$\mathbf{A} = \begin{pmatrix} -\lambda & 0 & \lambda & 0 \\ \mu & -(\mu + \lambda) & 0 & \lambda \\ 0 & \mu & -\mu & 0 \\ 0 & \mu & 0 & -\mu \end{pmatrix}$$

2. Il s’agit d’une CMTC finie et irréductible. Tous les états sont récurrents non nuls. La distribution limite existe donc nécessairement. Elle est égale à la distribution stationnaire π . En écrivant les équations d’équilibre en chaque noeud, on obtient :

$$\begin{aligned} \lambda\pi_{00} &= \mu\pi_{01} \\ (\lambda + \mu)\pi_{01} &= \mu\pi_{10} \\ \lambda\pi_{00} + \mu\pi_{11} &= \mu\pi_{10} \\ \lambda\pi_{01} &= \mu\pi_{11} \end{aligned} \tag{6.13}$$

avec la condition de normalisation :

$$\pi_{00} + \pi_{01} + \pi_{10} + \pi_{11} = 1$$

La résolution de trois équations indépendantes¹⁸ parmi les quatre équations (6.13) donne :

$$\begin{aligned} \pi_{01} &= \frac{\lambda}{\mu}\pi_{00} \\ \pi_{10} &= \frac{\lambda(\lambda + \mu)}{\mu^2}\pi_{00} \\ \pi_{11} &= \frac{\lambda^2}{\mu^2}\pi_{00} \end{aligned}$$

17. associé à l’ordre des nombres naturels dont l’écriture en base 2 est l’état

18. Les quatre équations (6.13) ne peuvent pas être linéairement indépendantes car 0 est nécessairement valeur propre de la matrice \mathbf{A}

et la condition de normalisation permet d'obtenir

$$\pi_{00} = \frac{\mu^2}{\lambda^2 + (\lambda + \mu)^2} = \frac{1}{\rho^2 + (1 + \rho)^2}$$

avec

$$\rho := \frac{\lambda}{\mu}$$

3. La probabilité de rejet d'un paquet (ou le taux de paquets perdus) est :

$$\pi_{10} + \pi_{11} = \frac{2\rho^2 + \rho}{\rho^2 + (1 + \rho)^2} = \frac{1}{1 + \frac{1 + \rho}{\rho(1 + 2\rho)}}$$

4. Dans ce cas, l'état du réseau est déterminé par l'état, vide ou occupé, des trois relais 1, 2 et 3. Il y a donc 8 états possibles. Nous les numérotons $d_1 d_2 d_3$ où $d_i \in \{0, 1\}$ ($i \in \{1, 2, 3\}$) désigne l'état du noeud i (0 pour vide et 1 pour occupé). En rangeant les états dans l'ordre "naturel"¹⁹ (000, 001, 010, 011, 100, 101, 110, 111), la chaîne de Markov qui correspond au problème est représentée à la figure 6.33. Le générateur infinitésimal correspondant s'écrit :

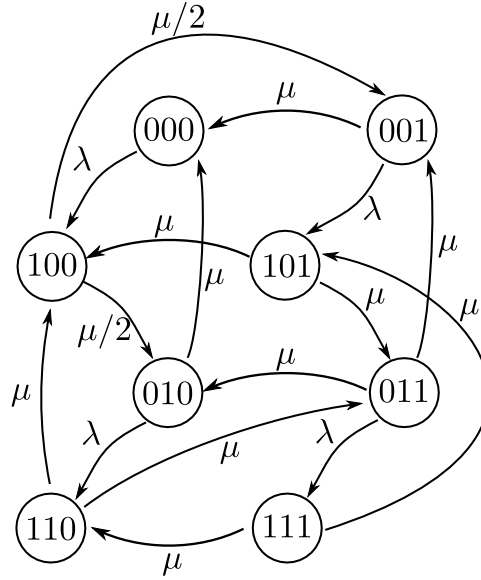


FIGURE 6.33 – La chaîne de Markov à temps continu qui modélise le système de transmission par paquets avec trois relais (huit états), dont deux en parallèle

19. associé à l'ordre des nombres naturels dont l'écriture en base 2 est l'état

$$\mathbf{A} = \begin{pmatrix} -\lambda & \lambda & 0 & 0 & 0 & 0 & 0 & 0 \\ \mu & -(\mu + \lambda) & 0 & 0 & 0 & \lambda & 0 & 0 \\ 0 & \mu & -(\mu + \lambda) & 0 & 0 & 0 & \lambda & 0 \\ 0 & \mu & \mu & -(2\mu + \lambda) & 0 & 0 & 0 & \lambda \\ 0 & \mu/2 & \mu/2 & 0 & -\mu & 0 & 0 & 0 \\ 0 & 0 & 0 & \mu & \mu & -2\mu & 0 & 0 \\ 0 & 0 & 0 & \mu & \mu & 0 & -2\mu & 0 \\ 0 & 0 & 0 & 0 & 0 & \mu & \mu & -2\mu \end{pmatrix}$$

6.4 Files et réseaux de files d'attente

Exercice 28

Dans une école, se proposent deux choix : soit une installer une imprimante, soit installer 3 imprimantes indépendantes, mais 3 fois moins rapides que dans le premier cas. Pour simplifier le calcul, le processus d'arrivée des requêtes d'impression sera modélisé comme un processus de Poisson. Les temps moyen de traitement des requêtes d'impression sont supposés distribués selon des lois exponentielles indépendantes.

1. Dans un premier temps, la configuration envisagée pour la solution à trois imprimantes est de les doter chacune d'un spooler d'impression indépendant. On suppose que l'installation des trois imprimantes peut être faite de manière à ce que les requêtes d'impression soit équitablement distribuée – en régime permanent – entre les trois imprimantes. Quel est le meilleur choix à faire, en terme de temps de réponse (confort de l'utilisateur) et en terme de stockage (nombre de requêtes stockées) entre les deux configurations ?
2. La solution avec une seule imprimante étant jugée trop fragile et inconfortable pour les usagers, on cherche une configuration plus efficace pour la solution à trois imprimantes. On regarde alors une solution où les trois imprimantes travaillent en parallèle, mais sont alimentées par un spooler d'impression unique. Comparer cette solution avec les deux précédentes en terme de temps de réponse.

Correction de l'Exercice 28

1. Les deux configurations d'imprimantes proposées sont représentées à la figure 6.34. La configuration de gauche est une file $M/M/1$. Pour étudier son fonctionnement en régime permanent (donc sur le long terme), nous devons supposer sa stabilité, c'est-à-dire $\lambda < \mu$. Si c'est le cas, le temps de service d'une requête (temps moyen de séjour d'un

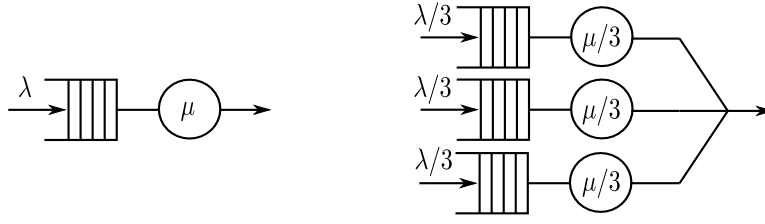


FIGURE 6.34 – Les deux configurations du système d'impression étudiées dans l'exercice 28 (premier point)

client dans le système) est :

$$R = \frac{1}{\mu(1 - \rho)} = \frac{1}{\mu} + \frac{\rho}{\mu(1 - \rho)}$$

avec pour taux d'utilisation de l'imprimante :

$$U = \rho := \frac{\lambda}{\mu}$$

La configuration de droite est la mise en parallèle de trois files $M/M/1$. La condition de stabilité du réseau est la stabilité de chacune des trois files, c'est-à-dire $\lambda/3 < \mu/3$. Le réseau est donc stable si $\lambda < \mu$, comme dans le cas de la première configuration. Désignons respectivement par \tilde{X}_i et \tilde{Q}_i , les débits moyens et les nombres de requêtes ($i \in \{1, 2, 3\}$) dans chacune de ces trois files (en régime stationnaire). En appliquant la loi de Little à l'ensemble du système constitué des trois files, on peut calculer le temps moyen de séjour d'un client :

$$\tilde{R} = \frac{\tilde{Q}_1 + \tilde{Q}_2 + \tilde{Q}_3}{\tilde{X}_1 + \tilde{X}_2 + \tilde{X}_3}$$

En régime stationnaire, la condition de stabilité donne

$$\tilde{X}_1 = \tilde{X}_2 = \tilde{X}_3 = \frac{\lambda}{3}$$

Par ailleurs, la formule qui donne le nombre moyen de clients dans une file $M/M/1$ permet d'écrire :

$$\tilde{Q}_1 = \tilde{Q}_2 = \tilde{Q}_3 = \frac{\tilde{\rho}}{1 - \tilde{\rho}}$$

avec

$$\tilde{\rho} := \frac{\lambda/3}{\mu/3} = \rho$$

On obtient donc

$$\tilde{R} = \frac{3\rho}{\lambda(1 - \rho)} = \frac{3}{\mu(1 - \rho)} = 3 \left(\frac{1}{\mu} + \frac{\rho}{\mu(1 - \rho)} \right) = 3R$$

La première configuration est donc meilleure en terme de temps moyen de séjour des requêtes d'impression dans le système. Elle est d'ailleurs préférable également en terme de nombre de requêtes présentes en moyenne dans le système puisque :

$$\tilde{Q} = \tilde{Q}_1 + \tilde{Q}_2 + \tilde{Q}_3 = \frac{3\rho}{1-\rho} = 3Q$$

si Q et \tilde{Q} désignent respectivement le nombre moyen de requêtes en impression dans le premier et dans le deuxième système.

2. La nouvelle configuration envisagée est représentée à la figure 6.35. Il s'agit cette fois d'une file de type $M/M/C$ avec $C = 3$ serveurs de taux de service $\hat{\mu} := \mu/3$ et une discipline de service $FCFS$. La condition

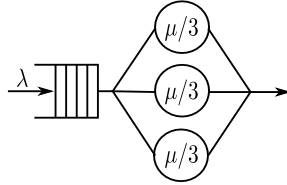


FIGURE 6.35 – La configuration envisagée pour le système d'impression dans l'exercice 28 (deuxième point)

de stabilité pour la file $M/M/3$ donne

$$\lambda < 3 \cdot \frac{\mu}{3} = \mu$$

C'est la même condition de stabilité que pour les deux premières configurations. En appliquant les formules qui donnent le régime stationnaire pour une file $M/M/C$ avec $C = 3$ et un ratio

$$\hat{\rho} := \frac{\hat{\lambda}}{\hat{\mu}} = \frac{\lambda}{\frac{\mu}{3}} = 3\rho$$

on obtient :

$$\hat{R} = \frac{1}{\hat{\mu}} + \frac{\hat{\rho}^3}{2\hat{\mu}(3-\hat{\rho})^2} \pi_0$$

avec

$$\pi_0 = \frac{1}{1 + \hat{\rho} + \frac{\hat{\rho}^2}{2} + \frac{\hat{\rho}^3}{2(3-\hat{\rho})}} = \frac{2(\rho-1)}{3\rho^2 + 4\rho + 2}$$

Au final, on obtient ainsi :

$$\begin{aligned}
 \hat{R} &= \frac{3}{\mu} \left(1 + \frac{3\rho^3}{2(1-\rho)^2} \pi_0 \right) \\
 &= \frac{3}{\mu} \left(1 + \frac{3\rho^3}{(1-\rho)(3\rho^2 + 4\rho + 2)} \right) \\
 &= \underbrace{\frac{3}{\mu} \left(1 + \frac{\rho}{1-\rho} \right)}_{3R} - \frac{3\rho}{\mu(1-\rho)} \left[1 - \frac{3\rho}{3\rho^2 + 4\rho + 2} \right] \\
 &= 3R - \frac{3\rho}{\mu(1-\rho)} \cdot \frac{3\rho^2 + \rho + 2}{3\rho^2 + 4\rho + 2}
 \end{aligned}$$

On montre alors que pour $0 < \rho < 1$, on a toujours :

$$R < \hat{R} < 3R = \tilde{R}$$

La dernière configuration est donc un meilleur compromis si on veut malgré tout une solution avec trois imprimantes.

Exercice 29

Dans cet exercice, on s'intéresse en particulier à l'effet du routage dynamique dans des files d'attente aux guichets des administrations, dans les grands magasins, etc. On dispose de deux serveurs (par exemple deux employés) qui traitent des requêtes similaires avec la même efficacité. Le traitement des requêtes (de durée variée, en général) nécessite un temps de service modélisé par une exponentielle de paramètre μ .

1. Dans un premier temps, on cherche à comparer les deux organisations représentées à la figure 6.36. Dans l'organisation à l'américaine (à

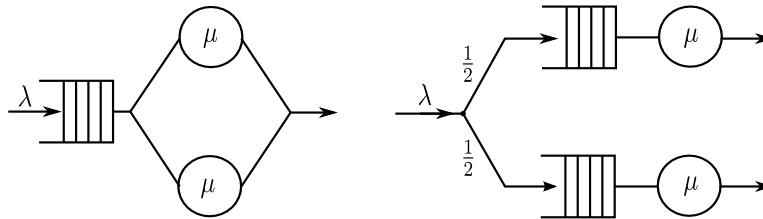


FIGURE 6.36 – Les deux organisations de files étudiées dans l'exercice 29 (premier point). Elles sont parfois appelées respectivement – improprement – à l'américaine (à gauche) et à la française (à droite)

gauche), les clients attendent dans un buffer unique et la discipline de service est *first come, first served*. Dans la file à la française (à droite),

les clients se répartissent aléatoirement dans deux files, de manière équiprobable. On posera pour simplifier les notations dans les calculs :

$$\rho := \frac{\lambda}{\mu}$$

- (a) donner les conditions de stabilité pour les deux organisations
 - (b) calculer pour les deux organisations le nombre moyen de clients présents et le temps moyen de séjour
 - (c) comparer les performances de ces deux organisations en terme de nombre de clients et de temps de séjour
 - (d) calculer et comparer la proportion du temps où au moins un employé est inactif, dans les deux organisations
2. Dans un deuxième temps, on cherche à représenter plus fidèlement l'organisation *à la française*. Dans la réalité, les clients ne se répartissent pas aléatoirement mais choisissent, en arrivant, de se ranger dans la file qui contient le moins de clients²⁰. Ce type d'organisation est représentée à la figure 6.37. Le routage devient donc dynamique : la

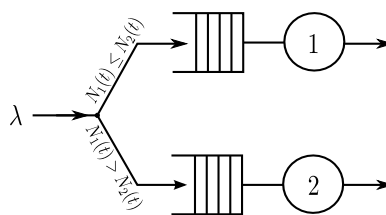


FIGURE 6.37 – Une organisations de file *à la française* avec un routage dynamique

route choisie par un client qui arrive au temps t dépend de l'état du système $(N_1(t), N_2(t))$ à cet instant.

- (a) donner une condition de stabilité pour cette organisation
- (b) donner un modèle de ce système sous la forme d'une chaîne de Markov à temps continu
- (c) écrire les équations d'équilibre qui permettent de calculer la distribution stationnaire de probabilité (sans les résoudre)
- (d) donner les expressions du nombre moyen de clients présents et du temps moyen de séjour des clients, en fonction des composantes (non calculées) de la distribution stationnaire de probabilité

20. un observateur attentif de l'organisation des files *à la française* pourrait critiquer cette représentation simpliste et affirmer – non sans raison – que dans ce type d'organisation, les clients changent de plus de file lorsqu'ils ont l'impression que l'autre file avance plus vite ...

Correction de l'Exercice 29

1. L'organisation "à l'américaine" correspond à une file $M/M/2$ qui est stable pour $\rho < 2$. Pour ce type de file (voir les résultats du cours sur la file $M/M/C$), on obtient :

$$\begin{aligned}\pi_0 &= \frac{2 - \rho}{2 + \rho} \\ R_a &= \frac{\rho^2}{\mu(2 - \rho)^2} \pi_0 + \frac{1}{\mu} = \frac{1}{\mu} \frac{4}{4 - \rho^2} \\ Q_a &= \frac{\rho^3}{(2 - \rho)^2} \pi_0 + \rho = \frac{4\rho}{4 - \rho^2}\end{aligned}$$

On vérifie bien sûr la loi de Little $Q_a = R_a \lambda$ qui nous permet de calculer (par exemple) Q_a à partir de R_a (ou le contraire). L'organisation "à la française" est celle déjà examinée dans l'exercice précédent (trois imprimantes en parallèle, sans spooler d'impression). En effet, le routage probabiliste entraîne des taux d'arrivées $\lambda \cdot \frac{1}{2} = \frac{\lambda}{2}$ dans chacune des deux files en parallèle. En notant Q_1 et Q_2 les nombres moyens de clients respectivement dans les files 1 et 2, et en notant de même X_1 et X_2 les débits moyens respectivement dans les files 1, et 2, on obtient, par conservation des clients :

$$\begin{aligned}Q_f &= Q_1 + Q_2 \\ X &= X_1 + X_2\end{aligned}$$

Les deux files simples sont stables si $\lambda/2 < \mu$, c'est-à-dire si $\rho < 2$ (même condition que dans l'organisation "à l'américaine"). Dans ce cas, on obtient pour les paramètres de performance asymptotiques des deux files simples de type $M/M/1$:

$$\begin{aligned}Q_1 &= Q_2 = \frac{\rho}{2 - \rho} \\ X_1 &= X_2 = \frac{\lambda}{2}\end{aligned}$$

La relation de Little $Q_f = R_f X$ permet alors d'obtenir

$$R_f = \frac{1}{\mu} \frac{2}{2 - \rho}$$

Nous pouvons maintenant répondre aux questions du point 1.

- (a) les deux organisations sont stables avec la même condition, à savoir $\lambda < 2\mu$

- (b) les nombres moyens de clients ont été calculés ci-dessus. On observe en particulier que

$$Q_a = \frac{4\rho}{4 - \rho^2} = Q_f \cdot \frac{2}{2 + \rho}$$

et

$$R_a = \frac{1}{\mu} \frac{4}{4 - \rho^2} = R_f \cdot \frac{2}{2 + \rho}$$

- (c) La file *américaine* est donc plus performante, aussi bien en ce qui concerne le nombre de client dans le système, qu'en ce qui concerne le temps moyen de séjour. Le facteur $\alpha := 1 + \rho/2$, d'amélioration de performance entre les deux types d'organisation, croît linéairement depuis $\alpha = 1$ (pour le cas limite $\rho = 0$ où il n'y a pas de client ou que le temps de service est nul) jusqu'à $\alpha = 2$ (à la limite de saturation $\rho \rightarrow 2$ du système)
- (d) La proportion du temps où un des deux serveur au moins est inoccupé s'écrit :

$$p := \pi_0 + \pi_1$$

Dans le cas de la “file américaine” ($M/M/2$), on a :

$$\pi_1 = \frac{\rho^1}{1!} \pi_0 = \rho \frac{2 - \rho}{2 + \rho}$$

et

$$p_a := \frac{(2 - \rho)(1 + \rho)}{2 + \rho}$$

Dans le cas de la “file française”, un des serveur est inactif si une des deux files est vide. Par facilité, on calcule la probabilité de l'évènement complémentaire :

$$P(\text{aucune file vide}) = P(\text{serveur 1 occupé}) \cdot P(\text{serveur 2 occupé}) = U^2$$

où U est le taux d'utilisation du serveur dans une file $M/M/1$ avec un taux d'arrivée $\lambda/2$ et un taux de service μ . On trouve finalement :

$$p_f := 1 - U^2 = 1 - \frac{\rho^2}{4} = \frac{4 - \rho^2}{4}$$

On a donc

$$p_f = \beta p_a$$

avec

$$\beta := \frac{(2 + \rho)^2}{4(1 + \rho)}$$

La proportion du temps où un serveur est inoccupé est plus grande dans l'organisation "à la française" que dans l'organisation "à l'américaine". Le facteur β d'augmentation croît de manière monotone de $\beta = 1$ (cas limite $\rho = 0$, quand il n'y a pas de client ou que le temps de service est nul) à $\beta = 4/3$ (à la limite de saturation du système)

2. La nouvelle configuration envisagée, avec routage dynamique, de la file "à la française" est représentée à la figure 6.37.
 - (a) La capacité maximale de service (débit moyen maximal de sortie) de ce réseau de deux files est 2μ . Il est atteint lorsque les deux serveurs sont occupés. Or, les deux serveurs sont nécessairement occupés dès qu'il y a plus d'un client dans le système. La file est donc instable si λ (taux global d'arrivée) est supérieur ou égal à 2μ et stable pour $\lambda < 2\mu$.
 - (b) L'état du système à un instant donné est un couple de valeurs $n(t) = (n_1(t), n_2(t)) \in \mathbb{N}^2$ où $n_1(t)$ et $n_2(t)$ désignent respectivement le nombre de clients dans la file 1 et dans la file 2, à l'instant $t > 0$. En effet, les probabilités de transition dépendent – à cause du routage dynamique – non seulement du nombre total de clients dans le réseau, mais également de la répartition de ces clients entre les deux files. Le processus stochastique reste néanmoins une chaîne à temps continu puisque le nouvel espace d'état $E := \mathbb{N}^2$ est lui aussi dénombrable. Par ailleurs, les probabilités de transition depuis un état $n = (n_1, n_2)$ vers un autre état ne dépendent que de cet état et de l'état de destination, pas du temps t , ni du chemin parcouru pour arriver à l'état $n = (n_1, n_2)$. Le processus considéré dans son ensemble (i.e. le réseau de deux files simples avec un routage dynamique) reste donc bien une chaîne de Markov homogène à temps continu. Nous pouvons dès lors spécifier le modèle en donnant tous les taux de transition. Une représentation commode du modèle, c'est-à-dire du graphe associé à la CMTC, est celle donnée à la figure 6.38.
 - (c) La CMTC de la figure 6.38 est irréductible. Sous l'hypothèse de stabilité $\lambda < 2\mu$, la distribution stationnaire existe, est unique, non nulle et égale à la distribution limite de probabilité. On peut donc définir les probabilités stationnaires de la manière suivante :

$$\pi_{m,n} := \lim_{T \rightarrow \infty} P(N_1(t) = m, N_2(t) = n), \forall m, n \in \mathbb{N}$$

Les équations d'équilibre en chaque noeud sont des récurrences qui lient, dans le cas général, la probabilité stationnaire d'un état avec les probabilités d'état de ses quatre voisins. Ces équations sont difficiles à écrire, et bien sûr bien davantage encore à résoudre.

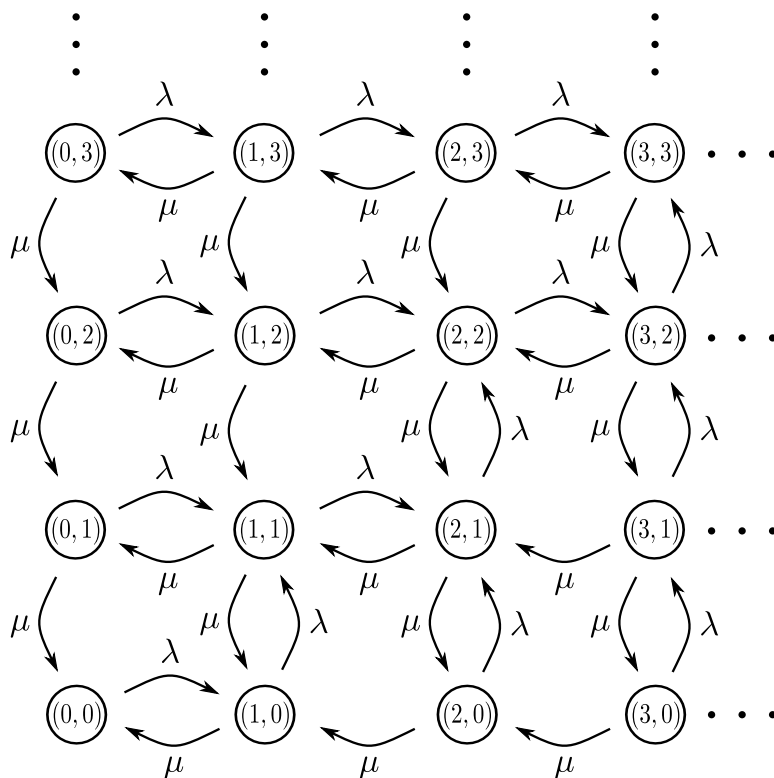


FIGURE 6.38 – Le graphe de la CMTC associée au réseau de deux files avec routage dynamique

On pourra toutefois se limiter en première approximation à la résolution d'un problème approché obtenu en limitant l'état à $\bar{E} := \{1, \dots, N\} \times \{1, \dots, N\}$ car par hypothèse de stabilité, les probabilités d'état sont décroissantes avec N . Les équations d'équilibre – quelles soient obtenues en exprimant l'équilibre en chaque noeud ou par la méthode des coupes – devront être complétées par la condition de normalisation :

$$\sum_{m,n \geq 0} \pi_{m,n} = 1$$

(d) Le nombre moyen de client dans le système s'écrit

$$\begin{aligned} Q &:= \sum_{m,n \geq 0} (m+n) \pi_{m,n} \\ &= \sum_{m,n \geq 0} m \pi_{m,n} + \sum_{m,n \geq 0} n \pi_{m,n} \\ &= \sum_{m \geq 0} m \sum_{n \geq 0} \pi_{m,n} + \sum_{n \geq 0} n \sum_{m \geq 0} \pi_{m,n} \end{aligned}$$

La CMTC de la figure 6.38 n'est pas symétrique – à cause du routage dynamique²¹ – et en général $\pi_{m,n} \neq \pi_{n,m}$. Une fois calculé Q , le temps moyen de séjour des clients dans le système s'obtient par la formule de Little :

$$R = \frac{Q}{\bar{X}} = \frac{Q}{\lambda}$$

Exercice 30 (La file M/M/C/C) On considère un système de type *commutateur téléphonique*, composé de C relais permettant de rediriger les appels vers la bonne destination. Chaque relais ne traite qu'un seul appel à la fois, avec un temps de service aléatoire de distribution exponentielle $\text{Exp}(\mu)$. Les appels arrivent selon un Processus de Poisson d'intensité λ , et sont ensuite dirigés vers un relais libre. Si un nouvel appel arrive alors que tous les relais sont occupés, il est perdu.

1. Donner l'expression (en fonction de λ , μ et n) des taux d'arrivées $\lambda(n)$ et des taux de départ $\mu(n)$ des appels dans le commutateur lorsque n appels sont en cours de traitement. Modéliser ce problème par une chaîne de Markov à temps continu (CMTC). Donner le graphe associé, les taux de transition et le générateur infinitésimal de la CMTC.
2. Quelle est la condition de stabilité de cette file ? Le système est-il ergodique ?
3. Calculer la distribution stationnaire de probabilité
4. Quel est le taux de perte (proportion d'appels perdus) ?
5. Donner l'expression du nombre moyen d'appels en cours de traitement, du temps moyen de séjour des appels dans le commutateur et du débit moyen de sortie des appels

Correction de l'Exercice 30

1. Le système considéré est en fait une file de type $M/M/C/C$ avec C serveurs qui opèrent en parallèle et une capacité limitée à C clients. Il n'y a donc pas de buffer dans cette file, représentée schématiquement à la figure 6.39. L'état du système est décrit par le nombre total de clients dans le système, $N(t) \in E := \{0, 1, \dots, C\}$. Lorsque $N(t) = n \leq C$, tous les clients sont en service et chacun est susceptible de quitter son serveur avec un taux exponentiel. Le taux de départ $\mu(n)$ – qui est aussi le taux de transition de l'état n vers l'état $n - 1$ – vaut donc

21. pour rétablir la symétrie, il faudrait que lorsque les deux files sont de même taille, le client entrant choisisse de manière équiprobable l'une ou l'autre file

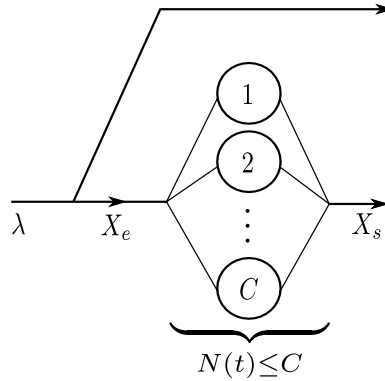


FIGURE 6.39 – Représentation schématique d’une file $M/M/C/C$ utilisée pour modéliser le commutateur téléphonique de l’exercice 30

$\mu(n) = n\mu$. Lorsque $N(t) = C$, les C serveurs sont occupés et le taux de départ vaut $\mu(n) = C\mu$. Le taux d’arrivée des clients dans le système – qui est aussi taux de transition de n à $n + 1$ – est constant et vaut $\lambda(n) = \lambda$, tant que $N(t) = n < N$. Pour $N(t) = C$, le taux d’arrivée est nul. On a donc :

$$\mu(n) = \mu n, \forall n \in \{0, 1, \dots, C\}$$

et

$$\lambda(n) = \begin{cases} \lambda & \text{si } n < C \\ 0 & \text{si } n = C \end{cases}$$

La CMTC qui correspond à cette file $M/M/C/C$ est donnée à la figure 6.40. Le *générateur infinitésimal* (infini) de la file $M/M/C$ s’écrit :

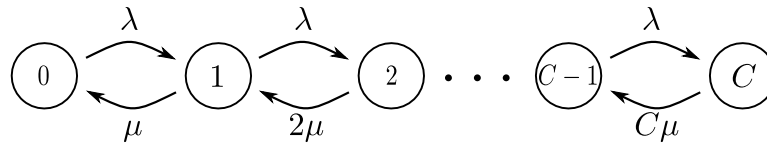


FIGURE 6.40 – La chaîne de Markov à temps continu associée à la file $M/M/C/C$

$$\mathbf{A} = \begin{pmatrix} -\lambda & \lambda & 0 & \cdots & \cdots & 0 \\ \mu & -\lambda - \mu & \lambda & 0 & \cdots & 0 \\ 0 & 2\mu & -\lambda - 2\mu & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & (C-1)\mu & -\lambda - (C-1)\mu & \lambda \\ 0 & \cdots & 0 & 0 & C\mu & -\lambda - C\mu \end{pmatrix}$$

2. La CMTC qui décrit la file $M/M/C/C$ est finie et irréductible. Tous les états sont donc nécessairement récurrents non nuls. La distribution stationnaire de probabilité existe, est unique et égale à la distribution limite. La file $M/M/C/C$ est donc nécessairement stable et ergodique.
3. la distribution stationnaire s'obtient à l'aide de la méthode des coupes. On obtient les relations suivantes :

$$\lambda \pi_{n-1} = n \mu \pi_n, \forall n \in \{1, \dots, C\}$$

En posant,

$$\rho := \frac{\lambda}{\mu}$$

on obtient directement :

$$\pi_n = \frac{\rho^n}{n!} \pi_0$$

La condition de normalisation donne :

$$\sum_{n=0}^C \pi_n = \pi_0 \sum_{n=0}^C \frac{\rho^n}{n!} = 1$$

c'est-à-dire

$$\pi_0 = \frac{1}{\sum_{n=0}^C \frac{\rho^n}{n!}}$$

4. En régime permanent, lorsqu'un appel arrive, il est perdu si il y a déjà C clients dans le système. La proportion d'appels perdus est donc la probabilité limite d'être dans l'état C , soit :

$$\begin{aligned} \pi_C &= \frac{\rho^C}{C!} \pi_0 \\ &= \frac{\frac{\rho^C}{C!}}{\sum_{n=0}^C \frac{\rho^n}{n!}} \\ &= \frac{1}{1 + C! \sum_{n=0}^{C-1} \frac{\rho^{n-C}}{n!}} \end{aligned}$$

5. Les débit moyen de sortie X_s et le débit moyen d'entrée X_e (dans le commutateur) sont égaux entre eux, mais ne sont pas égaux au taux d'arrivée λ des appels. En effet, une partie des appels sont perdus et

n'entrent pas dans le commutateur. On a :

$$\begin{aligned}
 X_s &= \sum_{n=0}^C n\mu\pi_n \\
 &= \mu\rho\pi_0 \sum_{n=1}^C \frac{\rho^{n-1}}{(n-1)!} \\
 &= \mu\rho\pi_0 \sum_{n=0}^{C-1} \frac{\rho^n}{n!} \\
 &= \mu\rho \left(1 - \frac{\rho^C\pi_0}{C!}\right) \\
 &= \lambda \left(1 - \frac{\rho^C\pi_0}{C!}\right) \\
 &= \lambda(1 - \pi_C)
 \end{aligned}$$

On retrouve bien sûr le résultat attendu : le débit moyen dans le commutateur est égal au taux d'arrivée λ diminué du taux d'appels perdus $\lambda\pi_C$. Le nombre moyen de d'appels dans le commutateur s'écrit :

$$\begin{aligned}
 Q &= \sum_{n=0}^C n\pi_n \\
 &= \rho\pi_0 \sum_{n=1}^C \frac{\rho^{n-1}}{(n-1)!} \\
 &= \rho\pi_0 \sum_{n=0}^{C-1} \frac{\rho^n}{n!} \\
 &= \rho \left(1 - \frac{\rho^C\pi_0}{C!}\right) \\
 &= \rho(1 - \pi_C)
 \end{aligned}$$

Le temps moyen de séjour des clients effectivement admis dans le commutateur est simplement le temps moyen de service puisqu'il n'y a pas de buffer :

$$R = \frac{1}{\mu}$$

On vérifie aisément la loi de Little $Q = RX_s$.

Exercice 31 (discipline de service dynamique) Dans un magasin, une caisse fonctionne en permanence. Parfois la file d'attente à cette caisse est trop longue. Une deuxième caisse est alors ouverte (partageant la même file) si le nombre de clients attendant dans la file (i.e. le buffer) est au moins égal

à K . Le premier client arrivé passe alors immédiatement à cette deuxième caisse (le buffer devient donc de taille $K - 1$ à cet instant). Quand le nombre de clients dans la file (i.e. le buffer) devient strictement inférieur à $K - 1$, la première des deux caisses qui termine le service en cours est fermée (et ne prend donc plus de client, jusqu'à ce qu'elle soit à nouveau ouverte). On suppose que le taux de service de la deuxième caisse est le même que celui de la première.

1. Modéliser cette file par une CMTC et calculer la distribution stationnaire de probabilité de cette CMTC
2. Cette file est-elle stable et ergodique ?
3. Calculer le nombre moyen de clients dans le système en régime permanent
4. Calculer, pour le cas $\lambda = \mu$, le seuil d'ouverture K de façon à ce que le nombre moyen de clients dans la file (buffer) soit – en régime stationnaire – inférieur ou égal à 4.
5. Calculer dans le cas général le taux d'utilisation²² de la caisse N°2.

Correction de l'Exercice 31

1. L'état du système est décrit par le nombre total de clients dans le système, $N(t) \in \mathbb{N}$. Le système considéré est une file à deux serveurs avec une discipline de service *dynamique*. En effet, les clients ne sont dirigés vers le serveur 2 que si $N(t) \geq K$ (condition d'ouverture de la caisse N°2). Si cette condition est satisfaite, la discipline de service est FCFS (*first come, first served*) : dès qu'un serveur (i.e. une caisse) se libère, le premier client dans la file (i.e. le premier arrivé) est servi. Cette file à discipline de service dynamique est représentée schématiquement à la figure 6.41. Lorsque deux caisses sont ouvertes (i.e. $N(t) \geq K$), le taux de service est 2μ . Sinon, il vaut μ . Le taux d'arrivée des clients dans le système vaut $\lambda(n) = \lambda$, quel que soit le nombre de client dans le système. La CMTC qui correspond à cette file est donnée à la figure 6.42. la distribution stationnaire s'obtient à l'aide de la méthode des coupes. On obtient les relations suivantes :

$$\begin{aligned} \lambda\pi_{n-1} &= \mu\pi_n & \text{pour } 0 \leq n \leq K \\ \lambda\pi_{n-1} &= 2\mu\pi_n & \text{pour } n \geq K + 1 \end{aligned}$$

En posant,

$$\rho := \frac{\lambda}{\mu}$$

22. Ce taux reflète par exemple la proportion du temps où l'employé de la caisse N°2 sera disponible pour d'autres tâches dans le magasin

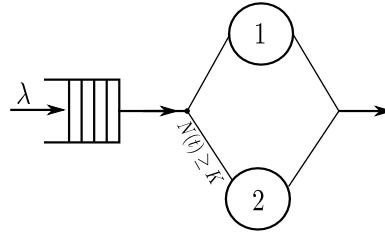


FIGURE 6.41 – Représentation schématique de la file à deux serveurs et discipline de service dynamique étudiée dans l'exercice 31

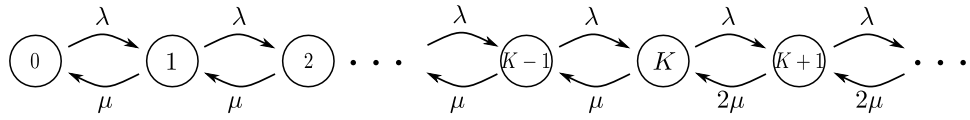


FIGURE 6.42 – CMTTC associée à la file à deux serveurs avec discipline de service dynamique étudiée à l'exercice 31

on obtient :

$$\pi_n = \begin{cases} \rho^n \pi_0 & \text{pour } 0 \leq n \leq K \\ \frac{\rho^n}{2^{n-K}} \pi_0 & \text{pour } n \geq K+1 \end{cases}$$

La condition de normalisation donne (dans le cas $\rho \neq 1$) :

$$\begin{aligned} \sum_{n \geq 0} \pi_n &= \sum_{n=0}^{K-1} \pi_n + \sum_{n \geq K} \pi_n \\ &= \left[\frac{1 - \rho^K}{1 - \rho} + \rho^K \underbrace{\sum_{n \geq K} \left(\frac{\rho}{2}\right)^{n-K}}_{\Sigma_\infty} \right] \pi_0 \\ &= \left[\frac{1 - \rho^K}{1 - \rho} + \frac{2\rho^K}{2 - \rho} \right] \pi_0 \\ &= \left[\frac{2 - \rho - \rho^K}{(1 - \rho)(2 - \rho)} \right] \pi_0 \\ &= 1 \end{aligned}$$

c'est-à-dire

$$\pi_0 = \frac{(1 - \rho)(2 - \rho)}{2 - \rho - \rho^K}$$

Il faut noter que la somme Σ_∞ ne converge que si $\rho < 2$, c'est-à-dire

si $\lambda < 2\mu$. Par ailleurs, dans le cas $\rho = 1$, on obtient :

$$\sum_{n=0}^{K-1} \pi_n = K$$

et

$$\pi_0 = \frac{1}{K+2}$$

2. La CMTC de la figure 6.42 irréductible. Elle est stable et ergodique pour $\lambda < 2\mu$. Si cette condition est satisfaite, tous les états sont donc nécessairement récurrents non nuls. La distribution stationnaire de probabilité existe, est unique et égale à la distribution limite.
3. En régime permanent, dans le cas général où $\rho \neq 1$, le nombre moyen de clients dans le système s'écrit :

$$\begin{aligned}
 Q &= \sum_{n \geq 0} n \pi_n \\
 &= \sum_{n=0}^{K-1} n \pi_n + \sum_{n \geq K} n \pi_n \\
 &= \left[\sum_{n=0}^{K-1} n \rho^n + \rho^K \sum_{n \geq K} n \left(\frac{\rho}{2} \right)^{n-K} \right] \pi_0 \\
 &= \left[\sum_{n=0}^{K-1} \rho \frac{d}{dx} \left(\frac{1-x^K}{1-x} \right) \Big|_{x=\rho} + \rho^K \sum_{n \geq K} n \left(\frac{\rho}{2} \right)^{n-K} \right] \pi_0 \\
 &= \left[\frac{\rho ((K-1)\rho^K - K\rho^{K-1} + 1)}{(1-\rho)^2} + \rho^K \left(\sum_{n \geq K} (n-K) \left(\frac{\rho}{2} \right)^{n-K} + K \sum_{n \geq K} \left(\frac{\rho}{2} \right)^{n-K} \right) \right] \pi_0 \\
 &= \left[\frac{\rho ((K-1)\rho^K - K\rho^{K-1} + 1)}{(1-\rho)^2} + \rho^K \left(\frac{2\rho}{(2-\rho)^2} + \frac{2K}{2-\rho} \right) \right] \pi_0 \\
 &= \frac{\rho^K (-(K-1)\rho^3 + K(3\rho^2 + 4\rho + 2)) + \rho(\rho^2 - 4\rho + 2)}{(1-\rho)(2-\rho)(2-\rho-\rho^K)}
 \end{aligned}$$

4. Dans le cas particulier $\lambda = \mu$, c'est-à-dire pour $\rho = 1$, on trouve

$$\begin{aligned}
 Q &= \sum_{n \geq 0} n \pi_n \\
 &= \sum_{n=0}^{K-1} n \pi_n + \sum_{n \geq K} n \pi_n \\
 &= \left[\frac{K(K-1)}{2} + \sum_{n \geq K} n \left(\frac{1}{2} \right)^{n-K} \right] \pi_0 \\
 &= \left[\frac{K(K-1)}{2} + \sum_{n \geq K} (n-K) \left(\frac{1}{2} \right)^{n-K} + K \sum_{n \geq K} \left(\frac{1}{2} \right)^{n-K} \right] \pi_0 \\
 &= \left[\frac{K(K-1)}{2} + 2 + 2K \right] \pi_0 \\
 &= \frac{K^2 + 3K + 4}{K + 2}
 \end{aligned}$$

On a $Q \leq 5$ si $K \leq 1 + \sqrt{7} \simeq 3.646...$ On choisira donc un seuil de $K = 3$ clients dans la file pour ouvrir la deuxième caisse.

5. Le taux d'utilisation de la deuxième caisse est

$$\begin{aligned}
 U_2 &= \sum_{n \geq K+1} \pi_n \\
 &= 1 - \sum_{n \leq K} \pi_n \\
 &= 1 - \pi_0 \sum_{n \leq K} \rho^n \\
 &= \frac{2(2 - \rho) - \rho^K(1 + \rho)^2}{2 - \rho - \rho^K}
 \end{aligned}$$

Exercice 32 (File avec plantage) On considère une file d'attente $M/M/1$ (de paramètres λ, μ) *modifiée*, où le serveur peut en “plantant” détruire tous les clients de la file. Après le “plantage”, le serveur reprend son service normalement. On suppose que le temps de bon fonctionnement du serveur suit une loi $Exp(\gamma)$. On suppose $\gamma > \mu$ et $\lambda < \mu$

1. Modéliser cette file par une chaîne de Markov à temps continu (CMTC) en choisissant comme état le nombre de clients dans le système au temps $t > 0$.
2. A l'aide du théorème des coupes, calculer le nombre moyen de clients dans la file *en régime stationnaire* en fonction de π_0 (on note π le vecteur des probabilités stationnaires et π_0 sa première composante)
3. On cherche à calculer le régime stationnaire

- En utilisant les équations d'équilibre, donner une relation liant π_n , π_{n+1} et π_{n-1}
- Montrer qu'il existe $r_1 > 1$ et $r_2 < 1$ tels que $\pi_n = Ar_1^n + Br_2^n$ (on pourra introduire $\rho = \frac{\lambda}{\mu}$ et $\alpha = \frac{\gamma}{\mu}$)
- La file considérée est-elle stable ? Pourquoi ? Que peut-on en conclure sur la valeur des paramètres A ou B ?
- Calculer π_n ($n \geq 0$), en utilisant la condition de normalisation du vecteur des probabilités stationnaires

Correction de l'Exercice 32

- L'état du système au temps $t > 0$ est le nombre total de clients dans le système, $N(t) \in \mathbb{N}$. La CMTC qui correspond à cette file avec "plantage" est donnée à la figure 6.43. Le *générateur infinitésimal* (infini)

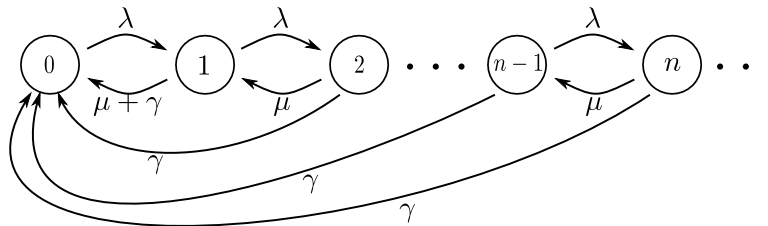


FIGURE 6.43 – La CMTC associée à la file avec "plantage"

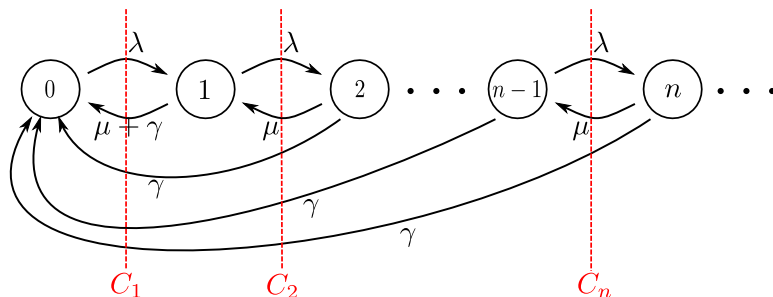
de cette file s'écrit :

$$\mathbf{A} = \begin{pmatrix} -\lambda & \lambda & 0 & \cdots & \cdots & \cdots \\ \mu + \gamma & -(\lambda + \mu + \gamma) & \lambda & 0 & \cdots & \cdots \\ \gamma & \mu & -(\lambda + \mu + \gamma) & \lambda & 0 & \cdots \\ \gamma & 0 & \mu & -(\lambda + \mu + \gamma) & \lambda & \ddots \\ \vdots & \vdots & \ddots & \ddots & \ddots & \ddots \end{pmatrix}$$

- A l'aide du théorème des coupes, en utilisant les coupes illustrées à la figure 6.44, on obtient les équations :

$$\begin{aligned} \lambda \pi_0 &= \mu \pi_1 + \gamma (\pi_1 + \pi_2 + \dots) \\ \lambda \pi_1 &= \mu \pi_2 + \gamma (\pi_2 + \pi_3 + \dots) \\ &\vdots \\ \lambda \pi_n &= \mu \pi_{n+1} + \gamma (\pi_{n+1} + \pi_{n+2} + \dots) \end{aligned}$$

En sommant ces équations, on obtient :

FIGURE 6.44 – Les coupes utilisées pour le calcul de Q dans l'exercice 32

$$\lambda \sum_{n \geq 0} \pi_n = \mu \sum_{n \geq 1} \pi_n + \gamma \sum_{n \geq 1} (n \pi_n)$$

Or

$$\begin{aligned} \sum_{n \geq 0} \pi_n &= 1 \\ \sum_{n \geq 1} \pi_n &= 1 - \pi_0 \\ \sum_{n \geq 1} (n \pi_n) &=: Q \end{aligned}$$

où Q désigne le nombre moyen de client dans la file en régime stationnaire. On obtient finalement :

$$Q = \frac{\lambda - \mu(1 - \pi_0)}{\gamma}$$

3. (a) Les équations d'équilibre en chacun des noeuds s'écrivent pour $n \geq 1$ (voir figure 6.43) :

$$\lambda \pi_{n-1} + \mu \pi_{n+1} = (\lambda + \mu + \gamma) \pi_n \quad (6.14)$$

avec (équilibre au noeud 0) :

$$\lambda \pi_0 = \mu \pi_1 + \gamma (\pi_1 + \pi_2 + \dots) = \mu \pi_1 + \gamma (1 - \pi_0) \quad (6.15)$$

et la condition de normalisation usuelle

$$\sum_{n \geq 0} \pi_n = 1 \quad (6.16)$$

La distribution stationnaire de probabilité $\{\pi_n\}_{n \geq 0}$ est donc l'unique solution de la récurrence linéaire d'ordre 2 à coefficients constants (6.14) qui satisfasse simultanément les conditions (6.15) et (6.16).

- (b) L'équation caractéristique associée à la récurrence (6.14) s'écrit

$$r^2 - (\rho + \alpha + 1)r + \rho = 0 \quad (6.17)$$

en posant $\rho := \frac{\lambda}{\mu} < 1$ et $\alpha := \frac{\gamma}{\mu} > 1$. Le discriminant de cette équation du second degré est

$$\Delta^2 = (\rho + \alpha + 1)^2 - 4\rho = (\rho - 1)^2 + \alpha^2 + 2\alpha + 2\alpha\rho > 0$$

Les deux solutions r_1 et r_2 de (6.17) sont donc réelles et distinctes. Leur produit satisfait

$$r_1 r_2 = \rho < 1$$

et elles s'écrivent

$$r_{1,2} = \frac{(\rho + \alpha + 1) \pm \Delta}{2}$$

où Δ désigne la solution positive de l'équation $\Delta^2 = (\rho + \alpha + 1)^2 - 4\rho$. Les solutions r_1 et r_2 de (6.17) sont donc toutes les deux positives avec $0 < r_1 < 1$ et $r_2 > 1$. La solution générale de la récurrence (6.14) s'écrit

$$\pi_n = Ar_1^n + Br_2^n \text{ pour } n \geq 1$$

- (c) Comme $\rho < 1$, même en l'absence de plantage du serveur, la file est stable. Les éventuels plantages du serveur ne peuvent que diminuer le nombre moyen de clients dans le système et ne peuvent donc remettre en question la stabilité de la file. En conséquence, les probabilités π_n sont bornées avec n et $B = 0$.
- (d) La condition de normalisation s'écrit donc

$$\begin{aligned} \sum_{n \geq 0} \pi_n &= \pi_0 + \sum_{n \geq 1} \pi_n \\ &= \pi_0 + A \sum_{n \geq 1} r_1^n \\ &= \pi_0 + A \frac{r_1}{1 - r_1} \\ &= 1 \end{aligned}$$

La condition (6.15) d'équilibre en 0 s'écrit

$$\begin{aligned} \rho\pi_0 &= \pi_1 + \alpha(1 - \pi_0) \\ &= Ar_1 + \alpha(1 - \pi_0) \end{aligned}$$

A l'aide des deux dernières équations, on trouve

$$\begin{aligned} \pi_0 &= \frac{1 - r_1 + \alpha}{1 - r_1 + \alpha + \rho} \\ A &= \frac{(1 - r_1)\rho}{(1 - r_1 + \alpha + \rho)r_1} \end{aligned}$$

Exercice 33 (File avec impatience) On considère une *file avec impatience*, i.e. une file dans laquelle un client qui arrive et voit n clients (dans le système) devant lui a une probabilité p_n de s'insérer dans la file (et une probabilité $1 - p_n$ de partir). Dans cet exercice, on considèrera que les clients arrivent selon un processus de Poisson de paramètre λ (avant de décider si ils entrent dans le système). Le temps de service d'un client est une variable aléatoire avec une distribution exponentielle de paramètre μ . La probabilité d'entrée effective dans le systèmes pour un nouveau client arrivant vaut

$$p_n := \frac{1}{1+n}, \forall n \geq 0$$

où n désigne le nombre de clients déjà dans le système lorsque le nouveau client arrive.

1. Modéliser cette file par une chaîne de Markov à temps continu
2. Calculer la distribution stationnaire de probabilité associée à cette chaîne
3. Calculer le taux moyen stationnaire d'entrée (X_e) et le taux moyen stationnaire de sortie (X_s). A partir de ces deux résultats, discuter la stabilité de la file (éventuellement, en fonction des valeurs des paramètres λ et μ)
4. calculer le nombre moyen de clients dans le système
5. calculer le temps de séjour moyen d'un client dans le système (lorsqu'il y entre effectivement)
6. la distribution stationnaire de cette file avec impatience est-elle la même que pour une file $M/M/\infty$? Les paramètres de performance qui en découlent sont-ils les mêmes que pour une file $M/M/\infty$?

Correction de l'Exercice 33

1. L'état du système au temps $t > 0$ est le nombre total de clients dans le système, $N(t) \in \mathbb{N}$. La CMTC qui correspond à cette file avec impatience est donnée à la figure 6.45. En effet, lorsqu'il y a n clients dans

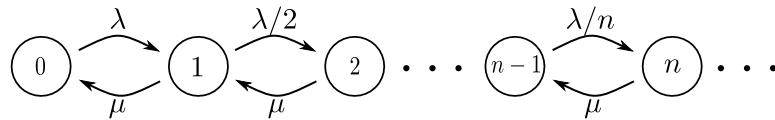


FIGURE 6.45 – CMTC associée à la file avec impatience de l'exercice 33

le système, le processus d'arrivées se subdivise en deux processus de

Poisson de taux respectifs λp_n (clients qui entrent effectivement dans le système) et $\lambda(1-p_n)$ (clients impatientes qui quittent immédiatement le système). Le *générateur infinitésimal* (infini) de cette CMTC s'écrit :

$$\mathbf{A} = \begin{pmatrix} -\lambda & \lambda & 0 & \cdots & \cdots & \cdots \\ \mu & -(\lambda + \mu) & \lambda & 0 & \cdots & \cdots \\ 0 & \mu & -(\frac{\lambda}{2} + \mu) & \frac{\lambda}{2} & 0 & \cdots \\ 0 & 0 & \mu & -(\frac{\lambda}{3} + \mu) & \frac{\lambda}{3} & \ddots \\ \vdots & \vdots & \ddots & \ddots & \ddots & \ddots \end{pmatrix}$$

2. A l'aide de la méthode des coupes, on obtient les équations :

$$\begin{aligned} \rho \pi_0 &= \pi_1 \\ \frac{\rho}{2} \pi_1 &= \pi_2 \\ &\vdots \\ \frac{\rho}{n} \pi_{n-1} &= \pi_n \\ &\vdots \end{aligned}$$

avec $\rho := \lambda/\mu$. En résolvant de manière itérative, il vient :

$$\pi_n = \frac{\rho^n}{n!} \pi_0, \text{ pour } n \geq 0$$

La condition de normalisation s'écrit

$$\sum_{n \geq 0} \pi_n = \left(\sum_{n \geq 0} \frac{\rho^n}{n!} \right) \pi_0 = e^\rho \pi_0 = 1$$

c'est-à-dire

$$\pi_0 = e^{-\rho}$$

Les probabilités stationnaires sont donc distribuées selon une loi de Poisson de paramètre ρ :

$$\pi_n = \frac{\rho^n e^{-\rho}}{n!}, \text{ pour } n \geq 0$$

3. Le taux moyen stationnaire d'entrée s'écrit :

$$\begin{aligned}
 X_e &= \sum_{n \geq 0} \frac{\lambda}{n} \pi_n \\
 &= \lambda \pi_0 \sum_{n \geq 0} \frac{\rho^n}{(n+1)!} \\
 &= \frac{\lambda}{\rho} \pi_0 \sum_{n \geq 1} \frac{\rho^n}{n!} \\
 &= \mu \pi_0 (e^\rho - 1) \\
 &= \mu e^{-\rho} (e^\rho - 1) \\
 &= \mu (1 - e^{-\rho})
 \end{aligned}$$

Le taux moyen stationnaire de sortie s'écrit :

$$\begin{aligned}
 X_s &= \sum_{n \geq 0} \mu \pi_n \\
 &= \mu (1 - \pi_0) \\
 &= \mu (1 - e^{-\rho})
 \end{aligned}$$

On trouve donc $X_e = X_s = \mu (1 - e^{-\rho})$, quels que soient les paramètres λ et μ . La file est inconditionnellement stable. L'impatience des clients revient à diminuer le taux effectif d'entrée dans le système de telle sorte que pour N suffisamment grand, on a :

$$n \geq N \Rightarrow \lambda(n) := \lambda p_n = \frac{\lambda}{1+n} < \mu$$

4. Le nombre moyen de clients dans le système s'écrit :

$$\begin{aligned}
 Q &= \sum_{n \geq 0} n \pi_n \\
 &= \pi_0 \sum_{n \geq 1} \frac{\rho^n}{(n-1)!} \\
 &= \rho \pi_0 \sum_{n \geq 1} \frac{\rho^{n-1}}{(n-1)!} \\
 &= \rho e^{-\rho} e^{\rho} = \rho \\
 &= \rho
 \end{aligned}$$

5. Le temps de séjour moyen des clients qui entrent effectivement dans le système se calcule grâce à la loi de Little :

$$R = \frac{Q}{X} = \frac{\rho}{\mu (1 - e^{-\rho})}$$

6. Bien que la distribution stationnaire de cette file avec impatience soit la même que la distribution stationnaire d'une file $M/M/\infty$ (i.e. une distribution de Poisson de paramètre ρ), ses paramètres de performance sont différents. En effet, si le processus d'arrivées des clients est un processus de Poisson de paramètre λ , le taux effectif d'entrée dans le système est pour la file avec impatience

$$X_e = \mu (1 - e^{-\rho}) = \lambda \left(1 - \frac{\rho}{2!} + \frac{\rho^2}{3!} - \dots \right) \neq \lambda \text{ pour } \rho \neq 0$$

Le nombre moyen de clients présents dans le système est par contre le même pour les deux types de files. Il découle de la loi de Little que les temps de séjour moyen des clients admis dans les deux types de file ne peuvent être égaux.

Exercice 34 (File simple à priorité) On considère un *commutateur téléphonique à priorités* capable de gérer simultanément jusqu'à C appels. Les appels sont subdivisés en trois classes : les appels de type 1, 2 et 3. Les appels de type r arrivent selon un processus de Poisson de taux λ_r (pour $r \in \{1, 2, 3\}$). Lorsqu'un appel arrive alors qu'il y a (strictement) moins de C appels déjà présents dans le commutateur, il est instantanément traité. Tous les appels nécessitent un temps de traitement aléatoire – indépendant du type d'appel – distribué suivant une loi exponentielle de paramètre μ . Lorsqu'un appel arrive alors qu'il y a déjà C appels en cours de traitement dans le commutateur, cet appel est dans certains cas mis en attente dans un buffer FCFS de capacité limitée et, dans d'autres cas, rejeté. Le buffer est subdivisé en trois zones distinctes de tailles respectives k_r pour $r \in \{1, 2, 3\}$. La zone r ne peut accueillir que des appels de type $l \leq r$. La zone 2, par exemple, est de capacité k_2 et peut accueillir des appels de types 1 et 2. Lorsqu'un appel arrive et que les clients déjà présents dans le buffer remplissent toutes les places disponibles pour ce type d'appel, l'appel arrivant est rejeté. Sinon, l'appel arrivant est placé en dernière position dans le buffer.

1. Représenter schématiquement la file simple qui correspond à ce problème. Est-il nécessaire de mémoriser le type des appels une fois ceux-ci admis dans le commutateur ?
2. Modéliser le commutateur comme une file markovienne monoclasse. Donner la chaîne de Markov à temps continu associée à cette file
3. La file est-elle stable et ergodique ?
4. Calculer la distribution stationnaire de probabilité
5. Quelles sont les probabilités de rejet pour les appels de chacun des trois types ?
6. Quels sont les nombres moyens d'appels satisfaits par unité de temps ? Quel est le débit moyen de sortie tous types d'appels confondus ?

7. Calculer le temps moyen d'attente (dans le buffer) pour chaque type d'appel
8. En déduire le nombre moyen d'appels de chaque type dans le buffer

Correction de l'Exercice 34

1. La file simple considérée est représentée à la figure 6.46. Une fois admis

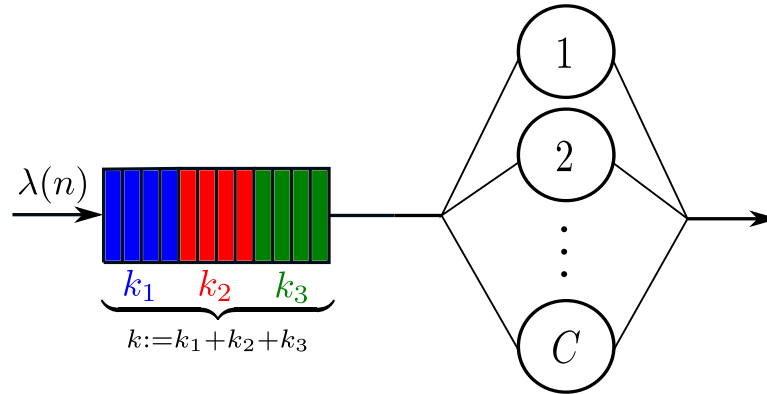


FIGURE 6.46 – Représentation schématique de la file simple à priorité de l'exercice 34. Il s'agit d'une file $M/M/C/C + k$ avec C serveurs et une capacité totale limitée à $C + k$ clients. La discipline de service est FCFS

dans le commutateur, le traitement des appels est identique, quel que soit leur type. Ils sont rangés dans le buffer si tous les serveurs sont occupés. La discipline de service est FCFS pour tous. Leur temps de traitement dans les serveurs est distribué selon une même loi exponentielle de paramètre μ , quel que soit leur type. Il n'est donc pas utile de distinguer les clients dans le commutateur selon leur type. Par contre, le taux effectif d'arrivée $\lambda(n)$ dépend de l'état du système, c'est-à-dire du nombre total $N(t) = n$ de clients présents dans le système au temps $t > 0$. Celui-ci peut être $\lambda(n) := \lambda_1 + \lambda_2 + \lambda_3$ lorsqu'il reste des places libres dans le buffer vert (de capacité k_3), $\lambda(n) := \lambda_1 + \lambda_2$ lorsque le buffer vert est plein mais pas le rouge, $\lambda(n) := \lambda_1$ lorsque les buffers vert et rouge sont pleins mais pas le bleu ou $\lambda(n) := 0$ lorsque les trois buffers sont pleins. On a :

$$\lambda(n) := \begin{cases} \lambda_1 + \lambda_2 + \lambda_3 & \text{si } 0 \leq n < C + k_3 \\ \lambda_1 + \lambda_2 & \text{si } C + k_3 \leq n < C + k_3 + k_2 \\ \lambda_1 & \text{si } C + k_3 + k_2 \leq n < C + k_3 + k_2 + k_1 \\ 0 & \text{si } n \geq C + k_3 + k_2 + k_1 \end{cases}$$

2. L'état du système au temps $t > 0$ est le nombre total de clients dans le système, $N(t) \in \{0, \dots, C + k\}$ avec $k := k_1 + k_2 + k_3$, le nombre maximum de clients dans le buffer. La CMTC qui correspond commutateur est donnée à la figure 6.47. Dans la rangée du haut, il y a moins

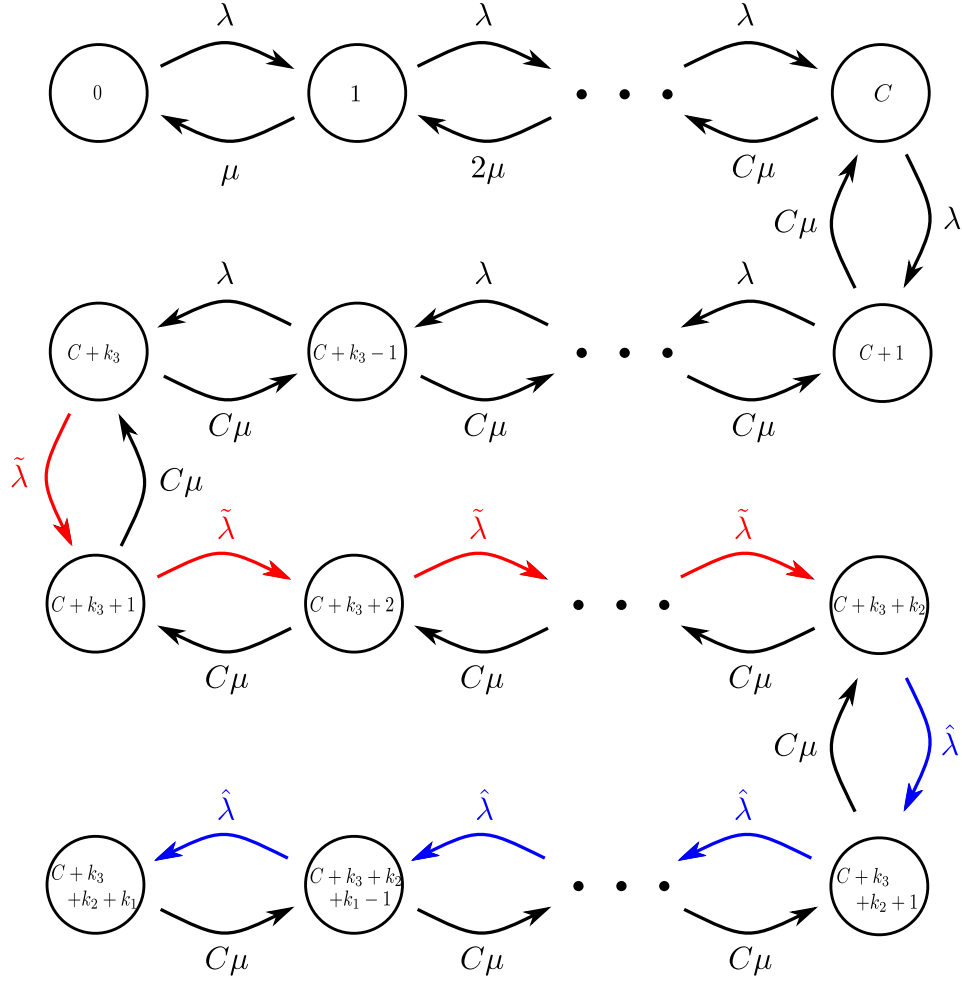


FIGURE 6.47 – La CMTC associée à la file $M/M/C/C + k$ qui représente le commutateur à priorité de l'exercice 34.

de C clients dans le système. Ils sont donc tous servis, dès leur arrivée, par un des serveurs disponibles. Dans la deuxième rangée (en partant du haut), tous les serveurs sont occupés et les clients qui arrivent sont placés dans le buffer, quel que soit leur type. Le taux de service est saturé à la valeur $C\mu$ (tous les serveurs sont occupés). Dans la deuxième rangée (en partant du haut) le buffer qui accepte tous les clients (de capacité k_3) est plein et les clients de type 3 sont donc rejetés à leur

arrivée. Dès lors, le taux effectif d'arrivées passe de $\lambda := \lambda_1 + \lambda_2 + \lambda_3$ à $\tilde{\lambda} := \lambda_1 + \lambda_2$. Dans la rangée du bas, les buffers de capacités respectives k_2 et k_3 sont pleins. Seuls les clients de type 1 sont encore acceptés. Le taux effectif d'arrivées passe donc de $\tilde{\lambda} := \lambda_1 + \lambda_2$ à $\hat{\lambda} := \lambda_1$, jusqu'à ce que le buffer de capacité k_1 soit lui aussi saturé. Tous les clients sont alors rejetés.

3. La CMTC associée à la file est irréductible et finie. Elle est donc nécessairement stable et ergodique. La distribution stationnaire de probabilité est la distribution limite.
4. En appliquant la méthodes des coupes, on obtient pour les C premières coupes les équations :

$$\begin{aligned}\lambda\pi_0 &= \mu\pi_1 \\ \lambda\pi_1 &= 2\mu\pi_2 \\ &\vdots \\ \lambda\pi_{C-1} &= C\mu\pi_C\end{aligned}$$

c'est-à-dire

$$\pi_q = \frac{\rho^q}{q!} \pi_0 \text{ pour } q \in \{1, \dots, C\} \quad (6.18)$$

avec $\rho := \lambda/\mu$. Pour les k_3 coupes suivantes, on obtient les équations :

$$\begin{aligned}\lambda\pi_C &= C\mu\pi_{C+1} \\ \lambda\pi_{C+1} &= C\mu\pi_{C+2} \\ &\vdots \\ \lambda\pi_{C+k_3-1} &= C\mu\pi_{C+k_3}\end{aligned}$$

c'est-à-dire

$$\pi_{C+p} = \frac{\rho^p}{C^p} \pi_C = \frac{\rho^C \rho^p}{C! C^p} \pi_0 \text{ pour } p \in \{1, \dots, k_3\} \quad (6.19)$$

Pour les k_2 coupes suivantes, on obtient les équations :

$$\begin{aligned}\tilde{\lambda}\pi_{C+k_3} &= C\mu\pi_{C+k_3+1} \\ \tilde{\lambda}\pi_{C+k_3+1} &= C\mu\pi_{C+k_3+2} \\ &\vdots \\ \tilde{\lambda}\pi_{C+k_3+k_2-1} &= C\mu\pi_{C+k_3+k_2}\end{aligned}$$

c'est-à-dire

$$\begin{aligned}\pi_{C+k_3+m} &= \frac{\tilde{\rho}^m}{C^m} \pi_{C+k_3} \\ &= \frac{\rho^{C+k_3} \tilde{\rho}^m}{C! C^{k_3+m}} \pi_0 \text{ pour } m \in \{1, \dots, k_2\}\end{aligned} \quad (6.20)$$

avec $\tilde{\rho} := \tilde{\lambda}/\mu = (\lambda_1 + \lambda_2)/\mu$. Enfin, pour les k_1 coupes suivantes, on obtient les équations :

$$\begin{aligned}\hat{\lambda}\pi_{C+k_3+k_2} &= C\mu\pi_{C+k_3+k_2+1} \\ \hat{\lambda}\pi_{C+k_3+k_2+1} &= C\mu\pi_{C+k_3+k_2+2} \\ &\vdots \\ \hat{\lambda}\pi_{C+k_3+k_2+k_1-1} &= C\mu\pi_{C+k_3+k_2+k_1}\end{aligned}$$

c'est-à-dire

$$\begin{aligned}\pi_{C+k_3+k_2+n} &= \frac{\hat{\rho}^n}{C^n} \pi_{C+k_3+k_2} \\ &= \frac{\rho^{C+k_3} \tilde{\rho}^{k_2} \hat{\rho}^n}{C!C^{k_3+k_2+n}} \pi_0 \text{ pour } n \in \{1, \dots, k_1\}\end{aligned} \quad (6.21)$$

avec $\hat{\rho} := \hat{\lambda}/\mu = \lambda_1/\mu$. On obtient comme précédemment la valeur de π_0 par la condition de normalisation

$$\begin{aligned}\sum_{l=0}^{C+k_3+k_2+k_1} \pi_l &= \sum_{q=0}^C \pi_q + \sum_{p=1}^{k_3} \pi_{C+p} + \sum_{m=1}^{k_2} \pi_{C+k_3+m} \\ &\quad + \sum_{n=1}^{k_1} \pi_{C+k_3+k_2+n} = 1\end{aligned}$$

En substituant les solutions (6.18), (6.19), (6.20) et (6.21) dans la condition de normalisation, on obtient la valeur de π_0 .

5. La probabilité de rejet pour les clients de type 1 est

$$IP_1 = \pi_{C+k_3+k_2+k_1} = \frac{\rho^{C+k_3} \tilde{\rho}^{k_2} \hat{\rho}^{k_1}}{C!C^{k_3+k_2+k_1}} \pi_0$$

La probabilité de rejet pour les clients de type 2 est

$$IP_2 = IP_1 + \sum_{n=0}^{k_1-1} \pi_{C+k_3+k_2+n}$$

où les valeurs de $\pi_{C+k_3+k_2+n}$ sont données par les équations (6.21). La probabilité de rejet pour les clients de type 3 est

$$IP_3 = IP_2 + \sum_{m=0}^{k_2-1} \pi_{C+k_3+m}$$

où les valeurs de π_{C+k_3+m} sont données par les équations (6.20).

6. Le nombre moyen d'appels satisfaits par unité de temps, toutes catégories confondues, est le débit moyen d'entrée X_e . La file étant finie, elle est nécessairement stable et $X_s = X_e$. Dans cet exemple, le débit moyen de sortie est plus facile à calculer. Il s'écrit :

$$\begin{aligned}
 X_s &= \sum_{q=0}^C q\mu\pi_q \\
 &= \mu\pi_0 \sum_{q=1}^C \frac{\rho^q}{(q-1)!} \\
 &= \lambda\pi_0 \sum_{q=0}^{C-1} \frac{\rho^q}{q!}
 \end{aligned}$$

7. Les temps moyen d'attente dans le buffer sont différents pour chaque type d'appel. Les clients de type 3 ne peuvent se trouver que dans la première zone (verte) du buffer. Un client de type 3 qui se trouve à la place $p \in \{1, \dots, k_3\}$ (en comptant depuis la sortie) du buffer devra attendre que p clients en cours de service aient terminé leur service. Le temps restant pour terminer le service de chaque client dans un serveur est distribué selon une exponentielle de paramètre μ (temps de vie résiduel). Comme il y a C serveurs occupés en parallèle (car le nombre de clients dans le buffer est plus grand que 1), le temps nécessaire pour que les services de p clients soient terminés est $p/(\mu C)$. Par ailleurs, quand il y a moins de $C + k_3$ clients dans le système, un tiers des clients qui y sont admis (en moyenne) sont de type 3. Le temps moyen d'attente des clients de type 3 dans le buffer (hors service) est donc :

$$\begin{aligned}
 R_3^w &= \frac{1}{3} \sum_{p=0}^{k_3-1} \pi_{C+p} \frac{p}{\mu C} \\
 &= \frac{1}{3\mu C} \sum_{p=0}^{k_3-1} p\pi_{C+p}
 \end{aligned}$$

où les valeurs de π_{C+p} sont données par les équations (6.19). Les autres clients de type 3 qui arrivent sont soit rejetés, soit admis directement dans un serveur. Dans les deux cas, ils passent un temps nul dans le buffer (qui ne contribue donc pas à la moyenne). Les clients de type 2, lorsqu'ils sont dans le buffer, peuvent être dans la première ou la deuxième zone de celui-ci. Ils constituent (en moyenne) un tiers des clients qui arrivent dans la zone verte (lorsque les C serveurs sont occupés et qu'il reste des places dans la zone verte) et la moitié (en moyenne) des clients qui arrivent dans la zone rouge (lorsque la zone verte est pleine et qu'il reste des places dans la zone rouge). Le temps

moyen des clients de type 2 dans le buffer s'écrit donc :

$$\begin{aligned} R_2^w &= \frac{1}{3} \sum_{p=0}^{k_3-1} \pi_{C+p} \frac{p}{\mu C} + \frac{1}{2} \sum_{m=0}^{k_2-1} \pi_{C+k_3+m} \frac{k_3+m}{\mu C} \\ &= R_3^w + \frac{1}{2\mu C} \sum_{m=0}^{k_2-1} (k_3+m) \pi_{C+k_3+m} \end{aligned}$$

où les valeurs de π_{C+k_3+m} sont données par les équations (6.20). Les clients de type 1, lorsqu'ils sont dans le buffer, peuvent être dans les trois zones de celui-ci. Ils constituent en moyenne un tiers des clients qui arrivent dans la zone verte (lorsque les C serveurs sont occupés et qu'il reste des places dans la zone verte), moitié des clients qui arrivent dans la zone rouge (lorsque la zone verte est pleine et qu'il reste des places dans la zone rouge) et la totalité des clients qui arrivent dans la zone bleue (lorsque la zone rouge est pleine et qu'il reste des places dans la zone bleue). Le temps moyen des clients de type 1 dans le buffer s'écrit donc :

$$\begin{aligned} R_1^w &= \frac{1}{3} \sum_{p=0}^{k_3-1} \pi_{C+p} \frac{p}{\mu C} + \frac{1}{2} \sum_{m=0}^{k_2-1} \pi_{C+k_3+m} \frac{k_3+m}{\mu C} \\ &\quad + \sum_{n=0}^{k_1-1} \pi_{C+k_3+k_2+n} \frac{k_3+k_2+n}{\mu C} \\ &= R_2^w + \frac{1}{\mu C} \sum_{n=0}^{k_1-1} (k_3+k_2+n) \pi_{C+k_3+k_2+n} \end{aligned}$$

où les valeurs de $\pi_{C+k_3+k_2+n}$ sont données par les équations (6.21).

8. On définit un système composé des clients de type 3 admis dans le commutateur et séjournant dans le buffer (donc en particulier dans la zone verte du buffer). On applique la loi de Little à ce système. Le débit stationnaire d'entrée des clients de type 3 dans ce système est $X_3^w = (1 - IP_3)\lambda_3$ puisqu'une proportion IP_3 des clients de type 3 – qui arrivent avec un taux λ_3 – sont rejetés. Le nombre moyen de clients de type 3 dans le buffer est doncé :

$$Q_3^w = R_3^w \cdot X_3^w = R_3^w \cdot (1 - IP_3)\lambda_3$$

On a de même

$$\begin{aligned} Q_2^w &= R_2^w \cdot X_2^w = R_2^w \cdot (1 - IP_2)\lambda_2 \\ Q_1^w &= R_1^w \cdot X_1^w = R_1^w \cdot (1 - IP_1)\lambda_1 \end{aligned}$$

Bibliographie

- [Altiok, 2012] Altiok, T. (2012). *Performance analysis of manufacturing systems*. Springer Science & Business Media.
- [Baynat, 2000] Baynat, B. (2000). *La théorie des files d'attente : des chaînes de Markov aux réseaux à forme produit*. Hermès.
- [Bolch et al., 2006] Bolch, G., Greiner, S., De Meer, H., and Trivedi, K. S. (2006). *Queueing networks and Markov chains : modeling and performance evaluation with computer science applications*. John Wiley & Sons.
- [Buzacott, 1967] Buzacott, J. A. (1967). Automatic transfer lines with buffer stocks. *The International Journal of Production Research*, 5(3) :183–200.
- [Buzacott and Shanthikumar, 1993] Buzacott, J. A. and Shanthikumar, J. G. (1993). *Stochastic models of manufacturing systems*. Pearson College Division.
- [Foata and Fuchs, 2004] Foata, D. and Fuchs, A. (2004). *Processus stochastiques : Processus de Poisson, chaînes de Markov et martingales*. Dunod.
- [Forestier, 1980] Forestier, J. (1980). Modélisation stochastique et comportement asymptotique d'un système automatisé de production. *Revue française d'automatique, informatique, recherche opérationnelle (RAIRO)*, 14(2) :127–143.
- [Gelenbe and Mitrani, 2010] Gelenbe, E. and Mitrani, I. (2010). *Analysis and synthesis of computer systems*, volume 4. World Scientific.
- [Menascé et al., 1994] Menascé, D. A., Almeida, V. A., and Dowdy, L. W. (1994). *Capacity Planning and Performance Modeling : from mainframes to client-server systems*. Prentice-Hall, Inc.
- [Onaga, 1969] Onaga, K. (1969). Maintenance and operating characteristics of communication networks. *Operations Research*, 17(2) :311–336.
- [Park, 1976] Park, S. (1976). *Queueing Systems : Vol. II, Computer Applications*. New York : John Wiley & Sons.
- [Stamp, 2017] Stamp, M. (2017). *Introduction to machine learning with applications in information security*. CRC Press.