الجمهورية الجزائرية الديمقراطية الشعبية
République algérienne démocratique et populaire
وزارة التعليم العالي والبحث العلمي
Ministère de l'enseignement supérieur et de la recherche scientifique
جامعة عين تموشنت بلحاج بوشعيب
Université –Ain Temouchent- Belhadj Bouchaib
Faculté des Sciences et de Technologie
Département des Mathématiques et de l'Informatique

Projet de Fin d'Etudes
Pour l'obtention du diplôme de Master en Mathématique
Domaine : Mathématique et Informatique
Filière : Mathématique
Spécialité : Probabilités et statistique appliquées
Thème

Etude de la stabilité des files d'attente

Présenté Par :
1) Melle BENDAHMA Saida Safaa

Devant le jury composé de :

M. HAMMOUDI Ahmed	Professeur	UAT.B.B (Ain Temouchent)	Président
Mme. SAKHI Hanane	M C B	UAT.B.B (Ain Temouchent)	Examinateur
Mme. MESSABIHI Aicha	M C B	UAT.B.B (Ain Temouchent)	Encadrant

Année Universitaire 2023/2024

Dédicace

Je dédie ce modeste travail

Á ma chère mère et à mon cher père.

Á ma soeure Safia et mon petit frère Tedj Elddin .

Á ma proche hanaâ.

BENDAHMA saida safaâ

Remerciements

Avant tout je remercie notre grand Dieu tout puissant pour exprimer ma reconnaissance envers sa grande générosité. Dieu m'a donné la volonté, la patience, la santé et la confiance durant toutes mes années d'études, ce qui ma donné le pouvoir de réaliser ce modeste travail.

Je tiens à exprimer ma profonde gratitude et sincères remerciements à Mme ***MESSABIHI Aicha*** *qui m'a encadré tout au long de la réalisation de ce travail . Je vous remercie pour votre précieuse présence et assistance, votre disponibilité et l'intérêt que vous avez manifesté pour ce modeste travail. Je vous remercie pour vos orientations et votre enthousiasme envers mon travail. Les judicieux conseils et rigueur que vous m'avez prodigué tout au long de ce travail. J'ai pris un grand plaisir de travailler avec vous.*

Mes remerciements vont également à tous les membres du jury, Monsieur ***HAMOUDI Ahmed*** *pour nous avoir fait l'honneur de pésider ce jury et Mme* ***SAKHI Hanane*** *pour examiner ce travail.*

Je tiens à remercier chaleureusement ma très chère famille sans exception pour leurs encouragement et leurs soutient et l'aide qu'ils m'ont apportés tout au long de ce travail peu importe la façon.

Enfin, je ne saurai oublier de remercier tous mes enseignants du département de Mathématiques, qui m'ont accompagné et aidé à m'améliorer durant mon cursus de formation.

Résumé

Ce travail explore en profondeur le concept des systèmes de files d'attente, avec un accent particulier sur les modèles markoviens et ses variantes et parmi les principales approche nous avona introduit la modélisation par les martingales. Le travail commence par donner des définitions de base nécessaires à l'analyse des systèmes de files d'attente. Une attention particulière est accordée à l'étude du comportement des clients (feedback, rappel, déroade et abandon) et de leur impact sur la performance des systèmes. En outre, quelques études antérieures sont présentées comme exemples pour montrer l'effet des différents paramètres du système sur les mesures de performance du modèle proposé. La stabilité des différents systèmes de files d'attente a été étudiée et nous avons abordé dans le troisième chapitre la méthode des équations d'équilibre (Chapman-Kolmogorov) pour étudier la stabilité d'un système de file d'attente.

Abstract

This work explores in depth the concept of queuing systems, with a particular emphasis on Markovian models and its variants and among the main approaches we introduced modeling using martingales. The work begins by giving basic definitions necessary for analysis queuing systems. Particular attention is given to the study of customer behavior (feedback, retriall, balking and reneging) and their impact on the performance of systems. In addition, some previous studies are presented as examples to show the effect of different system parameters on model performance measures propose. The stability of different queueing systems has been studied, and in the third chapter, we addressed the method of equilibrium equations (Chapman-Kolmogorov) to study the stability of a queueing system.

ملخص

يستكشف هذا العمل بعمق أنظمة الطوابير، مع التركيز بشكل خاص على النماذج الماركوفيانية وانواعها ومن بين الأساليب الرئيسية قمنا بتقديم النمذجة بواسطة المارتينجال. يبدا العمل بإعطاء التعريفات الأساسية اللازمة لتحليل انظمة الطوابير. يتم ايلاء اهتمام خاص لدراسة سلوك العملاء (التغذية الراجعة، التذكير، العزوف والعزوف العكسي) وتأثيرهم على اداء الانظمة. علاوة على ذلك، يتم تقديم بعض الدراسات السابقة كأمثلة لإظهار تأثير المعايير المختلفة للنظام على مقاييس الاداء للنموذج المقدم، تمت دراسة استقرار الانظمة المختلفة للطوابير، وقد تناولنا في المحور الثالث طريقة معادلات التوازن (شابمن كولموغوروف) لدراسة استقرار نظام الطوابير.

Table des matières

Introduction générale

Issue des travaux pionniers dErlang (1909) sur lanalyse des modéles pour la communication téléphonique, la théorie des files dattente est un vaste domaine scientifique à lévolution trés rapide qui couvre des domaines d'application trés larges. Elle se développe aujourdhui selon diffèrents axes . La théorie des files d'attente a pour objet l'analyse de l'évolution des files d'attente , l'élaboration et l'optimisation des indicateurs de mesure des performances. Le caractère massif des demandes de service, ainsi que la diversité des facteurs d'influence externe sur le systeme de file d'attente qui conduisent naturellement à une formulation dans le vocabulaire probabiliste.
La théorie de file d'attente est une discipline fondamentale des sciences de la gestion et de l'ingénierie qui étudie le comportement des systèmes dans lesquels des entités attendent pour être servies. Ces systèmes sont omniprésents dans notre vie quotidienne, allant des réseaux de télécommunications aux centres de santé, des services bancaires aux chaînes de production et jouent un rôle crucial dans la conception et l'optimisation des opérations. L'étude de file d'attente vise à comprendre les mécanismes complexes qui régissent le flux des entités à travers ces systèmes, en mettant l'accent sur des paramètres tels que le temps d'attente, le taux d'arrivée, la capacité du système et d'autres facteurs qui influencent l'efficacité opérationnelle. La modélisation mathématique et l'analyse quantitative sont des outils essentiels pour appréhender la dynamique des file d'attente et pour proposer des solutions efficaces aux défis opérationnels.

Plusieurs situations réelles peuvent être modélisées comme des systèmes de files d'attente associés à certains facteurs. Par exemple, dans les télécommunications, les transmissions des données de protocole sont parfois répétées. Ceci arrive fréquemment à cause de la médiocrité du service. Ces modèle de file d'attente ont été largement étudiés par un grand nombre de chercheurs qui sont traité sous le facteur de feedback. Takac's [**59**] a étudie la file M/M/1 avec feedback, il a déterminé le processus stationnaire de la longueur de la file et la distribution d'attente des clients dans le système. Nous avons aussi d'autres facteurs tels que : la dérobade (balking) dans une file d'attente qui se produit lorsqu'un client arrivant choisi de ne pas joindre la file, l'abandon (reneging) fait référence aux clients qui quittent la file d'attente sans être servi, le rappel (retrial) concerne les clients qui quittent la file vers une orbite pour être rappeler ultèrieurement.
L'objectif de ce manuscrit est consacré à l'étude de la stabilité de file d'attentes en spésifiant les différents facteurs . Ce manuscrit va être structuré autour de trois axes organisés comme suit :
Dans le premier chapitre, nous avons présenté les notions de bases de la théorie de file d'ttente :

la terminologie de la théorie de file d'attente et de certaines définitions et notations qui sont nécessaires dans l'étude des systèmes de files d'attente (la notation de Kendall [**9**], la formule de Little [**36**]). En suite, nous avons présenté les différents facteurs que peut prendre le client (impatience, rappel, feedback) et le serveur (vacance, homogene , hétérogéne).
Dans le deuxiéme chapitre , nous exposons comment peut on modéliser une file d'attente, nous abordons une modélisation par des processus stochastiques (markovien , non markovien et martingales) qui servent comme un outil mathématique de base pour résoudre les problemes d'attente. Nous abordons également leur propriétés fondamentales ainsi que quelques exemples.

Enfin, le dernier chapitre présente l'une des méthodes les plus utilisées pour étudier la stabilité des files d'attente connu sous le nom de l'équation de Chapman-Kolmogorov qui s'avère être un outil puissant. Cette équation relie les probabilités de transition entre les états du processus de Markov à différentes périodes de temps, offrant ainsi un moyen de déterminer la dynamique à long terme du système.

Chapitre 1

La théorie de file d'attente

1.1 Développement de la théorie de file d'attente

La théorie de file d'attente est une théorie de la recherche opérationnelle relevant du domaine de probabilité. Elle a été développé pour fournir des modèles mathématiques pour prédire l'évolution des files d'attente. Les origines du formalisme de file d'attente datent du début du XXème siècle, elle a commencé en 1909 avec les travaux de recherches de l'ingénieur danois Agner Krarup Erlang sur le trafic téléphonique de Copenhague pour déterminer le nombre de circuits nécessaires afin de fournir un service téléphonique acceptable. Entre 1909 et 1920, elle a étudié notamment les systèmes d'arrivée dans une file, les différentes priorités de chaque nouvel arrivant ainsi que la modélisation statistique des temps d'exécution. La première généralisation a été effectuée par Engset [**17**] dans le cas d'une source finie.
En 1928, d'autres modèles ont été publier par Fry sur les problèmes de téléttrafic. Parmi les précurseurs de cette période, nous citons Vaulot [**63**], O'Dell et Gibson [**43**]. Crommelin [**16**], Wilkinson [**66**].
À partir des années 30, Kolmogorov s'est intéressé sur les processus de naissance et de mort. En 1933, il publie son ouvrage où il propose son axiomatique de la théorie des probabilités qui est adopté jusqu'à aujourd'hui. D'autre part, Pollaczek remarque en 1934 que l'équilibre statistique ne permet pas toujours de décrire la micro-dynamique de tous les processus aléatoires en théorie de file d'attente, , Pollaczek et Khintchine s'intéressent au cas des durées de service arbitraires de la formule obtenue par d'Erlang dans le cas d'une distribution exponentielle, elles sont simplifiées de manière plus directe par Kendall en 1951 en utilisant la méthode de la chaîne de Markov induite.
L'analyse opérationnelle bien que souvent controversée constitue une approche complémentaire intéressant la théorie classique des files d'attente. Elle contribue à prouver la validité des résultats classiques pour les systèmes complexes lorsque les hypothèses conventionnelles ne peuvent être justifiées.

Dans les années 80, cette discipline devient beaucoup plus mathématique, et la littérature regorge d'articles décrivant des techniques ou des astuces mathématiques permettant de trouver des solutions exactes aux modèles, Les travaux de Lindley, Kingman et Marshall sont à l'origine

de nombreuses méthodes mathématiques plus complexes axées sur les questions d'approximation et de stabilité. Pour approfondir ce sujet, on peut se référer aux travaux de Borovkov [**14**] et Stoyan [**57**]. Le développement de l'approche des opérateurs dans la théorie de stabilité, connue sous le nom de méthode de stabilité forte, permet d'obtenir des estimations qualitatives de la stabilité avec un calcul précis des constantes, comme décrit par Aissani et Kartashov [**3**].

La théorie de file d'attente a évolué pour inclure des modèles plus complexes (rappel, feedback, dérobade, abandon,...) qui caractérisent les arrivées et (vacancies, homogénéité, hétérogénéité,...) pour caracteriser les serveurs.

1.2 Le formalisme de files d'attente

Dans ce chapitre, nous présentons le formalise file d'attent, la description s'échelonnera d'une file simple et de tous les paramètres qui permettent de la caractériser. Le formalisme va permettre de modèliser des phénomènes de partage de ressources où la ressource est une composante physique, logique ou humaine que les clients d'un système doivent obtenir afin de réaliser une activité.

1.2.1 Définition d'une file d'attente

Une file d'attente est un système caractérisé par un espace d'attente qui contient une ou plusieurs places et un espace de service composé d'un ou plusieurs serveurs. Les clients arrivent de l'extérieur à des instants aléatoires, ils attendent que l'un des serveurs soit libre pour pouvoir être servi puis quittent le système, les autres clients doivent attendre avant d'être servi formant ainsi une file d'attente, par exemple dans un réseau informatiques, le serveur représente le routeur et le client est présenté par le paquet.

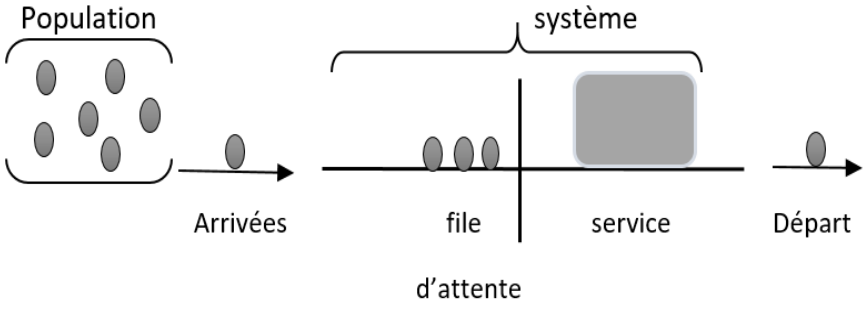

FIGURE 1.1 – Structure générale d'un système de file d'attente.

1.2.2 Structure et discipline de la file d'attente

Nombre de serveurs

Soit C le nombre de serveurs. Dès qu'un client arrive à la station, soit il y a un serveur de libre et le client entre instantanément en service, soit tous les serveurs sont occupés et le client se place dans le buffer en attente de libération d'un des serveurs. Si $C = 1$ le système de file d'attente fonctionnent avec serveur unique. Une station peut disposer de plusieurs serveurs en parallèle ou bien en série.

– **Système de file d'attente avec plusieurs serveurs en parallèles** : Dans ce type de file d'attente, on a C serveurs en parallèle. Le client entrant au système n'est pas obligé de visiter tous les serveurs. Si chaque serveur est doté d'un client, au moment de son arrivée, le client choisit d'attendre que l'un des serveurs soit libre pour le servir, si les serveurs ne sont pas tous occupés (le nombre de clients présents dans le système est inférieur à C), ce qui revient à dire tant que les serveurs ne sont pas tous occupé, alors la file ne se constitue pas et tout client arrivant est immédiatement pris en charge par l'un des serveurs libres.

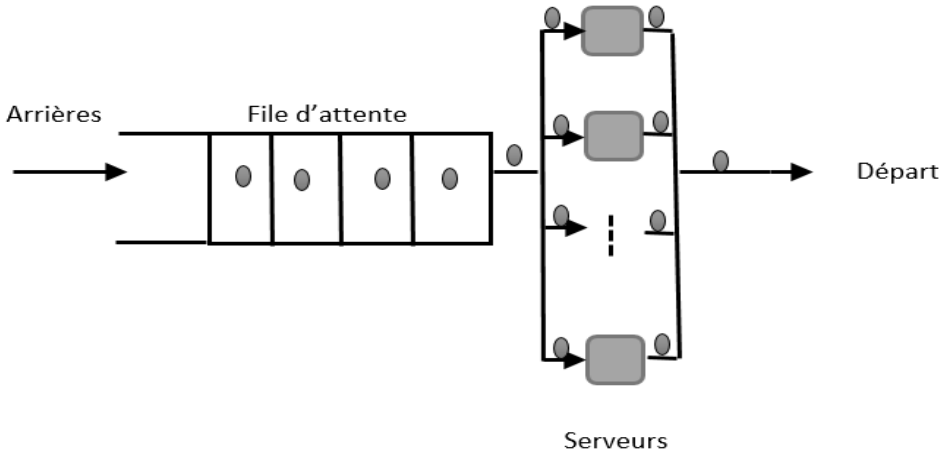

FIGURE 1.2 – Système de file d'attente à serveurs en parallèles.

– **Système de file d'attente avec plusieurs serveurs en série** : Le client entrant au système doit visiter plusieurs serveurs successifs dans un ordre fixe pour recevoir le service, ce type est appelé aussi système de file d'attente en cascade.

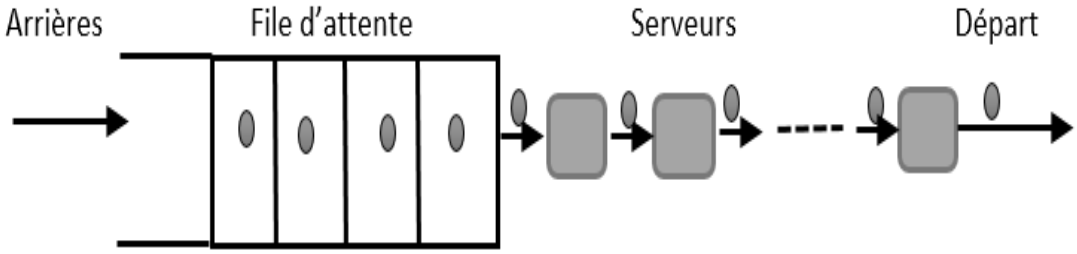

FIGURE 1.3 – Système de files d'attente à serveurs en série.

Capacité de la file d'attente

La capacité de la file peut être finie ou infinie. Lorsqu'elle est limitée et qu'un client arrive alors que cette file est pleine alors ce client est perdu ou rentre dans l'orbite si il s'agit des systèmes avec rappels. Par exemple : $M/D/1/\infty$ indique que c'est une file d'attente a capacité illimitée.

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 (voir[**9**]). Les disciplines les plus caurantes sont :

- **FIFO** (First In First Out) c'est la file standard dans laquelle les clients sont servis dans leur ordre d'arrivée.
- **LCFS** (Last Come First Served) Cela correspond à une file, dans laquelle le dernier client arrivé sera le premier traité.
- **RANDOM** (Aléatoire) Le prochain client qui sera servi est choisi aléatoirement dans la file d'attente.
- **Round-Robin** (Cyclique) Tous les client de la file d'attent entrent en service à tour de rôle, effectuent un quantum Q de leur temps de service soit totalement accompli. Cette discipline de service a été introduite afin de modéliser des systèmes informatiques.
- **PS** (Processor Sharing) Tous les clients sont servis simultanément en même temps.

1.3 Notation de Kendall-Lee

La notation introduite par David George Kendall [**9**] en 1953 de forme T/X/C/K permet de décrire les éléments qui constituent une file d'attente simple tels que :

- T : distribution d'interarrivée,
- X : distribution de service,
- C : nombre de serveurs,
- K : capacité de la file.

où T et X sont donnés par :

- M : loi exponentielle (markovienne),
- G : loi générale,
- GI : lois générale indépendantes,
- E_k : loi de Etlang-k,
- H_k : loi hyperexponentielle-k.

La Figure 1.4 suivante reprèsente un exemple d'une file $M/D/2/3$. Cette notation indique que une file d'attente à deux serveurs, la distribution des temps des interarrivées est exponentielle, le temps de service est déterministe, et la capacité de la file est trois clients.

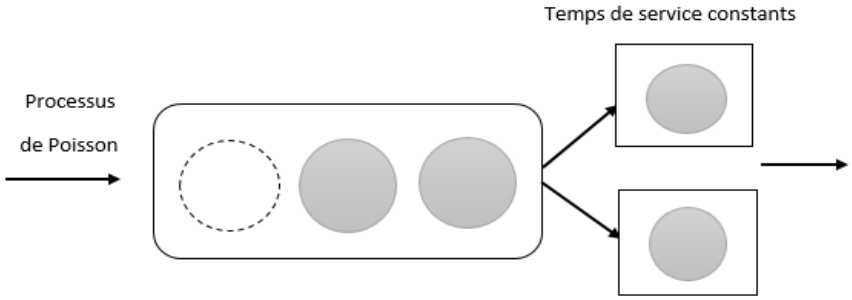

FIGURE 1.4 – Exemple d'une file M/D/2/3.

1.4 La loi de Little

La loi de Little est une relation empirique formulée par John Little en 1961. Elle ne concerne que le régime permanent du systéme. La loi de Little a une très grande importance dans l'analyse des systèmes de file d'attente.

Théoréme 1.4.1 (Formule de Little) *: La loi de Little établit une connexion entre trois concepts principaux dans la théorie des files d'attente (voir* **[36]***) : le nombre moyen de clients dans un système $\overline{L}$, le temps moyen passé dans le système $\overline{W}$, et est le taux dentrée dans le système λ, Le théorème de Little est souvent exprimé par l'èquation :*

$$\overline{L} = \lambda \overline{W}$$

En se concentrant sur l'attente dans la file (sans considérer le service) , la loi de Little permet de relier le nombre moyen de clients en attente $\overline{L_q}$ au temps moyen d'attente d'un client $\overline{W_q}$ avant le service :

$$\overline{L_q} = \lambda \overline{W_q}$$

En ne tenant compte que des serveurs, la loi de Little relie le nombre moyen de clients en service $\overline{L_s}$ au temps moyen de séjour d'un client $\overline{W_s}$ dans le service par la relation :

$$\overline{L_s} = \lambda \overline{W_s}$$

En d'autres termes, a partire se ces trois relation on peut déduire : $\overline{L} = \overline{L_s} + \overline{L_q}$ et $\overline{W} = \overline{W_s} + \overline{W_q}$

1.4.1 L'intensité du trafic

L'une des questions fondamentales dans l'analyse de file d'attente est la stabilité du système. Une file d'attente est dite stable si le nombre moyen de clients en attente reste borné au fil du temps, c'est-à-dire que la file d'attente ne devient pas infiniment longue. La stabilité est un aspect crucial pour garantir un bon service aux clients et éviter les engorgements du système. C'est pour cela, le problème de la dérivation des conditions de stabilité (intensité de trafic) est devenue ainsi un problème ouvert, surtout si on ne considère pas la supposition d'indépendance ou d'exponentialité des distributions des variables que régissent le système. Cela est d'essentiellement à la difficulté de dècrire la dynamique du système par une chaîne de Markov (voir [**56**]).

Exemple

En supposant que λ est le taux d'arrivée et μ le taux de service pour une file d'attente M/M/c, Une importante mesure est donnée à partir de ces deux paramètres qui est le taux d'utilisation du système (intensité de trafic) représenté par le rapport entre la demande et le taux de service (voir [**8**]) :

$$\rho = \frac{\lambda}{c\mu}, \text{ où } c \text{ est le nombre de serveurs dans le système.}$$

La condition nécessaire pour dire que le système M/M/c est stable est l'existence d'une intensité de trafic inférieure strictement à 1 ($\rho < 1$), c'est à dire $\lambda < c\mu$.

1.5 Files d'attente avec feedback

Le mot feedback est un mot qui caractérise le client qui quitte la file du à plusieurs facteurs, soit par l'insuffisance du nombre de serveurs, soit par une qualité de service médiocre ou bien le système est mal géré. Dans ce cas , le client retourne à la file pour demander son service.

1.5.1 Modèle d'attente M/M/1 avec Bernoulli feedback

Considérons un système de file d'attente M/M/1 avec Bernoulli feedback. Cette dernière peut modéliser un guichet unique où chaque client reçoit un service dont la durée est une variable exponentielle de paramètre μ et le processus d'arrivée des clients dans la file est un processus de Poisson de taux λ, $N(t)$ est le nombre de clients arrivant pendant un intervalle de temps $[0, t]$ suit une distribution de Poisson. Après avoir obtenu un service avec une probabilité β, le client peut rejoindre le système en tant que client Bernoulli feedback pour recevoir un autre service supplémentaire avec une probabilité $1 - \beta$. Sinon, il quitte définitivement le système, avec une probabilité β.

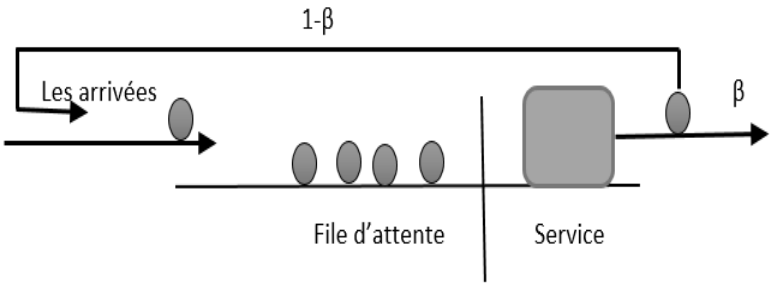

FIGURE 1.5 – Representation d'une file d'attente M/M/1 avec Bernoulli feedback.

1.6 Comportement des clients dans une file d'attente

Pour avoir une bonne modélisation d'un systeme de file d'attente, nous prenons en considération la dynamique du client dans le systeme de file d'attente. Génŕalement, dans une files d'attente classique le client ettend dans le buffer jusqu'a obtenir un service, s non i le client active une durée de temps sur laquelle il va choisir son comportement. Les files d'attente avec clients impatients (balking et/ou reneging) sont devenues un domaine vital dans la théorie des files d'attente. Ces systèmes ont une large application dans l'ingénierie des télécommunications, les réseaux informatiques, les systèmes de production et d'autres systèmes stochastiques.

1.6.1 File d'attente avec client impatients

Dans un système de file le d'attente, les clients sont dits impatients lorsqu'ils quittent le système avant d'être servis. Cela peut arriver soit dès l'arrivée, s'ils jugent la file trop importante, on parle alors de découragement, soit après avoir attendu et c'est alors un abandon. Les systèmes de file d'attente avec clients impatients (dérobade et abandon) apparaissent dans de nombreuses situations de la vie réelle, leur application est potentielle dans différents domaines tels que les systèmes de communication, les centres d'appels,...etc. Les clients impatients, découragés soit par la qualité de service soit par la longueur de la file d'attente ou abandonnés carrément la file sont devenus le but de plusieurs études.

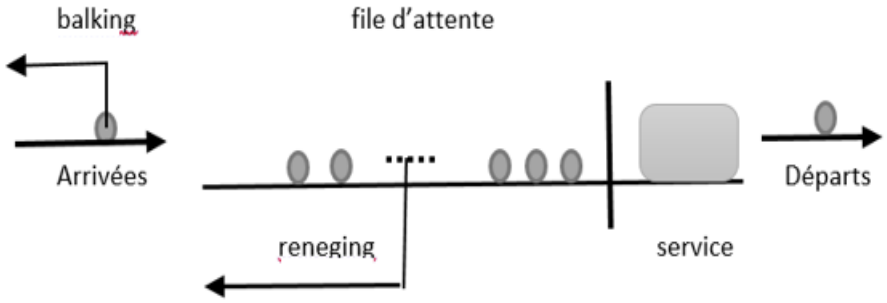

FIGURE 1.6 – système de file d' attente avec impatience.

Système de file d'attente avec dérobade

Les systèmes de file d'attente avec dérobade sont des modèles largement utilisés dans des problèmes de la vie réelle tels que les centrales d'appels téléphoniques, les urgences des hôpitaux, A son arrivée au système, le client découragé par la longueur de la file d'attente par exemple, décide de ne pas joindre la file d'attente, c'est le cas du dérobade. Cette impatience a poussé beaucoup de mathématiciens à l'étude de ce phénomène afin de trouver des solutions[**23**]
Haight [**25**] est le premier qui a travaillé sur les file d'attente avec balking. Il a supposé que le client admet une valeur de seuil N qui est la longueur de la file avant qu'il n'arrive au service, si le client observe que la longueur de la file est inférieur à N alors il rejoint la file d'attente dans le cas contraire il quitte la file.

Pour donner une bonne modélisation mathématique pour ce système de files d'attente avec dérobade il faut décrire ce modèle comme une file d'attente classique en ajoutant le caractère dérobade que subit le client impatient et le modéliser mathématiquement. Dans ce modèle, un client entrant dans le système qui trouve k clients devant lui. Il devient impatient et décide de ne pas joindre la file. Notons b_k la probabilité qu'un client rejoint le système lorsque il y a k clients dans le système au moment de son arrivée.
Rao [**48**], Ancker et Grafian [**5**] ont proposé l'hypothèse pour qu'un client se dérobe (balk) du système avec une probabilité de balking égale à

$$b_k = \frac{k}{N} \qquad\qquad k = 0, 1, 2, ..., N. \tag{1.1}$$

où n est le nombre de clients dans le système, N la taille de la file d'attente.
Singh [**52**] a supposé que si le client trouve les deux serveurs occupés alors il se dérobe du système avec une probabilité

$$b_k = 1 - r. \tag{1.2}$$

Où $0 \leq r \leq 1$. Et la même probabilité de balking se trouve dans Subba Rao [**58**]. Par ailleurs,

Van Tits et Van Der Veeken [**62**] ont proposé que la probabilité de balking est égale à

$$b_k = \frac{k}{k+1}. \tag{1.3}$$

Où k est la longueur de la file d'attente au moment de l'arrivé de client.
Récemment Lozano et Moreno [**37**] ont proposé la probabilité de balking suivante

$$b_k = 1 - r^k. \tag{1.4}$$

Où $0 \leq r \leq 1$, k la taille du système.

Modèle de file d'attente M/M/1 avec dérobade
Nous traitons une file d'attente $M/M/1$ avec dérobade etudié par [**27**]. Si un client arrive et trouve k clients déjà dans la file d'attente, il devient impatient et décide de ne pas joindre la file. La probabilité qu'un client reste dans le système malgré la présence de k clients dans la file est notée b_k. Ce modèle est défini par

des taux de naissances

$$\lambda_k = \lambda b_k, \qquad k = 0, 1, ...$$

avec

$$b_k = \frac{1}{k+1} \qquad k = 0, 1, ... \tag{1.5}$$

Les probabilités $\mathbb{P}_k$ de trouver k clients dans le système sont données par :

$$\mathbb{P}_k = \frac{\rho^k}{k!} e^{-\rho}.$$

La condition de stabilité pour ce modèle est que l'espérance de ρ soit finie, ce qui signifie que $E(\rho) < \infty$.

Paramètres de performances

Nombre moyen de clients

$\overline{L}_s$ Nombre moyen de clients dans le système :

$$\overline{L}_s = \rho.$$

$\overline{L}_q$ Nombre moyen de clients dans la file d'attente :

$$\overline{L}_q = \rho + e^{-\rho} - 1.$$

Temps moyen de séjour :

$$\overline{W}_s = \frac{\rho}{\mu(1 - e^{-\rho})}$$

Système de files d'attente avec abandon

Dans cette partie, nous évoquons un autre type de patience qui est représenté par le reneging. Les clients peuvent devenir impatients et quittent le système sans obtenir le service lorsque le temps d'attente est intolérable (long).

Deux catégories d'impatience se présentent (voir [**54**]) :

Impatience sans abandon : Dans ce cas, le client trouve qu'il a patienté trop longtemps par rapport au service qu'il a demandé mais il reste toutefois dans la file d'attente. Cette information peut être obtenue par le biais d'un questionnaire de satisfaction. Ce cas se produit surtout pour des services ou des biens qui font face à une pénurie ou encore à un monopole. Par exemple lorsqu'un client achète une voiture neuve, il n'est pas inhabituel qu'il ait à attendre plusieurs mois avant d'obtenir le modèle qu'il avait demandé. Cette situation décrit le phénomène de pénurie. En effet, les véhicules étant pour la plupart fabriqués à la demande du client pour respecter exactement le panel d'options qu'il a choisi, cela entraine un long délai de production. Dans ce cas, le client va éventuellement négocier un rabais si les délais de livraison ne sont pas respectés. Il s'impatiente mais va rarement annuler sa commande pour se tourner vers un concurrent, il ne va donc pas abandonner. La situation de monopole se produit par exemple à un péage autoroutier lors des heures de pointe. Les véhicules forment des files d'attente à chaque portique de péage, les automobilistes peuvent s'impatienter mais ils n'ont pas d'autre choix que de passer par ce péage. Là encore, ils peuvent s'impatienter mais n'abandonneront pas.

Impatience avec abandon : Dans ce cas, les clients qui s'impatientent quittent immédiatement le système. Cela se produit lorsque la ressource demandée est accessible à un faible coût pour le client ou lorsque le nombre de demandes pour cette ressource varie fortement. Dans ce cas, le client peut avoir intérêt à abandonner sa demande pour la reformuler plus tard. C'est ce qui se produit souvent dans les centres d'appels téléphoniques. Les clients qui s'impatientent raccrochent, car ils n'ont pas été servis assez vite.

Les principales règles de reneging dans beaucoup de recherches sont : Le cas le plus simple lorsque T est une constante fixe (Boots et Tijms [**13**], Xiong et Altiok [**67**]). Dans d'autres travaux le temps d'attente maximal T est supposé distribuer selon la loi exponentielle de paramètre i où i est le nombre de clients dans le système (Rao [**48**], Wang et Chang [**64**]). Le cas où le temps d'attente est arbitrairement distribuée se trouve dans (Andreas et Manfred [**6**], Ward et Glynn [**65**], Zeltyn et Mandelbaum [**72**]). De plus, la lenteur du service (Omarah [**45**]) et la panne de service (Blackburn [**11**], Nasrallah [**41**]) peuvent provoquer l'impatience (Perel et Yachiali [**46**]). Il y a également d'autres hypothèses particulières concernant le comportement du reneging. Adan et al. [**2**] ont examiné le cas où les clients abandonnent la file simultanément, par exemple le cas des systèmes à distance où les clients abondonnent le système une fois l'installation de transport est disponible.

1.6.2 File d'attente avec rappel

Les systèmes de file d'attente avec rappel sont des systèmes utilisés dans la modélisation des réseaux de télécommunication et dans les systèmes informatiques. C'est le cas pour les appels téléphoniques par exemple, entre deux rappels successifs, le client en question se trouve en orbite.

Ce phénomène de file d'attente avec rappels ou avec répétition d'appels [**10**] sont caractériseés par la propriété suivante : un client arrivant dans le système et qui trouve tous les serveurs et les positions d'attente occupés quitte le système définitivement ou rappelle ultérieurement à des instants aléatoires. Un client qui attend pour rappeler est dit en orbite. Les résultats analytiques et les techniques utilisées pour ses modèles sont résumés dans les articles de synthèse de Yang et Tempteton [**68**] et Falin [**19**] ainsi que dans la monographie de Falin et Templeton [**20**].
Un système de files d'attente est composé de c serveur avec $c \geq 1$, d'un buffer de capacité $K-c$ ($K \geq c$) et d'une orbite de capacité N. les arrivées des clients dans le système sont aléatoires et les temps de service distribués selon une loi donné, mais au moment de son arrivée, un client qui trouve les serveurs occupés, soit il rejoint la file d'attente soit il quitte l'espace de service pour renouveler sa demande de service après une durée de temps aléatoire. La capacité de l'orbite peut être finie ou infinie.

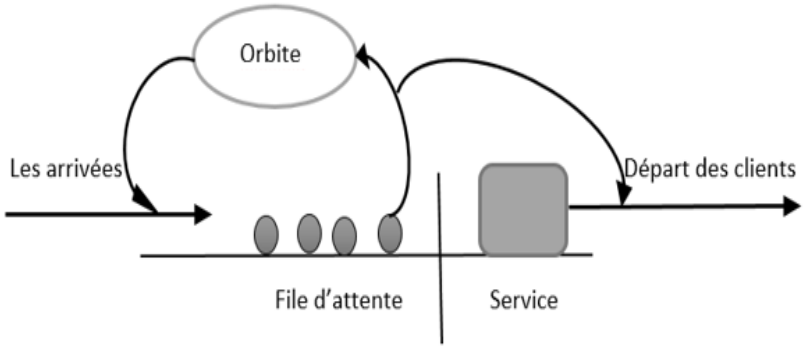

FIGURE 1.7 – système de file d' attente avec rappel.

Modèle de files d'attente M/M/1 avec rappels

Nous considèront un système de file d'attente sans positions d'attente traité dans [21]. Il y a un seul serveur qui assure le service. Les clients primaires arrivent selon un processus de Poisson de taux $\lambda > 0$. Les durées de service suivent une loi exponentielle de fonction de répartition $B(x) = 1 - e^{-\mu x}, x \geq 0$, avec une moyenne $\frac{1}{\mu}$. Les intervalles du temps entre deux rappels consécutifs sont également exponentiels de paramètre $\theta > 0$. Nous supposons que les durées entre deux rappels consécutifs ainsi que entre deux arrivées primaires successives sont mutuellement indépendantes. L'état du système peut être décrit par le processus

$$\{K(t), N_o(t), t_0 \geq 0\} \tag{1.6}$$

est de Markov d'espace d'états $\{0, 1\} \times \mathbb{N}$, Où

$$K_n(t) = \begin{cases} 0, & \text{si le serveur est libre} \\ 1, & \text{sinon} \end{cases}$$

Et $N_o(t)$ est le nombre de clients en orbite à l'instant t.

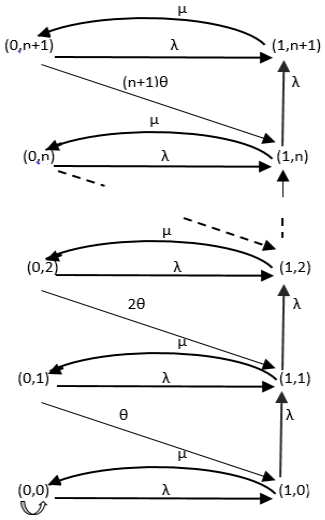

FIGURE 1.8 – La structure générale du modèle M/M/1 avec rappels.

Supposons que le régime stationnaire existe ($\lambda < \mu$). A partir du Figure 1.8, les équations d'équilibre statistique sont :

$$\begin{cases} 0 = \mu\pi_{1,n} - (\lambda + n\theta)\pi_{0,n} \\ 0 = \lambda\pi_{0,n} + (n+1)\theta\pi_{0,n+1} + \lambda\pi_{1,n-1} - (\lambda + \mu)\pi_{1,n}. \end{cases}$$

pour

$$K(t) = i, \qquad i = 0, 1$$

et

$$N_o(t) = n, \qquad n \geq 0,$$

on $\pi_{in} = \lim_{t\to\infty} \mathbb{P}(K(t) = i, N_o(t) = n)$, représentent la distribution stationnaire conjointe de l'état du serveur et du nombre de clients en orbite, est donnée par

$$\pi_{0n} = \frac{\rho}{n!\theta^n} \prod_{k=0}^{n-1} (1 + k\theta)(1 - \rho)^{\frac{\lambda}{\theta}+1}, \tag{1.7}$$

$$\pi_{1n} = \frac{\rho^{n+1}}{n!\theta^n} \prod_{k=1}^{n} (\lambda + k\theta)(1 - \rho)^{\frac{\lambda}{\theta}+1}. \tag{1.8}$$

La distribution stationnaire existe si $\rho = \dfrac{\lambda}{\mu} < 1$.

1.7 Mécanisme des serveurs dans une file d'attente

1.7.1 File d'attente avec serveur en vacance

L'absence temporaire due aux serveurs pendant une certaine période dans un système d'attente au moment de l'achèvement d'un service, qu'il y ait ou non des clients en attente dans une file est appelée vacance des serveurs. Les arrivées qui attendent le service peuvent bénéficier du service une fois que le serveur a terminé sa période de vacance. Il existe de nombreuses situations qui peuvent entraîner un serveur en vacance, par exemple la maintenance du système, une panne de machine, le partage de ressources, des serveurs cycliques (où le serveur est censé servir plus d'une file d'attente) (voir [**8**]). Les modèles de file d'attente avec vacances du serveur ont trouvé une large applicabilité dans de nombreux systèmes à temps réel (réseau informatique et système de communication,...). Les file d'attente avec vacances et clients impatients ont joué un rôle important dans les situations de congestion quotidienne et industrielle des systèmes informatiques. Un grand nombre de travaux de recherche traitaient ces modèles. À partir d'année 2003, Zhang et Tian ont analysé les systèmes d'ttente multi-serveurs avec vacances [**31**], Artalejo et Gomez-corral [**7**] ont étudié le système de files d'attente M/G/1 avec rappel et serveurs en vacance. Yue [**70**]ont examiné plus en détail le modèle de files d'attente M/M/2 avec vacancies multiple de Kumar et Madheswari [**34**], Madan et al. [**38**] ont étudié une file d'attente à serveur unique avec des vacances de serveur de type phase optionnelle basée sur un service exhaustif. Nous considérons deux types de vacances :

- Vacance simple (SV) : Lorsque le système est vide les serveurs sortent en vacance. S'ils reviennent du vacance et trouvent le système vide, ils resentent inactifs jusqu'à ce que le premier client arrive.
- Vacance multiple (MV) : Si les serveurs reviennent du vacance trouvant la file d'attente vide, ils partent immédiatement tous ensemble pour de nouvelles vacancies.

1.7.2 File d'attente avec serveurs hétérogènes ou homogénes

Dans la recherche, la plupart des travaux sont réalisés pour les files d'attente multi-serveur avec homogénéité de service. Ceci n'est valable que lorsque le processus est à contrôle mécanique ou électrique. Mais en réalité ce mode n'est pas toujours valide, le service peut être long, court, ou normal. C'est à dire que le serveur ne peut pas fournir un service de même taux pour tous les clients, l'exemple le plus concrêt si le serveur est humain, donc ce cas on ne peut pas garder le même taux de service tout au long de son service vue la nature humaine, il perd son efficacité de service instantanément si la file est trop longue ainsi nous sommes entrain de parler de files d'attente avec serveurs hétérogènes. Dans la vie réelle, les modèles des systèmes de files d'attente avec différentes intensités de service sont utilisés pour l'étude des processus de télécommunication, en informatique, en industrie,...,etc.

L'hétérogénéité du service est une caractéristique commune á de nombreuses situations réelles de files d'attente multi-serveurs. Les mécanismes de services hétérogénes sont des méthodes de planification inestimables qui permettre aux clients de recevoir une qualitéde service différente.

Le service hétérogene est clairement une caractéristique principale du fonctionnement de presque tous les systemes de fabrication. Le rôle de la qualité et la performance des services sont des aspects cruciaux dans la perception des clients et les entreprises doivent leur accorder une attention particuliere lors de la conception et de la mise en uvre de leurs opérations. c'est pour cette raison que les files d'attente avec des serveurs hétérogenes ont reu une attention considérable dans la littérature. Morse [**40**] est le premier qui a introduit la notion d'hétérogénéité dans le service et a obtenu des résultats en régime permanent pour son modèle de files d'attente. Saaty [**51**] a examiné le problème de Morse et a obtenu les expressions explicites des probabilités en régime stationnaire et du nombre moyen dans le système. Ancker et Cafarian [**5**] ont étudié l'état stationnaire de la file $M/M/S/N$ avec abandon et serveur hétérogènes. Ils ont dérivé les mesures de performance de ce système et une comparaison était faite entre système avec serveurs hétérogènes et système avec serveurs homogènes équivalent. Sharma et Dass [**53**] ont analysé les distributions stationnaire pour le système de files d'attente $M/M/2/N$ à serveurs hétérogènes.

Ensuite, Rykov [**49**] a étudié le contrôle monotone d'un système de files d'attente avec serveurs hétérogènes . Dans Al Seedy [**4**], l'auteur proposait la solution transiente pour la file $M/M/2$ avec dérobade. Kumar et Madheswari [**34**] ont analysé un système de files d'attente $M/M/2$ avec vacances multiple en utilisant la méthode de la matrice géométrique. Dans Kumar et al [**35**], les auteurs ont dérivé la solution transiente pour un système de files d'attente avec deux serveurs hétérogènes avec catastrophe on définissant une suite de fonction génératrice. Artalejo et Gomez-corral [**7**] ont analysé le système de files d'attente $M/G/1$ avec rappel et serveurs en vacance.

Rykov et Efrosinin [**50**] ont étudié le contrôle optimal de système de files d'attente avec serveurs hétérogènes.

Krishnamoorthy et Sreenivasan [**33**] ont présenté un modèle de files d'attente $M/M/2$ à deux serveurs hétérogènes dont l'un des serveurs reste inactif et le second serveur est en vacance lorsque le système est vide. Sridhar et Pitchai [**55**] ont analysé le système de files d'attente $M/M/2$ avec serveurs hétérogènes et vacances de serveurs, l'état stationnaire de ce système est étudié par la méthode de la matrice géométrique. Rajan [**47**] a considéré un système de files d'attente $M/M/2$ avec catastrophe dont le premier serveur est tout le temps disponible et le deuxième serveur disponible par intermittence. Ce modèle est résolu par la technique de la matrice géométrique. Kalyanaraman et Kalaiselvi [**29**] ont présenté l'analyse d'un système de files d'attente à deux serveur hétérogènes dont un des serveurs admet un seuil de service.

Chapitre 2

Modélisation de file d'attente

Selon le concept du systéme de file d'attente, nous pouvons adopter les modélisations suivantes :

2.1 Modélisation par chaine de Markov

Le formalisme des chaines de Markov est un outil trés efficace pour l'analyse des systèmes de file d'attente. L'objectif pragmatique dans cette section est de montrer comment modèliser un systéme de files d'attente pour analyser ou pour prévoir son évolution.
Nous notons dans ce chapitre par $\{X_n\}_{n\in\mathbb{N}}$ une suite de variables aléatoires à valeurs dans l'ensemble E espace d'état de dimension finie ou infinie.

2.1.1 Chaîne de Markov à temps discret

Définition 2.1.1 [9] *La suite de variable aléatoire $\{X_n\}_{n\in\mathbb{N}}$ est une chaîne de Markov à temps discret si $\forall i, j \in E$ et pour tout $n \in \mathbb{N}$:*

$$\mathbb{P}[X_n = j | X_{n-1} = i_{n-1}, X_{n-2} = i_{n-2}, ..., X_0 = i_0] = \mathbb{P}[X_n = j | X_{n-1} = i_{n-1}] \tag{2.1}$$

Dans ce cas on peut définir la probabilité de transition de l'état i vers un état j et dire que la chaine de Markov est homogène si

$$p_{ij} = \mathbb{P}[X_n = j | X_{n-1} = i] \qquad \forall n \in \mathbb{N} \tag{2.2}$$

Proposition 2.1.1 *La matrice de transition $P = [p_{ij}]_{i,j\in E}$ est une matrice carrée d'ordre fini ou infini, vérifie les propriétés suivantes :*

$$\sum_{j\in E} p_{ij} = 1, \qquad \forall i \in E$$

$$p_{ij} \geq 0, \qquad \forall (i,j) \in E^2$$

$$P = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1j} & \cdots \\ p_{21} & p_{22} & \cdots & p_{2j} & \cdots \\ \vdots & \vdots & \ddots & p_{3j} & \cdots \\ p_{i1} & p_{i2} & \cdots & p_{ij} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

Exemple 2.1.1 *Nous allons détailler le cas le plus simple d'une chaîne de Markov à deux états fini. Considérons un téléphone qui peut se trouver dans deux états : "0" la probabilité pour le téléphone est libre, est "1" pour le téléphone est occupé, L'ensemble des états du téléphone est* $E = \{0, 1\}$. *Si à un moment n, le téléphone est libre, il devient occupé au moment* $n + 1$ *avec la probabilité* α, *ou il reste libre avec une probabilité de* $1 - \alpha$.
Si à un moment n le téléphone est occupé, alors au moment $n + 1$ *il devient libre avec la probabilité* β *ou il reste occupé avec une probabilité de* $1 - \beta$.
$X_n : \Omega \to \{0, 1\}$ *est une variable aléatoire pour* $n = 0, 1, ...$ *sous les hypothèses suivantes :*

$$p_{00} = \mathbb{P}[X_{n+1} = 0 | X_n = 0] = 1 - \alpha, \quad p_{01} = \mathbb{P}[X_{n+1} = 1 | X_n = 0] = \alpha,$$
$$p_{10} = \mathbb{P}[X_{n+1} = 0 | X_n = 1] = \beta, \qquad p_{11} = \mathbb{P}[X_{n+1} = 1 | X_n = 1] = 1 - \beta,$$

sa matrice de transition est

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

avec α *et* β *des réels dans* $[0, 1]$. *Le graphe de cette chaîne est illustré dans la Figure 2.1.*

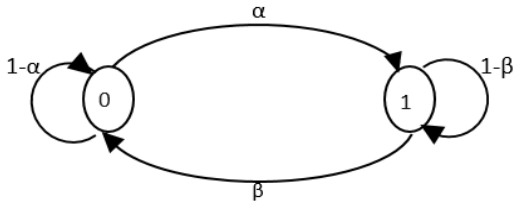

FIGURE 2.1 – Représentation graphique d'exemple 2.1.1.

2.1.2 Processus de Markov à temps continu

Définition 2.1.2 *Le processus stochastique* $\{X(t)\}_{t \geq 0}$ *est une chaîne de Markov à temps continu si* $\forall n$ *et* $\forall t_0 < t_1 < ... < t_n$ *:*

$$\mathbb{P}[X(t_n) = j | X(t_{n-1}) = i_{n-1}, X(t_{n-2}) = i_{n-2}, ..., X(t_0) = i_0] = \mathbb{P}[X(t_n) = j | X(t_{n-1}) = i_{n-1}]$$

De la même façon évoquée dans la définition précedante nous disons que la chaîne de Markov à temps continu est homogène si la porobabilité de transition $\mathbb{P}_{ij}(t)$ est définie par :

$$\mathbb{P}_{ij}(t) = \mathbb{P}[X(s+t) = j | X(s) = i] \qquad \forall s \geq 0 \qquad (2.3)$$

Proposition 2.1.2 – *Nous associons à une CMTC une matrice $Q = [q_{ij}]_{i,j \in E}$ dite générateur infinitisémal, qui vérifie les propriétés suivantes :*

$$q_{ij} = \mu_{ij}, \qquad \forall i \neq j$$

$$q_{ii} = -\sum_{i \neq j} \mu_{ij},$$

μ_{ij} est le taux de transition de l'état i vers l'état j.

$$Q = \begin{pmatrix} q_{11} & q_{12} & \cdots & q_{1j} & \cdots \\ q_{21} & q_{22} & \cdots & q_{2j} & \cdots \\ \vdots & \vdots & \ddots & p_{3j} & \cdots \\ q_{i1} & q_{i2} & \cdots & q_{ij} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

- *Le temps passé dans un état i d'une CMTC a une distribution exponentielle de taux μ_i.*
- *Les transitions d'un état i vers les autres états n sont probabilistes notées $\mathbb{P}_{in}$.*
- *Le taux de transition de l'état i vers létat j est $\mu_{ij} = \mu_i \mathbb{P}_{ij}$.*

2.1.3 Classification des états

Définition 2.1.3 (Irréductibilité) [9] *Une Chaîne de Markov est dite irréductible ssi de tout état i on peut atteindre tout état j (le nombre des étapes est fini)*

Remarque 2.1.1 [71] *$p_{ii}(0) = 1$ ainsi un état i est toujours accessible à partir de lui-même (état absorbant) : $i \to i$.*

Définition 2.1.4 *Les états i et $j \in E$ communiquent si $i \to j$ et $j \to i$. En notation : $i \leftrightarrow j$.*

Définition 2.1.5 (La probabilité de transition) [9][26] *Nous définons p_{ij}^m, la probabilité de transition de l'état i à l'état j en m étapes :*

$$p_{ij}^m = \mathbb{P}[X_{n+m} = j | X_n = i] \qquad \forall n \in \mathbb{N}$$

$$p_{ij}^{(0)} = p_{ij}(0) = \begin{cases} 1, & si\ i = j \\ 0, & si\ i \neq j \end{cases}$$

Théoréme 2.1.1 [71] *$i \leftrightarrow j$ est une relation d'équivalence, à savoir :*

1. $i \leftrightarrow i$ *(Réflexivité).*
2. $i \leftrightarrow j \Leftrightarrow j \leftrightarrow i$ *(Symétrie).*
3. $i \leftrightarrow j$ *et* $j \leftrightarrow k \Rightarrow i \leftrightarrow k$ *(Transitivité).*

Preuve. On suppose que : $i \leftrightarrow j, j \leftrightarrow k$ donc il existe m et n tel que $p_{ij}^m > 0$ et $p_{jk}^n > 0$ d'où

$$\begin{aligned} p_{ik}^{m+n} &= \mathbb{P}[X_{m+n} = k | X_0 = i] \\ &\geq \mathbb{P}[X_{m+n} = k, X_m = j | X_0 = i] \\ &= \mathbb{P}[X_m = j | X_0 = i] \mathbb{P}[X_{m+n} = k | X_m = j] \\ &= p_{ij}^m p_{jk}^n > 0. \end{aligned}$$

La relation de communication ($\leftrightarrow$) forme une relation d'équivalence car elle satisfait les propriétés de réflexivité de symétrie et de transitivité. Par conséquence, elle divise l'espace des états E en classes d'équivalence disjointes appelées classes de commmunication.

Définition 2.1.6 **[71]** *Une classe de communication C est ouverte s'il existe un état $i \in C$ et un état $k \notin C$ tels que $i \to k$. Sinon, une classe de communication est appelée fermée. Si une chaîne de Markov arrive une fois dans une classe de communication fermée, elle restera dans cette classe pour toujours.*

Définition 2.1.7 (La périodicité) *Soit $i \in E$, on appelle période de i, et on note $d(i)$. Le PGCD de tous les entiers $n \geq 1$, est définie par :*

$$d(i) = PGCD\{n \geq 1, p_{ii}(n) > 0\}.$$

Si $d(i) = 1$ l'état i est dit apériodique.

Lemme 2.1.1 **[71]** *Si l'état $i \in E$ est apériodique et $i \leftrightarrow j$, alors j est également apériodique.*

Définition 2.1.8 *Pour tout états j, le temps de premier retour en j est définit par*

$$T_j = min\{n \geq 1; X_n = j\} \in \mathbb{N} \cup \{+\infty\}.$$

Définition 2.1.9 (Récurrent) *On dit que l'état j est récurrent si partant de l'état j, la probabilité que la chaîne de Markov retourne à l'état j infini de fois est égale à 1*

$$\mathbb{P}(T_j < +\infty | X_0 = j) = 1.$$

Définition 2.1.10 (Transient) *Un état j est dit transient s'il n'est pas récurrent :*

$$\mathbb{P}(T_j < +\infty | X_0 = j) < 1.$$

Remarque 2.1.1 *Une chaîne irréductible ne contient aucun état absorbant ou transient.*

Théoréme 2.1.2 [71]
Si $i \in E$ est un état récurrent et $j \leftrightarrow i$, alors j est également récurrent.
Si $i \in E$ est un état transitoire et $j \leftrightarrow i$, alors j est également transitoire.

Définition 2.1.11 [56] *Notée $f_{ij}^{(n)}$ la probabilité d'aller de l'état i vers l'état j en n étapes*
On définit :

$f_{ij}^{(n)} = \mathbb{P}(T_j = n | X_0 = i) = \mathbb{P}(X_n = j, X_{n-1} \neq j, ..., X_1 \neq j | X_0 = i)$.
f_{jj} la probabilité de revenir en j après l'avoir quitté : $f_{jj} = \sum_{n=1}^{\infty} f_{jj}^{(n)}$
et M_j le temps moyen de retour en j : $M_j = \mathbb{E}[T_j] = \sum_{n=1}^{\infty} n f_{jj}^{(n)}$.

Théoréme 2.1.3 [9] *un état j est dit :*
- *transitoire si* $\sum_{n=1}^{\infty} f_{jj}^{(n)} = \mathbb{P}[T_j < \infty] < 1$
- *récurrent si* $\sum_{n=1}^{\infty} f_{jj}^{(n)} = \mathbb{P}[T_j < \infty] = 1$, *de plus il existe deux sortes d'états récurrents*
  L'état récurrent nul si le temps moyen de retour est infini : $M_j = \infty$.
  L'état récurrent positif si le temps moyen de retour est fini : $M_j < \infty$.

Théoréme 2.1.4 [56]
Les états d'une chaîne de Markov finie, irréductible et apériodique sont ergodiques.

Le nombre de visites d'un état i par la chaîne de Markov $\{X_n\}_{n\geq 0}$ est donné par

$$N_i = \sum_{n\geq 1} \mathbb{I}_{\{X_n = i\}}.$$

L'espérance du nombre de visites à partir de l'état i vers l'état i est donnée par

$$\mathbb{E}_i(N_i) = \sum_{n=1}^{\infty} p_{ij}(n).$$

Proposition 2.1.3 *Un état i est récurrent si et seulement si*

$$\mathbb{P}_i(N_i = \infty) = 1.$$

Lemme 2.1.2 *Pour tout état i de E, il n'existe que deux possibilités :*
- *i est transitoire si et seulement si*

$$\sum_{n\geq 1} p_{ii}(n) < \infty,$$

– *i est récurrent si et seulement si*

$$\sum_{n\geq 1} p_{ii}(n) = \infty.$$

Exemple 2.1.2 *Nous illustrons dans la Figure 2.2, une transition d'une chaîne de Markov*

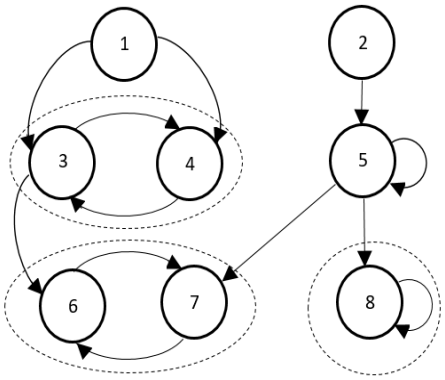

FIGURE 2.2 – Graphe de transition.

l'état 1 et l'état 2 sont chacun des états non retournés, les états 3 et 4 constituent une classe communicante qui n'est pas fermée, l'état 5 est un état de retour et constitue une classe communicante non fermée, les états 6 et 7 forment ensemble une classe communicante fermée, et l'état 8 est un état de retour qui forme une classe communicante fermée à un seul état. L'état 8 est un état absorbant. Ainsi, les états sont partitionnés comme suit :

$$\{1\}, \{2\}, \{3, 4\}, \{5\}, \{6, 7\}, \{8\}.$$

Les états de 1 à 5 sont des états transitoires, et les états de 6 à 8 sont des états positivement récurrents. L'ensemble des états récurrents $\{6, 7, 8\}$ *est fermé et contient deux sous-ensembles propres fermés* $\{6, 7\}$ *et* $\{8\}$.

2.1.4 Distributions stationnaires

La distribution stationnaire est une des caractéristiques les plus faciles à calculer des processus de Markov. L'étude de cette idée a comencé par le calcul d'Erlang (voir **[18]**).

Définition 2.1.12 *On dit qu'un vecteur de probabilité* $\pi = [\pi_0, \pi_1, ...]$, *est une distribution stationnaire de la chaîne* $(X_n)_n$ *de matrice de transition* P, *si et seulement si*

$$\forall i \in E, \pi_i \geq 0 \qquad et \qquad \sum_{i\in E} \pi_i = 1. \tag{2.4}$$

On a

$$\pi = \pi.P$$

Ou de façon équivalente, si pour tout $j \in E$

$$\pi_j = \sum_{i \in E} \pi_i p_{ij}.$$

Théoréme 2.1.5 **[12]** *Pour toute chaîne de Markov irréductible sur un espace d'états fini E, il existe une unique mesure de probabilité invariante π.*

Théoréme 2.1.6 **[10]** *Soit π la distribution stationnaire d'une chaîne de Markov si $\pi(0) = \pi$ alors $\forall n$, la loi $\pi(n)$ des états au temps n est invariante pour tout n,*

Définition 2.1.13 **[15]** *Si π est une distribution stationnaire, alors pour tout état récurrent j :*

$$\pi_j = \frac{1}{M_j}$$

où M_j est le temps moyen de retour à j :

Définition 2.1.14 (Distribution limite) **[56]** *Soit P la matrice de probabilité de transition d'une chaîne de Markov homogène en temps discret et soit $\pi(0)$ une distribution de probabilité initiale, si $\lim_{n\to\infty} P^{(n)} = \lim_{n\to\infty} P^n$ existe, alors la distribution de probabilité est égale à*

$$\pi = \lim_{n\to\infty} \pi(n) = \lim_{n\to\infty} \pi(0)P^{(n)} = \pi(0) \lim_{n\to\infty} P^{(n)} = \pi(0) \lim_{n\to\infty} P^n$$

Définition 2.1.15 **[15]** *Une chaîne transiente infinie n'admet pas de distribution stationnaire.*

Démonstration
On suppose qu'il existe une distribution stationaire π : alors $\pi P = \pi$.
Pour tout état j et pour tout entier n, $\pi P^n = \pi$, $\pi_j = \sum_i \pi_i p_{ij}^n$; les état étant transitoires, tous les $p_{ij}^{(n)}$ tendent vers 0.

$$\forall j, \qquad \pi_j = \lim_{n\to\infty} \sum_i \pi_i p_{ij}^{(n)} = \sum_i \lim_{n\to\infty} \pi_i p_{ij}^{(n)} = 0,$$

ce qui est contradictoire avec $\sum_j \pi_j = 1$.

2.1.5 Processus de naissance et de mort

Les processus de naissance et de mort sont des processus stochastiques à temps continu a valeurs dans $\mathbb{N}$ et à espace détats discret définit comme étant des processus sans mémoire.

Définition 2.1.16 **[9]** *Un processus stochastique est une famille de variables aléatoires $X(t), t \in E$ où chaque variable alétoire $X(t)$ est indexée par le paramètre $t \in E$. Si F est un ensemble de $\mathbb{R}_+$, alors t signifie temps.*

- *Nous disons que $X(t)$, $t \in F$ est un processus à temps discret, si E est dénombrable, i.e $E \subseteq \mathbb{N}$.*
- *Nous disons que $X(t)$, $t \in F$ est un processus à temps continu, si E est un intervalle de $[0, \infty)$*

$X(t)$ définit l'état du processus à un instant donné, l'ensemble des valeurs que peut prendre le processus à chaque instant est appelé espace d'état qui peut également être soit discret (fini ou infini dénombrable) ou continu, donc nous écrivons $(X_n)_{n\in\mathbb{N}}$ pour le processus à temps discret et $(X_t)_{t\in\mathbb{R}^+}$ pour le processus à temps continu.

Définition 2.1.17 *On peut réaliser un processus de naissance et de mort sous les hypoyhèses suivantes :*

- *Les arrivées et les départs d'entités obéissent à des lois exponentielles de taux λ_n et μ_n.*
- *La probabilité que deux événements se produisent dans un intervalle de temps est négligeable.*
- *Il y a une transition vers un état voisin, soit par l'arrivée d'un client (naissance), soit par le départ d'un client (mort).*
- *Si tous les λ_n sont nuls, on parle de processus de mort.*
- *Si tous les μ_n sont nuls, on parle de processus de naissance.*

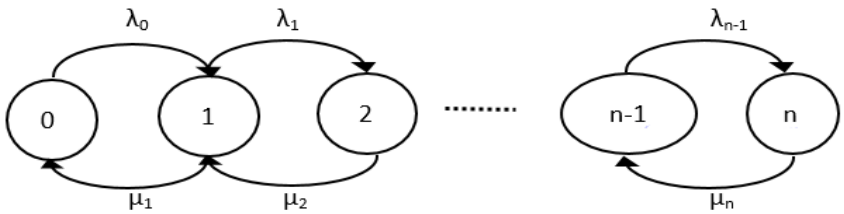

FIGURE 2.3 – Graphe de transition d'un processus de naissance et de mort.

Dans le cas où les λ_i et les μ_i sont strictement positifs, alors tous les états communiquent entre eux et la chaîne est irréductible, pour déterminer sa distribution stationnaire π on résout le système d'équation $\pi Q = 0$, où Q est le générateur infitésimal,

$$\begin{cases} \lambda_0\pi_0 - \mu_1\pi_1 = 0 \\ \lambda_{i-1}\pi_{i-1} - (\lambda_i + \mu_i)\pi_i + \mu_{i+1}\pi_{i+1} = 0 \qquad \forall i = 1, ..., n. \end{cases}$$

2.1.6 Quelque modèles de files d'attente

Les files d'attente de type Markovien sont des cas particuliers très importants de processus de naissance et de mort. L'étude détaillée sera effectuée dans la section suivante.

Modèle d'attente M/M/1

Nous considèrons un système formé d'une file de capacité infinie et d'un unique serveur. La discipline de service de la file est FIFO. Le processus d'arrivée des clients dans la file est

un processus de Poisson de taux λ et le taux de service est distribué selon la loi exponentielle de paramètre μ, (voir Figure 2.4). Ce système est connu sous le nom de file $(M/M/1)$ est un modèle de base le plus élémentaire de la théorie des file d'attente (voir [**9**]).

La file est suggérée comme étant un processus de naissance et de mort pour lequel :

$$\lambda_n = \lambda, \qquad \forall n \geq 0.$$

$$\mu_n = \begin{cases} \mu, & n > 0, \\ 0, & \text{si } n = 0. \end{cases}$$

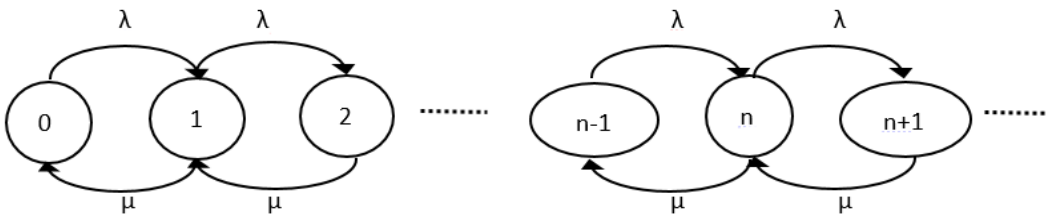

FIGURE 2.4 – Graphe de transition d'une file M/M/1.

Analyse du régime permanent

Nous avons vu au chapitre précédent que la condition de stabilité d'une file simple est $\lambda < \mu$. On note π_n la probabilité stationnaire d'être dans l'état n, ou encore la probabilité pour que le système contienne n clients en régime permanent. Les probabilités π_n Peuvent être calculées par le système d'équations linéiare $\pi Q = 0$ et $\sum_{n=1}^{\infty} \pi_n = 1$, où $\pi = [\pi_0, \pi_1, ...]$ est le vecteur des probabilités stationnaires et Q est le générateur infinitésimal de la chaine de Markov à temps continu(CMTC) :

$$Q = \begin{pmatrix} -\lambda & \lambda & 0 & \dots & & \\ \mu & -(\lambda+\mu) & \lambda & 0 & \dots & \\ 0 & \mu & -(\lambda+\mu) & \lambda & 0 & \dots \\ \vdots & 0 & \mu & -(\lambda+\mu) & \lambda & 0 \\ & \vdots & 0 & \mu & -(\lambda+\mu) & \lambda \\ & & \vdots & 0 & \mu & \ddots \end{pmatrix}$$

A l'aide de l'équation linéaire $\pi Q = 0$, on obtient les équations dites de balances locales :

$$\begin{cases} \pi_0 \lambda & = & \pi_1 \mu \\ \pi_1 (\lambda + \mu) & = & \pi_0 \lambda + \pi_2 \mu \\ \vdots & & \\ \pi_n (\lambda + \mu) & = & \pi_{n-1} \lambda + \pi_{n+1} \mu, \qquad \text{pour tout } n \geq 1 \end{cases}$$

On resoudre le système on trouve :

$$\begin{cases} \pi_1 = \dfrac{\lambda}{\mu}\pi_0 \\ \pi_2 = \left(\dfrac{\lambda}{\mu}\right)^2 \pi_0, & \text{pour } n = 1 \\ \pi_n = \left(\dfrac{\lambda}{\mu}\right)^n \pi_0, & \text{pour } n > 1 \end{cases}$$

Pour achever le calcul, on utilise la condition de normalisation de probabilités, en remplacant P_n par sa valeure calculée précédement (voir **[9]**), nous permet d'obtenir :

$$\pi_0 = \frac{1}{\sum_{n=0}^{\infty} \left(\dfrac{\lambda}{\mu}\right)^n}. \tag{2.5}$$

En notant $\rho = \dfrac{\lambda}{\mu}$, alors $\pi_0 = 1 - \rho$
à condition que la série converge, ce qui est vrai si $\rho < 1$. Dans ce cas :

$$\pi_n = (1-\rho)\rho^n \qquad \text{pour tout } n \geq 0$$

Paramètres de performances

Débit :

Noté D représente la probabilité que le système contient au moins un client

$$D = \mathbb{P}([\text{file non vide}])\mu = \sum_{n=1}^{\infty} \pi_n \mu = (1-\pi_0)\mu = \rho\mu = \lambda.$$

Nombre moyen de clients :

Le nombre moyen de clients $\overline{L}$ se calcule à partir des probabilités stationnaires :

$$\begin{aligned} \overline{L} &= \sum_{n=1}^{\infty} n\pi_n \\ &= (1-\rho)\sum_{n=1}^{\infty} n\rho^n \\ &= \rho(1-\rho)\frac{\partial}{\partial\rho}(1+\rho^2+\rho^3...) \\ &= \rho(1-\rho)\frac{\partial}{\partial\rho}\left(\frac{1}{1-\rho}\right). \end{aligned}$$

Soit,

$$\overline{L} = \frac{\rho}{1-\rho}. \tag{2.6}$$

Temps moyen de séjour :

Ce parèmtre noté $\overline{W_s}$ est obtenu en utilisant la loi de Little :

$$\overline{W_s} = \frac{\overline{L}}{D} = \frac{1}{\mu(1-\rho)}. \tag{2.7}$$

Proposition 2.1.4 [69]

Soit $\{X_t\}$ le nombre de clients à l'instant t dans une file $M/M/1$ pour laquelle les temps séparant deux arrivées sont exponentiels de paramètre λ et les temps de service sont exponentiels de paramètre μ. Le processus $\{X_t, t \geq 0\}$ est dit :

- *transiant si $\lambda > \mu$,*
- *récurrent positif si $\lambda < \mu$,*
- *récurrent nul si $\lambda = \mu$.*

Exemple 2.1.3 *Simulation on considére une file d'attente $M/M/1$,de taux de service vaut 1, et les taux d'arrivée valent respectivement 1.05, 0.95 et 1. La Figure 2.5 réprésente la simulation de cette file dans les trois cas transient, récurrent positif et récurrent nul.*

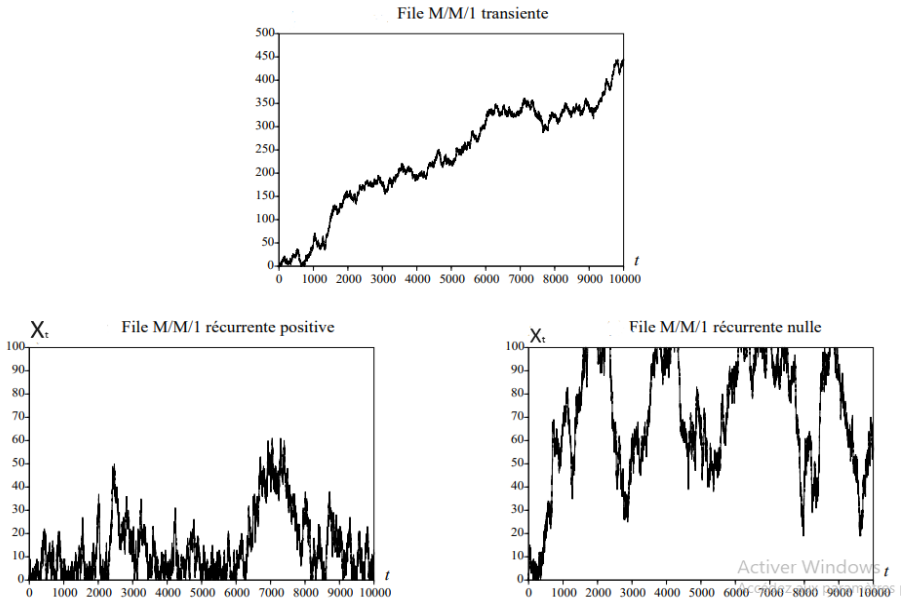

FIGURE 2.5 – Simulation File M/M/1.

Modèle d'attente M/M/1/K

Dans un file d'attente M/M/1 avec une capacité finie, Quand un client arrive alors qu'il y a déjà K clients présents dans le système, On a toujours les hypothèses : λ le taux du processus de Poisson pour les arrivées et μ le paramètre de la distribution exponentielle pour les temps de service. Soit K la capacité de la file d'attente : c'est le nombre maximal

de clients qui peuvent être présents dans le système, soit en attente, soit en service. Ce processus est considéré comme un processus de naissance et de mort avec :
– un taux de naissance $\lambda_n = \lambda$, pour tout $n < K$,
– un taux de mort $\mu_n = \mu$, pour $n \neq 0$.

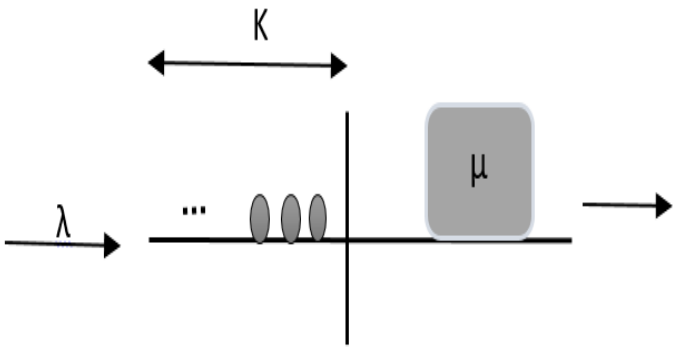

FIGURE 2.6 – Repésentation schématique d'une file M/M/1/K.

Analyse du régime permanent

Soit $\pi_n, n = 0, 1..., K$, la probabilité stationnaire que le système contienne n clients. Ces probabilités peuvent être calculées en écrivant les équations d'équilibre du système :

$$\begin{cases} \pi_0\lambda &= \pi_1\mu \\ \pi_{n-1}\lambda &= \pi_n\mu, \qquad \text{pour tout } n = 1, 2, ..., K \end{cases}$$

En utilisant la condition de normalisation, on obtient :

$$\pi_0 = \frac{1}{\sum_{n=0}^{K} \rho^n} = \frac{1-\rho}{1-\rho^{K+1}}.$$

$$\pi_n = \frac{(1-\rho)\rho^n}{1-\rho^{K+1}}. \tag{2.8}$$

Paramètres de performances

Débit :

$$D = \mathbb{P}([\text{file non vide}])\mu = \sum_{n=1}^{K} \pi_n\mu = (1-\pi_0)\mu = \frac{\rho-\rho^{K+1}}{1-\rho^{K+1}}\mu.$$

Modèle d'attente M/M/c

Pour ce modèle de file d'attente, on considère un système de c serveurs indentiques et indépedants les uns des autres, et une salle d'attente de capacité infinie. Le processus d'arrivée des clients poissonien de taux λ et temps de service esponentiel de taux μ. le processus modélisant le nombre de clients dans le système est un processus de naissance et de la mort avec :

$$\lambda_n = \lambda,$$
$$\mu_n = \begin{cases} n\mu, & \text{pour } n = 1, 2, ..., c-1 \\ c\mu, & \text{sinon } n \geq c \end{cases}$$

Voir Figure 2.7.

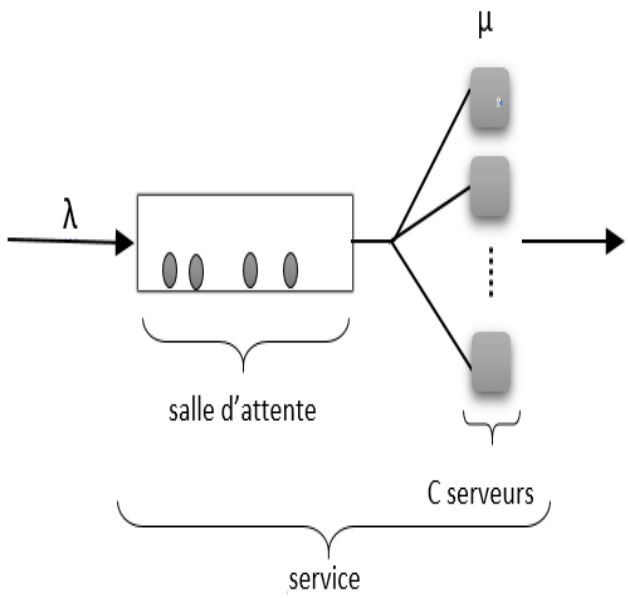

FIGURE 2.7 – Repésentation schématique d'une file M/M/c.

Analyse du régime permanent

La condition de stabilité d'une file comportant c serveurs est $\lambda < c\mu$ et exprime le fait que le nombre moyen de clients qui arrivent à la file par unité de temps doit être inférieur au nombre moyen de client que les serveurs de la file sont capables de traiter par unité de temps, dans ce cas on peut calculer la probabilité stationnaire π_n pour que le système contienne n clients (voir [**9**]). Alors

$$\pi_n = \begin{cases} \dfrac{\rho^n}{n!}\pi_0, & \text{pour } 1 \leq n \leq c-1 \\ \dfrac{\rho^n}{c!c^{n-c}}\pi_0, & \text{pour } n \geq c \end{cases}$$

Avec

$$\pi_0 = \left[\sum_{n=0}^{c-1}\frac{\rho^n}{n!}+\left(\frac{\rho^c}{c!}\right)\left(\frac{1}{1-\frac{\rho}{c}}\right)\right]^{-1}.$$

Lorsque $c=1$, on retrouve bien les résultat de la file $M/M/1$

Paramètres de performances

Débit :

Le service s'effectue avec un taux $n\mu$ quqnd le système comtient moins de c clients, et avec un taux $c\mu$ dans chaque état où le système contient plus de c client :

$$D=\sum_{n=1}^{c-1}\pi_n n\mu+\sum_{n=c}^{\infty}\pi_n c\mu.$$

Nombre moyen de clients :

$\overline{L_q}$ le nombre moyen de clients en attente dans la file,

$$\begin{aligned}\overline{L_q} &= \sum_{n=c}^{\infty}(n-c)\pi_n \\ &= \frac{\rho^{c+1}}{(c-1)!(c-\rho)^2}\pi_0.\end{aligned}$$

$\overline{L_s}$ le nombre moyen de clients dans le système,

$$\overline{L_s} = \frac{\rho^{c+1}}{(c-1)!(c-\rho)^2}\pi_0+\rho.$$

Temps moyen de séjour :

$$\overline{W_s} = \frac{\rho^c}{\mu(c-1)!(c-\rho)^2}\pi_0+\frac{1}{\mu}.$$

Modèle d'attente $M/M/\infty$

Ce système composé d'un nombre illimité de serveurs identiques et indépendants les uns des autres. Dés qu'un client arrive, il rentre donc instantanément en service. Dans celle file les clients arrivent d'un processus de poisson de taux λ et les temps de service sont exponentiels de taux μ. C'est un processus de naissance et de mort avec :

$$\begin{cases}\lambda_k=\lambda \\ \mu_k=k\mu \qquad \text{pour } k=0,1,2...\end{cases}$$

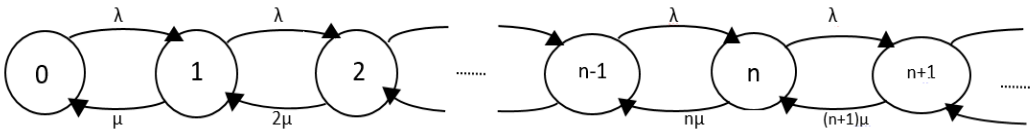

FIGURE 2.8 – Repésentation schématique d'une file $M/M/\infty$.

Analyse du régime permanent

Pour π_n la probabilité stationnaire d'être dans l'état n. on peut déduire les équations de frontière :

$$\begin{cases} \pi_0\lambda &= \pi_1\mu \\ \pi_{n-1}\lambda &= \pi_n n\mu, \qquad \text{pour } n = 1, 2, ... \end{cases}$$

La condition de normalisation nous donne alors immédiatement π_0 :

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

$$\pi_n = \frac{\rho^n}{n!} e^{-\rho} \qquad \text{pour } n = 1, 2, ... \tag{2.9}$$

Paramètres de performances

Débit :

Le service s'éffectue avec un taux $n$dans chaque état où le système contient n clients :

$$d = \sum_{n=1}^{\infty} \pi_n n\mu = \mu e^{-\rho} \sum_{n=1}^{\infty} \frac{\rho^n}{(n-1)!} = \mu e^{-\rho} \rho e^{\rho} = \lambda.$$

Nombre moyen de clients :

$$\overline{L} = \sum_{n=1}^{\infty} n\pi_n = e^{-\rho} \sum_{n=1}^{\infty} \frac{\rho^n}{(n-1)!} = e^{-\rho} \rho e^{\rho} = \rho.$$

Temps moyen de séjour :

$$\overline{W}_s = \frac{\overline{L}}{d} = \frac{\rho}{\lambda} = \frac{1}{\mu}.$$

2.2 Files d'attentes non markoviennes

Les files d'attente sont supposées non markoviennes si le temps des interarrivées ou la durée de services ne suit pas la loi exponentielle. Ce facteur rend l'étude des systèmes plus complexe et délicate. Pour éliminer cet impact nous faisons intervenir les méthodes citées dans Abbas [**1**] :

- Méthode des étapes d'Erlang : Cette métode consiste à approximer toute loi de probabilité qui possède une transformation de Laplace rationnelle par une loi de Cox qui possède la propriété d'absence de mémoire par étape.
- Méthode de la chaine de Markov induite : Son principe est de choisir une suite d'instants t_i pour tout $i = 1, 2, ..., k$ tels que la chaîne induite $X_{t_k}, t_k \geq 0$ est une chaine de Markov homogène.
- Méthode d'approximation :Nous caractérisons l'état du systme étudié par :
  - Des méthodes asymptotiques décrivant l'état du système.
  - L'éstimation par bornes de certaines de ces caractéristiques.
- Simulation : La simulation est une technique de modélisation. Elle permet de présenter le fonctionnement d'un système composé de différents centres d'activité, de mettre en évidence les caractéristiques et de d'écrire la circulation des différents objets traités par ces processus et enfin observer le comportement du système.

2.3 Modélisation par les martingales

Les martingales sont devenues une approche principale en théorie de file d'attente. Elle simule d'une part l'aléatoire du phénomène mais aussi son évolution dans le temps. L'utilité de cette approche est consacré plutot à l'analyse des problèmes plus géeraux. Dans cette section nous nous intéressons aux martingales à temps discrêt.

Les martingales sont souvent associées à des stratégies secrètes et énigmatiques visant à remporter des jeux de hasard. Par exemple nous désignons par Y_n la fortune d'un joueur après la $n^{\text{ème}}$ partie et F_n représente son état à propos du jeu à ce moment là. L'égalité

$$\mathbb{E}(Y_{n+1}/F_n) = Y_n$$

nous informe que la fortune espérée après la prochaine partie est la même que sa fortune actuelle. Une martingale est ainsi un jeu équitable. Pour chaque étape n, le joueur gagne une somme φ_n si $\varphi_n \geq 0$ ou perd une somme $-\varphi_n$ si $\varphi_n < 0$. La fortune du joueur au temps n sera alors notée.

$$\forall n \geq 0,\ Y_n = Y_0 + \sum_{i=1}^{n} \varphi_i.$$

où Y_0 représente la fortune initiale.

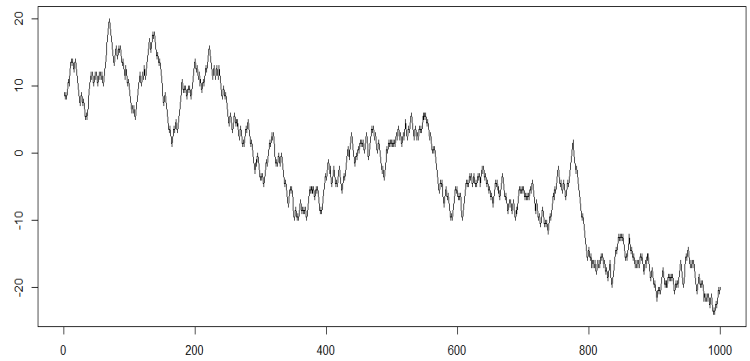

FIGURE 2.9 – Le graphe représente une marche alétoire.

La Fgure 2.9 représent des réalisations de la marche aléatoire $\{Y_n\}_{n\geq 0}$ décrivant la fortune d'un joueur après 1000 pas en partant initialement de $Y_0 = 10$ (voir [**12**]).

2.3.1 Quelques définitions et concepts de base

Soit $(\Omega, F, \mathbb{P})$ un espace probabilisé

a. Filtration

Nous appelons une filtration $\{F_n\}_{n\geq 0}$ sur $(\Omega, F, \mathbb{P})$ toute suite croissante de sous tribus de F,pour tout $n \in \mathbb{N}$.

Nous disons qu'un processus $\{Y_n\}_{n\in\mathbb{N}}$ est adapté à une filtration $\{F_n\}_{n\geq 0}$, si $\forall n \in \mathbb{N}$:Y_n est F_n mesurable.

Si $\{Y_n\}_{n\in\mathbb{N}}$ est un processus stochastique défini sur $(\Omega, F, \mathbb{P})$ alors$F_n^Y = \sigma(Y_0, ..., Y_n)$ est la filtration naturelle du processus $\{Y_n\}_{n\geq 0}$.

b. Les martingales

Nous disons qu'une suite de variables aléatoires $Y = \{Y_n\}_{n\in\mathbb{N}}$ est une F_n-martingale si pour tout n :

- La suite de variables aléatoires $\{Y_n\}_{n\in\mathbb{N}}$ est adaptée à filtration $(F_n)_{n\geq 0}$.
- (Y_n) est integrable ($\mathbb{E}(|Y_n|) < \infty$).
- $\forall n, \mathbb{E}(Y_{n+1}/F_n) = Y_n$, p.s.

Dans le cas où $\mathbb{E}(Y_{n+1}/F_n) \leq Y_n$ p.s, nous obtenons une sur-martingale et si $\mathbb{E}(Y_{n+1}/F_n) \geq Y_n$, nous obtenons une sous-martingale.

On dit qu'un processus Y est une martingale s'il est à la fois surmartingale et sous-martingale.

c. Temps d'arrêt

La notion de temps d'arrêt joue un rôle central dans l'analyse des processus aléatoires. C'est la vraie notion de temps aussi bien pour les développements mathématiques que pour la modélisation.

Une variable aléatoire $T : \Omega \to \mathbb{N} \cup \{\infty\}$ est dite un temps d'arrêt si $(T \leq n)$ est (F_n)-mesurable.

2.3.2 Martingales fermées

Définition 2.3.1 [10] *Un processus aléatoire $X = (X_n)_{n\geq 0}$ est une martingale fermée s'il existe une v.a. réeelle intégrable Y telle que $X_n = \mathbb{E}[Y|F_n]$ pour tout $n \in \mathbb{N}$.*

Théoréme 2.3.1 *Toute martingale fermée est uniformément intégrable*

Théoréme 2.3.2 (Théorème d'arrêt de Doob) [61] *Si le processus M_n est une martingale integrable et T est un temps d'arrêt régulier alors pour tout couple de temps d'arrêt T_1 et T_2 tel que $T_1 \leq T_2 \leq T$ presque sûrement les variables aléatoires X_{T_1} et X_{T_2} existent, sont intégrales et vérifient $X_{T_1} = \mathbb{E}[X_{T_2}/F_{T_1}]$.*

Pour construire une martingale à partir d'une chaîne de Markov $(X_n)_{n\geq 0}$ de matrice de transition P sur un espace d'états E dénombrable (voir **[24]**, **[12]**). Une fonction harmonique h pour P

$$\forall i \in E,\ h(i) = \sum_{j\in E} P_{ij}h(j) = \mathbb{E}[h(X_{n+1})|X_n = i].$$

Si $\mathbb{E}(\|h(X_n)\|) < \infty$ pour tout n, alors $\{h(X_n)\}_{n\geq 0}$ est une martingale pour la filtration F_n. et on à :

- si $Ph \geq h$, i.e $\{h(X_n)\}_{n\geq 0}$ est une sous-martingale,
- si $Ph \leq h$, i.e $\{h(X_n)\}_{n\geq 0}$ est une surmartingale.

2.3.3 Analyse du système $M/G/1$ par la méthode des martingale

Considérons le système de file d'attente $M/G/1$ (voir **[60]**). Le processus d'arrivée des clients, noté $(N_t)_{t\geq 1}$ suit un processus de Poisson de paramètre λ. Les temps de service, indépendants les uns des autres et du processus d'arrivée sont des variables aléatoires identiquement distribuées. Nous notons $B(\cdot)$ la distribution des temps de service et $B^*(\cdot)$ sa transformée de Laplace.

On note X_n le nombre de clients en attente (ou en train de se faire servir) juste aprèes le $n^{\text{ème}}$ départ.

$$X_{n+1} = X_n + N_{n+1} - \mathbb{I}_{\{X_n\neq 0\}} \qquad \forall n \in \mathbb{N} \tag{2.10}$$

où $\mathbb{I}_{\{X_n\neq 0\}}$ est la fonction indicatrice de l'évènement E.

Le processus $(N_t)_{t\geq 1}$ suit une loi de Poisson de paramètre λt, alors

$$\mathbb{P}(T_n = k) = \int_0^\infty \frac{(\lambda t)^k}{k!} e^{-\lambda t} dB(t) \tag{2.11}$$

En introduisant la fonction génératrice

$$\alpha(z) = \sum_{k=0}^{\infty} z^k \mathbb{P}(T_n = k) \tag{2.12}$$

pour $\mathbb{P}(T_n = k) = \alpha_k$, alors

$$\forall k \in \mathbb{N}, \forall 0 < \alpha_k \leq 1, \forall 0 < z \leq 1 : |\alpha_k z^k| \leq |z^k| \leq \infty$$

donc la série $\alpha(z) = \sum_{k=0}^{\infty} z^k \alpha_k$ converge.

$$\begin{aligned}
\alpha(z) &= \sum_{k=0}^{\infty} z^k \int_0^{\infty} \frac{(\lambda t)^k}{k!} e^{-\lambda t} dB(t) \\
&= \int_0^{\infty} e^{-\lambda t} \sum_{k=0}^{\infty} z^k \frac{(\lambda t)^k}{k!} dB(t) \\
&= \int_0^{\infty} e^{-\lambda t} e^{\lambda t z} dB(t) \\
&= \int_0^{\infty} e^{-\lambda t(1-z)} dB(t) \\
&= B^*(\lambda(1-z)).
\end{aligned}$$

Chapitre 3

La stabilité stochastique

Dans ce chapitre, nous allons présenter l'une des méthodes les plus importantes utilisées dans l'étude de la stabilité des modèles de file attente.

La méthode de Chapman-Kolmogorov (voir [**56**], [**44**], [**69**]) basée sur les équations fondamentales de la théorie des probabilités offre une approche analytique pour étudier la dynamique des systèmes de file d'attente. Nous examinerons comment cette méthode permet d'analyser les probabilités de transition entre les états du système et comment elle peut être utilisée pour évaluer la stabilité d' un processus de naissance et de mort.

3.1 Equations de Chapman-Kolmogorov

Dans une file d'attente markovienne, les événements futurs dépendent uniquement de l'état présent du système, ce qui rend la méthode de Chapman-Kolmogorov particulièrement adaptée pour étudier les transitions probabilistes entre les états de la file d'attente et prédire son comportement à long terme. On utilise cette méthode pour évaluer la stabilité du système, aprés nous prenons un exemple d' un modèle classique de file d'attente markovienne.

Soit $q_{ij}(t)$ le taux auquel les transitions se produisent de l'état i à l'état j au temps t.

$$q_{ij}(t) = \lim_{h \to 0} \frac{p_{ij}(t, t+h)}{h} \qquad \text{pour } i \neq j$$

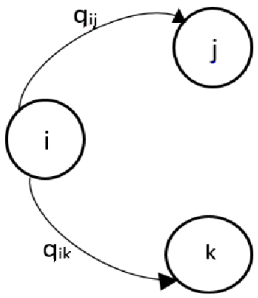

FIGURE 3.1 – La transition de l'état i à l'instant t vert un autre éta j ou bien k à l'instant $t+h$.

Et

$$p_{ij}(t, t+h) = q_{ij}(t)h + o(h) \qquad \text{pour } i \neq j \tag{3.1}$$

Cela signifie simplement que en termes d'ordre $o(h)$, la probabilité qu'une transition se produise de l'état i au temps t à l'état j dans les prochaines h unités de temps est égale au taux de transition au temps t multiplié par la durée de la période de temps h. En appliquant le principe de conservation de la probabilité et l'équation (3.1), nous trouvons

$$1 - p_{ii}(t, t+h) = \sum_{j \neq i} p_{ij}(t, t+h) \tag{3.2}$$

$$= \sum_{j \neq i} q_{ij}(t)h + o(h). \tag{3.3}$$

avec

$$\lim_{h \to 0} \frac{o(h)}{h} = 0$$

En divisant par h et en prenant la limite lorsque $h \to 0$, nous obtenons

$$\lim_{h \to 0} \left(\frac{1 - p_{ii}(t, t+h)}{h} \right) = \lim_{h \to 0} \left(\frac{\sum_{j \neq i} q_{ij}(t)h + o(h)}{h} \right) = \sum_{j \neq i} q_{ij}(t).$$

Pour les processus de Markov , le taux de transition correspondant au système restant en place est défini par l'équation

$$q_{ii}(t) = -\sum_{j \neq i} q_{ij}(t). \tag{3.4}$$

Remarque 3.1.1 *Si le taux de transition dans une chaîne de Markov est non homogène cela veut dire que la probabilité de transition peut dépendre du temps t, on peut consulter* **[56]** *pour comprendre le concept.*

Lorsque l'état i est un état absorbant alors $q_{ii}(t) = 0$. Le fait que $q_{ii}(t)$ soit négatif ne devrait pas être surprenant. Cette quantité représente un taux de transition . Étant donné que le système se trouve dans l'état i au temps t, la probabilité qu'il passe à un état différent j augmente avec le temps, tandis que la probabilité qu'il reste dans l'état i doit diminuer avec le temps. Il est approprié dans le premier cas que la dérivée au temps t soit positive et dans le second cas qu'elle soit négative. La substitution de l'équation (3.4) dans l'équation (3.2) fournit l'analogie de l'équation (3.1). Pour une chaîne de Markov a temps continu homogène on a :

$$q_{ij} = \lim_{h\to 0}\left(\frac{p_{ij}(h)}{h}\right) \tag{3.5}$$

$$q_{jj} = \lim_{h\to 0}\left(\frac{p_{jj}(h)-1}{h}\right) \tag{3.6}$$

Soit $\{X_t, t \geq 0\}$ une chaîne de Markov à temps continu homogène et supposons qu'au temps $t = 0$, la chaîne de Markov soit dans l'état non absorbant i. Soit T_i la variable aléatoire qui décrit le temps jusqu'à ce qu'une transition hors de l'état i se produise. Alors

$$\begin{aligned}
\mathbb{P}\{T_i > s+t|X_0 = i\} &= \mathbb{P}\{T_i > s+t|X_0 = i, T_i > s\}\mathbb{P}\{T_i > s|X_0 = i\} && (3.7)\\
&= \mathbb{P}\{T_i > s+t|X_s = i\}\mathbb{P}\{T_i > s|X_0 = i\} && (3.8)\\
&= \mathbb{P}\{T_i > t|X_0 = i\}\mathbb{P}\{T_i > s|X_0 = i\} && (3.9)
\end{aligned}$$

On suppose $H'(t) = \mathbb{P}\{T_i > t|X_0 = i\}, t > 0$, alors l'équation (3.7) devient

$$H'(s+t) = H'(t)H'(s)$$

Cette équation est satisfaite si et seulement si $H'(t) = e^{-\mu_i t}$ pour un paramètre positif $\mu_i > 0$ et $t > 0$. Ainsi, le temps de séjour dans l'état i doit être distribué de manière exponentielle. L'équation (3.7) offre une perspective différente à ce sujet. Si l'on sait que la chaîne de Markov a démarré au temps $t = 0$ dans l'état i et n'a pas quitté l'état i jusqu'au temps s, c'est-à-dire $\mathbb{P}\{T_i > s|X_0 = i\} = 1$, alors

$$\mathbb{P}\{T_i > s+t|T_i > s\} = \mathbb{P}\{T_i > t\}$$

ainsi la variable aléatoire continue T_i est sans mémoire.

Les équations de Chapman-Kolmogorov pour une $CMTC$ non homogène $\{X_t, t \geq 0\}$ peuvent être obtenues directement à partir de la propriété de Markov **[56]**. Elles sont définies par

$$p_{ij}(s,t) = \sum_k p_{ik}(s,u)p_{kj}(u,t) \qquad \text{pour } i,k,j = 0,1,... \text{ et } s \leq u \leq t$$

En passant de l'état i à l'instant s à l'état j à l'instant t avec $(s < t)$, nous devons passer par un état intermédiaire k à un moment intermédiaire u. Lorsque la chaîne de Markov en temps continu est homogène, l'équation de Chapman-Kolmogorov peut être écrite comme suit :

$$p_{ij}(t+h) = \sum_k p_{ik}(t)p_{kj}(h) \qquad \text{pour } t,h \geq 0 \tag{3.10}$$

$$= \sum_{k\neq j} p_{ik}(t)p_{kj}(h) + p_{ij}(t)p_{jj}(h). \tag{3.11}$$

Ainsi, en divisant par h, nous obtenons

$$\frac{p_{ij}(t+h)-p_{ij}(t)}{h} = \sum_{k\neq j} p_{ik}(t)\frac{p_{kj}(h)}{h} + p_{ij}(t)\frac{p_{jj}(h)}{h} - \frac{p_{ij}(t)}{h}$$

$$= \sum_{k\neq j} p_{ik}(t)\frac{p_{kj}(h)}{h} + p_{ij}(t)\frac{p_{jj}(h)-1}{h}.$$

Lorsque h tend vers 0 et d'après les équation (3.5), (3.6) nous obtenons l'équation différentielle suivante :

$$\frac{\partial p_{ij}(t)}{\partial t} = \sum_{k\neq j} p_{ik}(t)q_{kj} + p_{ij}(t)q_{jj}.$$

En d'autres termes :

$$\frac{\partial p_{ij}(t)}{\partial t} = \sum_k p_{ik}(t)q_{kj} \qquad \text{pour } i,k,j = 0,1,...$$

Ces équations sont appelées les équations de Kolmogorov. En forme matricielle, elles sont écrites comme suit :

$$\frac{\partial p(t)}{\partial t} = p(t)Q. \tag{3.12}$$

La matrice $Q(t)$ dont l'élément $ij^{\text{ème}}$ est $q_{ij}(t)$ est appelée le générateur infinitésimal ou la matrice de taux de transition pour la chaîne de Markov continue. En forme matricielle, cela se présente comme suit :

$$Q(t) = \lim_{h\to 0}\left(\frac{\mathbb{P}(t,t+h)-I}{h}\right)$$

Lorsque la chaîne de Markov en temps continu est homogène, les taux de transition q_{ij} sont indépendants du temps, et la matrice des taux de transition est simplement écrite comme Q. De manière similaire, en écrivant l'équation (3.10) sous la forme

$$p_{ij}(t+h) = \sum_k p_{ik}(h)p_{kj}(t) \qquad \text{pour } h \geq 0$$

Nous pouvons dériver de manière similaire les équations de Kolmogorov, qui sont

$$\frac{\partial p_{ij}(t)}{\partial t} = \sum_k q_{ik}(t)p_{kj}(t) \qquad \text{pour } i,j = 0,1,...$$

Ou sous forme matricielle,

$$\frac{\partial P(t)}{\partial t} = QP(t)$$

La solution des équations de Kolmogorov (3.12) est donnée par l'exponentielle matricielle.

$$P(t) = P(0)e^{Qt} = ce^{Qt} = e^{Qt} = \left(I + \sum_{n=1}^{\infty} \frac{Q^n t^n}{n!} \right) \tag{3.13}$$

avec une constante d'intégration $c = P(0) = I$.
Les équations de Kolmogorov vers l'avant est de la forme

$$\frac{\partial P(t)}{\partial t} = P(t)Q(t)$$

et vers l'arrière sont données par

$$\frac{\partial P(t)}{\partial t} = Q(t)P(t)$$

3.1.1 Analyser la stabilité d'un processus de naissance et de mort

Nous prenons une description des mouvements possibles d'un processus de naissance et de mort sur un petit intervalle de temps (voir **[69]**, **[56]**).
Nous considèrons un processus de naissance et de mort $\{X_t, t \geq 0\}$ de taux de naissance $(\lambda_n)_{n\in\mathbb{N}}$ et un taux de mort $(\mu_n)_{n\in\mathbb{N}^*}$. Supposons que le processus à l'instant t se trouve dans l'état n. Le prochain saut sera vert $n-1$, où $n+1$, (voir la Figure 3.2). Le taux net de flux de probabilité dans n est trouvé en calculant le flux à travers cette frontière, en utilisant des signes opposés pour l'entrée et la sortie.

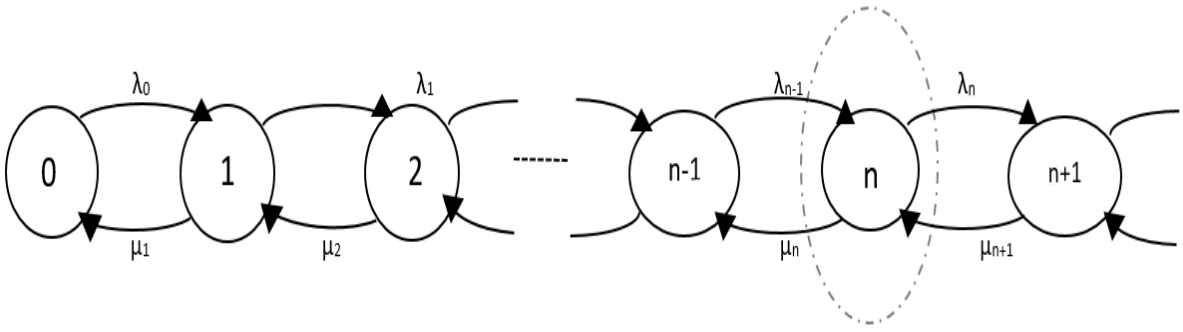

FIGURE 3.2 – Diagramme de transition d'un processus de naissance et de mort a l'instant t et $t+1$ vers l'état n.

À partir d'équation (3.1), on définit les probabilités de localisation de X_{t+h}, sont donnés :

$$\begin{aligned}
\mathbb{P}[X_{t+h} = n+1 | X_t = n] &= h\lambda_n + o(h) \\
\mathbb{P}[X_{t+h} = n-1 | X_t = n] &= h\mu_n + o(h) \\
\mathbb{P}[X_{t+h} = n | X_t = n] &= 1 - h(\lambda_n + \mu_n) + o(h)
\end{aligned}$$

Ces hypothèses signifient que la probabilité de deux naissances ou plus, de deux morts ou plus, ou de naissances et de morts presque simultanées dans un petit intervalle de temps h est négligeable, c'est-à-dire de l'ordre de $o(h)$. On déduit une relation entre la loi de X_{t+h} et la loi de X_t

$$\mathbb{P}[X_{t+h} = n] = h\lambda_{n-1}\mathbb{P}[X_t = n-1] + h\mu_{n+1}\mathbb{P}[X_t = n+1] + (1 - h(\lambda_n + \mu_n))\mathbb{P}[X_t = n] + o(h)$$

$$\begin{aligned}
\frac{1}{h}(\mathbb{P}[X_{t+h} = n] - \mathbb{P}[X_t = n]) &= \lambda_{n-1}\mathbb{P}[X_t = n-1] + \mu_{n+1}\mathbb{P}[X_t = n+1] \\
&\quad -(\lambda_n + \mu_n)\mathbb{P}[X_t = n] + \frac{o(h)}{h}
\end{aligned}$$

Posons $p_n(t) = \mathbb{P}[X_t = n]$, cette derniere équation devient

$$\frac{1}{h}(p_n(t+h) - p_n(t)) = \lambda_{n-1}p_{n-1}(t) + \mu_{n+1}p_{n+1}(t) - (\lambda_n + \mu_n))p_n(t) + \frac{o(h)}{h}$$

Lorsque h tend vers 0, on obtient léquation différentielle suivante.

$$p_n^{'}(t) = \lambda_{n-1}p_{n-1}(t) - (\lambda_n + \mu_n)p_n(t) + \mu_{n+1}p_{n+1}(t)$$

Pour l'état $n = 0$, l'équation est légèrment defférente.

$$p_0^{'}(t) = -\lambda_0 p_0(t) + \mu_1 p_{1(t)}$$

On a obtenu un système d'équations différentielles sous formes des équations directes de Chapman-Kolmogorov.

$$\begin{cases} p_0^{'}(t) = -\lambda_0 p_0(t) + \mu_1 p_{1(t)}, \\ p_n^{'}(t) = \lambda_{n-1}p_{n-1}(t) - (\lambda_n + \mu_n)p_n(t) + \mu_{n+1}p_{n+1}(t), \qquad \forall n \geq 1. \end{cases}$$

Si la loi de probabilité initiale $(p_n(0))_{n\in\mathbb{N}}$ est donnée, le système de Chapman Kolmogorov admet une unique solution $(p_n(t))_{n\in\mathbb{N}}$.
Une mesure stationnaire est une loi de probabilité $\pi = (\pi)_{n\in\mathbb{N}}$ telle que si la loi de X_0 est π, alors la loi de X_t reste π pour tout $t > 0$. Une telle mesure est donc nécessairement une solution constante du système de Chapman-Kolmogorov :

$$(S) \begin{cases} 0 = -\lambda_0 \pi_0(t) + \mu_1 \pi_{1(t)} \\ 0 = \lambda_{n-1}\pi_{n-1}(t) - (\lambda_n + \mu_n)\pi_n(t) + \mu_{n+1}\pi_{n+1}(t), \qquad \forall n \geq 1. \end{cases}$$

Il est facile de résoudre le système (S) pour tout $n \geq 1$,

$$\pi_{n+1} = \frac{\lambda_n + \mu_n}{\mu_{n+1}}\pi_n - \frac{\lambda_{n-1}}{\mu_{n+1}}\pi_{n-1},$$

et

$$\pi_1 = \frac{\lambda_0}{\mu_1}\pi_0.$$

Nous obtenons par itération,

$$\pi_2 = \frac{\lambda_1 + \mu_1}{\mu_2}\pi_1 - \frac{\lambda_0}{\mu_2}\pi_0 = \frac{\lambda_1 + \mu_1}{\mu_2}\left(\frac{\lambda_0}{\mu_1}\pi_0\right) - \frac{\lambda_0}{\mu_2}\pi_0 = \frac{\lambda_1\lambda_0}{\mu_2\mu_1}\pi_0.$$

De même.

$$\pi_3 = \frac{\lambda_2\lambda_1\lambda_0}{\mu_3\mu_2\mu_1}\pi_0.$$

On en déduit immédiatement l'expression de π_n en fonction de π_0.

$$\pi_n = \frac{\lambda_{n-1}\lambda_{n-2}...\lambda_0}{\mu_n\mu_{n-1}...\mu_1}\pi_0 = \pi_0\prod_{i=1}^{n}\frac{\lambda_{i-1}}{\mu_i}, \qquad n \geq 1$$

Le système (S) admet donc toujours une solution unique à un coefficient multiplicatif près. Mais pour que ce soit une loi de probabilité, il est nécessaire que la somme des coefficients soit égale à 1. C'est le cas si et seulement si la série de terme général $\prod_{i=1}^{n}\frac{\lambda_{i-1}}{\mu_i}$ converge.

On suppose $U_n = \prod_{i=1}^{n}\frac{\lambda_{i-1}}{\mu_i}$, alors $\pi_n = \pi_0 U_n$.

$(\pi_n)_{n\geq 0}$ est une loi de probabilité, alorsi on a :

$$\begin{aligned}\sum_{n=0}^{\infty}\pi_n = 1 \quad &\Rightarrow \quad \pi_0 + \pi_0\sum_{n=1}^{\infty}U_n = 1\\ &\Rightarrow \quad \pi_0(1 + \sum_{n=1}^{\infty}U_n) = 1\\ &\Rightarrow \quad \pi_0 = \frac{1}{1 + \sum_{n=1}^{\infty}U_n}.\end{aligned}$$

π_0 est défini s'il existe un entier k constant tel que

$$\frac{\lambda_k}{\mu_{k+1}} < 1$$

$$\sum_{n=1}^{\infty} U_n < \infty (converge).$$

Dans ce cas, il existe un point dans l'espace d'états tel que le taux d'arrivées dans tout état à partir de ce point ou au-delà est inférieur au taux de départs de cet état.

3.1.2 La file M/M/1

Prenons comme exemple une file d'attente $M/M/1$ traité par [**56**], avec

$$\begin{cases} \lambda_n = \lambda \\ \mu_n = \mu \end{cases}$$

et $\dfrac{\lambda}{\mu} < 1$ Nous trouvons :

$$1 + \sum_{n=1}^{\infty} U_n = 1 + \sum_{n=1}^{\infty} \prod_{i=1}^{n} \frac{\lambda_{i-1}}{\mu_i} = 1 + \sum_{n=1}^{\infty} \prod_{i=1}^{n} \frac{\lambda}{\mu} = 1 + \sum_{n=1}^{\infty} \left(\frac{\lambda}{\mu}\right)^n = \sum_{n=0}^{\infty} \rho^n = \frac{1}{1-\rho} < \infty$$

ce qui montre que la file d'attente $M/M/1$ est stable lorsque $\rho < 1$. Remarquez également que

$$\beta = \sum_{n=0}^{\infty} \left(\frac{1}{\lambda_n U_n}\right) = \frac{1}{\lambda} \sum_{n=0}^{\infty} \frac{1}{\rho^n} = \infty$$

3.1.3 Résolution des équations Chapman-Kolmogorov

Exemple 3.1.1 *[Processus de Markov à deux états] On suppose que $(X_t)_{t\geq 0}$ est une chaîne de Markov homogène à valeurs dans $E = \{e_1, e_2\}$, de générateur $Q = \begin{pmatrix} -\lambda & \lambda \\ \mu & -\mu \end{pmatrix}$ avec $\lambda, \mu > 0$.*

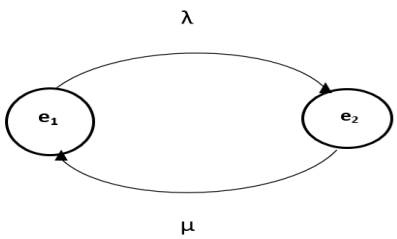

FIGURE 3.3 – Graphe de transition.

Première méthode :

On a P_t est une matrice stocastique (transition) définie par :

$$P_t = \begin{pmatrix} p_{11}(t) & p_{12}(t) \\ p_{21}(t) & p_{22}(t) \end{pmatrix}$$

On a $P_t' = P_t Q$ d'où :

$$P_t' = \begin{cases} p_{11}'(t) = -\lambda p_{11}(t) + \mu p_{12}(t) & (1) \\ p_{12}'(t) = \lambda p_{11}(t) - \mu p_{12}(t) & (2) \\ p_{21}'(t) = -\lambda p_{21}(t) + \mu p_{22}(t) & (3) \\ p_{22}'(t) = \lambda p_{21}(t) - \mu p_{22}(t) & (4) \end{cases}$$

De plus, $\sum_{j\in E} \mathbb{P}_{ij} = 1$ alors :

$$\begin{cases} p_{11}(t) + p_{12}(t) = 1 & (5) \\ p_{21}(t) + p_{22}(t) = 1 & (6) \end{cases}$$

D'aprés les équations (1) et (5) on trouve :

$$p_{11}'(t) = -\lambda p_{11}(t) + \mu(1 - p_{11}(t))$$

Donc,

$$p_{11}'(t) + (\lambda + \mu) p_{11}(t) = \mu$$

$$p_{11}(t) = \frac{\mu}{\lambda + \mu} + c e^{-(\lambda+\mu)t}, \qquad c \in \mathbb{R}.$$

On sait que $p_{11}(0) = 1$ donc $c = \frac{\lambda}{\lambda + \mu}$ d'où :

$$p_{11}(t) = \frac{1}{\lambda + \mu}(\mu + \lambda e^{-(\lambda+\mu)t})$$

Et d'après l'équation (5), on a alors :

$$p_{12}(t) = \frac{1}{\lambda + \mu}(\lambda - \lambda e^{-(\lambda+\mu)t})$$

les équations (3) et (6) donnent :

$$p_{21}'(t) = \lambda p_{21}(t) + \mu(1 - p_{21}(t)) \Rightarrow p_{21}'(t) + (\lambda + \mu) p_{21}(t) = \mu$$

d'où

$$p_{21}(t) = \frac{\mu}{\lambda + \mu} + c' e^{-(\lambda+\mu)t}, \qquad c' \in \mathbb{R}.$$

Pour $p_{21}(0) = 0$, *on obtient* $c' = -\dfrac{\mu}{\lambda+\mu}$ *alors :*

$$p_{21}(t) = \frac{1}{\lambda+\mu}(\mu - \mu e^{-(\lambda+\mu)t})$$

Et d'après (6), on a :

$$p_{22}(t) = \frac{1}{\lambda+\mu}(\lambda + \mu e^{-(\lambda+\mu)t})$$

et nous retrouvons P_t *:*

$$P_t = \frac{1}{\lambda+\mu}\begin{pmatrix} \mu + \lambda e^{-(\lambda+\mu)t} & \lambda - \lambda e^{-(\lambda+\mu)t} \\ \mu - \mu e^{-(\lambda+\mu)t} & \lambda + \mu e^{-(\lambda+\mu)t}. \end{pmatrix}$$

Deuxième méthode :

Dans cette méthode, nous utilisons équation l'exponentielle matricielle (3.13) :
Pour $Q = \begin{pmatrix} -\lambda & \lambda \\ \mu & -\mu \end{pmatrix}$ *alors,*

$$\begin{aligned} Q^2 &= \begin{pmatrix} -\lambda & \lambda \\ \mu & -\mu \end{pmatrix} \times \begin{pmatrix} -\lambda & \lambda \\ \mu & -\mu \end{pmatrix} = \begin{pmatrix} \lambda^2 + \lambda\mu & -\lambda^2 - \lambda\mu \\ -\mu^2 - \lambda\mu & \mu^2 + \lambda\mu \end{pmatrix} \\ &= -(\lambda+\mu)Q \\ Q^3 &= Q \times Q^2 = -(\lambda+\mu)Q^2 = (\lambda+\mu)^2 Q \\ &\cdots \\ Q^k &= [-(\lambda+\mu)]^{k-1} Q. \end{aligned}$$

d'aprés l'équation (3.13) :

$$\begin{aligned} P(t) &= I + \sum_{k=1}^{\infty} \frac{[-(\lambda+\mu)]^{k-1} Q t^k}{k!} = I - \frac{1}{\lambda+\mu}\sum_{k=1}^{\infty} \frac{[-(\lambda+\mu)t]^k}{k!} Q \\ &= I - \frac{1}{\lambda+\mu}(e^{-(\lambda+\mu)t} - 1)Q = I + \frac{1}{\lambda+\mu}Q - \frac{1}{\lambda+\mu}Qe^{-(\lambda+\mu)t} \\ &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} \frac{-\lambda}{\lambda+\mu} & \frac{\lambda}{\lambda+\mu} \\ \frac{\mu}{\lambda+\mu} & \frac{-\mu}{\lambda+\mu} \end{pmatrix} - \begin{pmatrix} \frac{-\lambda}{\lambda+\mu} & \frac{\lambda}{\lambda+\mu} \\ \frac{\mu}{\lambda+\mu} & \frac{-\mu}{\lambda+\mu} \end{pmatrix} e^{-(\lambda+\mu)t} \\ &= \begin{pmatrix} \frac{\mu}{\lambda+\mu} & \frac{\lambda}{\lambda+\mu} \\ \frac{\mu}{\lambda+\mu} & \frac{\lambda}{\lambda+\mu} \end{pmatrix} + \begin{pmatrix} \frac{\lambda}{\lambda+\mu} & \frac{-\lambda}{\lambda+\mu} \\ \frac{-\mu}{\lambda+\mu} & \frac{\mu}{\lambda+\mu} \end{pmatrix} e^{-(\lambda+\mu)t} \end{aligned}$$

finalement nous trouvons :

$$P_t = \frac{1}{\lambda+\mu}\begin{pmatrix} \mu + \lambda e^{-(\lambda+\mu)t} & \lambda - \lambda e^{-(\lambda+\mu)t} \\ \mu - \mu e^{-(\lambda+\mu)t} & \lambda + \mu e^{-(\lambda+\mu)t}. \end{pmatrix}$$

Nous remarquons bien que la méthode de Chapman Kolmogorov est basée sur l'hypothèse fondamentale de la propriété markovienne, qui stipule que l'état futur d'un système markovien dépend uniquement de son état présent et non pas de son histoire passée. Cependant, dans la file d'attente non markoviennes, cette propriété n'est pas satisfaite. Dans de tels systèmes, l'état futur peut dépendre non seulement de l'état actuel, mais aussi de l'historique complet du système. Dans ces cas, d'autres méthodes doivent être utilisées pour analyser la stabilité du système comme la méthode au sens probabiliste qui fait intervenir la fonction de Liapunov aussi on peut utiliser la méthode des limites fluides. (voir [**32**], [**22**])

Conclusion générale

Dans ce travail, nous avons résumé les principaux idées de structure des files d'attente. Ce travail a pésenté les types de files d'attente en mettant certain aspects en evidences et leur impact sur le systeme, ainsi que les différentes modélisation et ses propriétés de ces systèmes, ce manuscrit explore une méthode pratique qui souligne l'importance de la stabilité et la dérivation de ses condition qui est utilisé jusqu'a aujourd'hui pour la recherche dans ce domaine .

Bibliographie

[1] Abbas. A.J., Contribution a la modelisation et a la commande par les reseaux de Petri VOD : application a la minimisation des temps de correspondances des systemes de transport public, (2003).

[2] Adan .I., Economou .A., Kapodistria .S., Synchronized reneging in queueing systems with vacations, Queueing Systems, 62, 1-33, (2009).

[3] Aissani .A., Kartashov .N.V., Ergodicity and stability of markov chains with respect to operator topology in the space of transition kernels, Sciences U.S.S.R, serie A 11, 3-5, (1983).

[4] Al-Seedy .R.O., A transient solution of the non-truncated queue$M/M/2$ with balking and an additional server for longer queues (Krishnamoorthi discipline), Applied Mathematics and Computation , 3, 763-769, (2004).

[5] Ancker.C.J. and Gafarian.A.V., Queuing with reneging and multiple heterogeneous servers, Nav. Res. Logist. Q., 10, 125-145, (1963).

[6] Andreas, Manfred., On the M(n)/M(n)/s queue with impatient calls, Performance Evaluation, 35, 1-18, (1999).

[7] Artalejo.J. R., Gomez-Corral.A., Advances in retrial queues, European Journal of Operation Research, 189(3), 1042-1233, (2008).

[8] Baruah.B, Analysis of some batch arrival queueing systems with balking, renrging, random breakdowns, fluctuating modes of service and Bernoulli schedulled server vacation, (2017).

[9] Baynat.B., La théorie des files d'attente : des chaînes de Markov aux réseaux à forme produit. Hermès, (2000).

[10] Berdjoudj.L., Séminaire Mathématique de Béjaia (LaMOS), Sur l'analyse de stabilité des systèmes de files d'attente avec rappels via les martingales, Volume 1, (2003).

[11] Blackburn .J.D., Optimal control of a single-server queue with balking and reneging, Management Science, 19, 297-313, (1972).

[12] Bodineau.T.,Modélisation de phénomènes aléatoires : introduction aux chaînes de Markov et aux martingales, (2023).

[13] Boots .N., Tijms .H., An M/M/c queue with impatient customers, TOP, 7, 213-220, (1999).

[14] Borovkov .A., A Stochastic Processes in queueing theory, Springer-Verlag, Berlin, (1976).

[15] Caumel. Y.,Probabilité et Processus Stochastiques. Springer-Verlag FRance, (2011).

[16] Crommelin .C.D., Delay probability formulae, Post Office Electrical Engineers Journal, 26, 266-274, (1934).

[17] Engset. T., Emploi duu calcul des probabilites pour la determination du nombre des selecteurs dans les bureaux telephoniques centraux,Rev.Gen.D'elect, 9, 40-138, (1918).

[18] Erlang .A. K., The theory of probabilities and telephone conversations. Nyt Tidsskrift for Matematik B, (1909).

[19] Falin G.I., A survey of retrial queues. Queueing Systems, 7 : 127-168, (1990).

[20] Falin G.I et J.G.C. Templeton., Retrial Queues. Chapman and Hall, (1997).

[21] Federgruen.A et L.Green. Queueing systems with service interruptions. Research Working. Columbia University,30 : 5-84, (1984).

[22] Foss, S.G. and Konstantopoulos, T., An overview of some stochastic stability methods, Journal of Operation Research Society Japan, 47, No.4, 275- 303, (2004) .

[23] Garnett.O., A. Mandelbaum, and M. Reiman. Designing a call center with impatient customers. Manufacturing Service Operations Management, 208-227, (June 2002).

[24] Gilles Pagès., Martingales à temps discret, Théorème de convargence, Appllication à la loi forte des grands nombres, (2005-2006).

[25] Haight. F. A., Queuing with balking, I, Biometrika, 44, 360-369. (1957).

[26] Ibe. O., Markov processes for stochastic modeling.(second esition), (2013).

[27] Sztrik.j., Basic Queuing theory. p 25, (2016).

[28] Jinan.A.N., Al-obedy Publié dans Ibn AL-Haitham Journal For Pure and Applied Science, pp.333, (2023).

[29] Kalyanaraman .R., Kalaiselvi .S., Heterogeneous server queue with a threshold on slow server, International Journal of Recent Technology and Engineering (IJRTE), 7, 2277-3878, (2019).

[30] Kleinrock.L., Queueing système volume 1 : Theory (1975).

[31] Ke, Wu and Zhang : Recent Developments in Vacation Queueing Models : A Short Survey IJOR Vol. 7, No. 4, (2010).

[32] Kernane.T.Stabilité de modèles de files d'attent, (2007).

[33] Krishnamoorthy .A. , Sreenivasan .C., An $M/M/2$ queueing system with heterogeneous servers including one with working vacation, International Journal of Stochastic Analysis 1-16, (2012).

[34] Kumar .B.K., Madheswari .S.P., An $M/M/2$ queuing system with heterogeneous server and multiple vacations, Mathematical and Computer Modelling, 41, 1415-1429, (2005).

[35] Kumar.B.K., Madheswari.S.P., Venkatakrishnan.K.S., Transient solution of an $M/M/2$ queue with heterogeneous servers subject to catastrophes, Information and Management Sciences, 63-80, (2007).

[36] Little .John.D.C., A proof of the queueing formula L = λW, Oper. Res, 9(3), 383-387, (1961).

[37] Lozano .M., Moreno .P., A discrete time single-server queue with balking : economic applications, Applied Economics, 40, 735-748, (2008).

[38] Madan .K.C., Abu-Dayyeh .W., Taiyyan .F., A two server queue with Bernoulli schedules and a single vacation policy, Applied Mathematics and Computation, 145, 59-71, (2003).

[39] Meyn.S., Tweedie.R., Markov Chains and Stochastic Stability, (second edition 2009),(first edition 1993).

[40] Morse .P.M., Queues inventories and maintenance, Operations Research Society Of America, 359-520, (1958).

[41] Nasrallah .W.F., How pre-emptive priority affects completion rate in an M/M/1 queue with poisson reneging, European Journal of Operational Research, 193, 317-320, (2009).

[42] Newell.G.F., Applications of Queueing Theory Second edition(First published 1971), LONDON(1982).

[43] O'Dell.G.F., Gibson.W.W., Automatic trunking in theory and practice. Inst.P.O.Elect.Engrs,41, 107, (1926).

[44] Oliver.I., Markov Processes for Stochastic Modeling, (2013).

[45] Omarah .A.S.R., A transient solution of queues with variable channel considering balking concept when S = 2 and [sigma] = 1, Applied Mathematics and Computation, 174, 337-344, (2006).

[46] Perel .N. , Yechiali .U., Queues with slow servers and impatient customers, European Journal of Operational Research, 201, 247-258, (2010).

[47] Rajan .V., Queuing analysis of markovian queue Having two heterogeneous servers with catastrophes using matrix geometric technique, International Journal of Statistics and Systems, 12, 205-212, (2017).

[48] Rao .S.S., Queueing mdels with balking, reneging and interruptions, Operations Research, 13, 596, (1965)

[49] Rykov .V., Monotone control of queueing systems with heterogeneous servers, Queueing Sys, 37(4), 391-403, (2001).

[50] Rykov .V., Efrosinin .D.V., On the slow server problem, Automa. Remote Cont, 70(12), 2013-2023, (2009).

[51] Saaty .T.L., Elements of queueing theory with applications, McGraw-Hill, (1961).

[52] Singh .V.P., Two-server markovian queues with balking : heterogeneous vs. homogeneous servers, Operations Research, 18, 145-159, (1970).

[53] Sharma .D. P., Dass .S., Multiserver markovian queues with finite waiting space, Sankhya, B-50, 328-331, (1998).

[54] Salch .A., Thèse Ordonnancement stochastique avec impatience , (2013).

[55] Sridhar .A., Allah Pitcha .R., Analyses of a Markovian queue with two heterogeneous servers and working vacation, International Journal of Applied Operational Research, 5, 1-15, (2015).

[56] Stewart, William J. Probability, Markov chains, queues and simulation : the mathematical basis of performance modeling, (1946).

[57] Stoyan .D., Comparaison methods of queue and other stochastic models, Wileys, new york, (1983).

[58] Subba Rao .S., Queuing with balking and reneging in M/G/1 systems, Metrika, 12, 173-188, (1967) .

[59] Takacs .L., A Single server queue with feedback, The Bell System Tech. Journal, 42, 134-149, (1963).

[60] Takhedmit.B., Analyse des systèmes de files d'attente par la méthode des Martingales, (2012).

[61] Touzi.N., Martingales en temps discret et chaînes de Markov, (2009).

[62] Van Tits .M. H. L., Van der Veeken .H. J. M., Simulation of a queueing problem with balking, Simulation, 35, 88-93, (1980).

[63] Vaulot .E., Application du calcul des probabilities a l'exploitation telephonique, Annales des Postes, Télégraphes et Téléphones , 14, 136-156, (1925).

[64] Wang .K-H., Chang .Y-C., Cost analysis of a finite M/M/R queueing system with balking, reneging and server breakdowns, Mathematical Methods of Operations Research, 56(2), 169-180, (2002).

[65] Ward .A.R., Glynn .P.W., A diffusion approximation for a GI /GI /1 queue with balking or reneging, Queueing Systems, 50, 371-400, (2006)

[66] Wilkinson .R.I., The interconnection of telephone systems, The Bell System Technical Journal, 10, 531-564, (1931).

[67] Xiong .W., Altiok .T., An approximation for multi-server queues with deterministic reneging times, Annals of Operations Research, 172, 143-151, (2009).

[68] Yang .T.et J and Templeton G.C., A survey on retrial queues. Queueing systems, 2 : 201-233, (1987).

[69] Ycart.B, Files d'attente.Cahier de Mathématiques Appliquées n^o14, (2004)

[70] Yue .D., Yue .W., A heterogeneous two-server network system with balking and a Bernoulli vacation schedule, in : Proceedings of the 4th International Conference on Queueing Theory and Network Applications, Article No. 20, (2009).

[71] Zakhar. Kabluchko. Stochastic Processes (Stochastik II), (2014).

[72] Zeltyn .S., Mandelbaum .A., Call Centers with Impatient Customers : Many-Server Asymptotics of the M/M/n + G Queue, Queueing Systems, 51, 361-402, (2005).