N° d'ordre :

REPUBLIQUE ALGERIENNE DEMOCRATIQUE & POPULAIRE

MINISTERE DE L'ENSEIGNEMENT SUPERIEUR & DE LA RECHERCHE SCIENTIFIQUE

UNIVERSITE DJILLALI LIABES
FACULTE DES SCIENCES EXACTES
SIDI BEL ABBES

THÈSE DE DOCTORAT

Présentée par

KADI MOKHTAR

Spécialité : *Mathématiques*

Option : probabilités- *Statistique*

Intitulée

Estimation et analyse d'un système de files d'attente avec dérobades.

Soutenue le : 21/06/2017

Devant le jury composé de :

Président	***Aboubaker MECHAB***	***Univ. Djillali Liabes SBA (MCA)***
Examinateurs :	***Toufik GUENDOUZI***	***Univ. Moulay Tahar Saida (Pr)***
	Abdeldjebbar KANDOUCI	***Univ. Moulay Tahar Saida (Pr)***
	Fethi MADANI	***Univ. Moulay Tahar Saida (MCA)***
Directeur de thèse:	***Abbes RABHI***	***Univ. Djillali Liabes SBA (MCA)***
Co-Directeur de thèse:	***Amina Angelika BOUCHENTOUF***	***Univ. Djillali Liabes SBA (MCA)***

dédicace

A la mémoire de mon Père lhaj Mohamed.
à tous les membres de ma famille sans aucune exception ...

Remerciement

Par ces quelques lignes, je tiens à remercier toutes les personnes qui ont participé de près ou de loin au bon déroulement de cette thèse, en espérant n'avoir oublié personne ...
Je tiens tout d'abord à remercier le directeur de cette thèse Mr Abess RABHI (Maitre de conférence l'université Djillali liabes Sidi Bel Abbes) d'avoir bien assuré la direction et l'encadrement de mes travaux de thèse. Je vous remercie pour vos conseils, vos encouragements.
Je tiens à remercier mon co-directeur de thèse Melle Amina Angelika BOUCHENTOUF (Maitre de conférence à l'université Djillali liabes Sidi Bel Abbes), pour le temps et la patience qu'elle a manifestées à mon égard durant cette thèse en me fournissant d'excellentes conditions logistiques. Je garderai dans mon coeur votre générosité, votre compréhension et votre efficacité. Pour tout ce que vous m'avez donné, je vous remercie très sincèrement.
Mes sincères remerciements et ma gratitude vont aussi à Mr Aboubaker MECHAB (Maitre de conférence à l'université Djillali liabes Sidi Bel Abbes) pour avoir accepté de juger ce travail et d'en présider le jury de soutenance. Que vous soyez assuré de mon entière reconnaissance.
Je remercie tous ceux sans qui cette thèse ne serait pas ce qu'elle est : aussi bien par les discussions que j'ai eu la chance d'avoir avec eux, leurs suggestions ou contributions. Je pense ici en particulier à Mr Abdeldjebbar KANDOUCI (Professeur à l'université Dr Tahar Moulay Saida), qui de plus m'a fait l'honneur d'être parmi les rapporteurs de cette thèse.
Mr Toufik GUENDOUZI (Professeur à l'université Dr Tahar Moulay Saida) et Mr Fethi MADANI (Maitre de conférence à l'université Dr Tahar Moulay Saida) ont accepté d'être les rapporteurs de cette thèse, et je les en remercie, de même que pour leur participation au Jury. Ils ont également contribué par leurs nombreuses remarques et suggestions à améliorer la qualité de cette thèse, et je leur en suis très reconnaissant.

Résumé

Dans cette thèse, nous étudions différents modèles de systèmes de files d'attente. En premier lieu, nous considérons une file d'attente $M/M/2/N$ avec dérobade, abandon et deux serveurs hétérogènes. En utilisant le processus de Markov, nous développons d'abord les équations des probabilités d'état stable. Ensuite, nous donnons diverses mesures de performance du système. Enfin, nous présentons quelques exemples numériques pour démontrer comment les différents paramètres du modèle influencent sur le comportement du système.

En second lieu, nous traitons un modèle de file d'attente avec rappel et dérobade, où avec une certaine probabilité, les clients résidant dans l'orbite peuvent abandonner le système après un certain temps aléatoire. Pour ce système nous dérivons la formule analytique la fonction génératrice de la probabilité conjointe du nombre de clients dans l'orbite et du temps de service restant pour un client entrain d'être servi. De plus nous obtenons la fonction génératrice de la probabilité du nombre total de clients dans le système. Notons que ces fonctions génératrices pourraient être utilisées pour obtenir des mesures de performance importantes telles que la taille moyenne du système et le temps d'attente moyen.

Abstract

In this thesis, we study diiferent models of queueing systems. First, we consider a $M/M/2/N$ queueing system with balking, reneging, feedback and two heterogeneous servers. By using the Markov process, we first develop the equations for steady-state probabilities. Then we give some system performance measures. Finally, we present some numerical examples to demonstrate how the various model parameters influence the behavior of the system.
Secondly, we study a retrial queueing model with balking, such that with a certain probability customers in the orbit can leave the system after certain random time. For this system, we derive analytic formulas for both the joint probability generating function of the remaining service time of the customer currently in the server and the number of customers in the orbit, and the probability generating function of the total number of customers in the system. Note that these generating functions could be used to obtain important performance measures such as average system size and average waiting time.

Table des matières

Introduction

La théorie des files d'attente a commencé en 1909 avec les travaux de recherches de l'ingénieur danois Agner Krarup Erlang (1878,1929) sur le trafic téléphonique de Copenhague pour déterminer le nombre de circuits nécessaires afin de fournir un service téléphonique acceptable. Par la suite, les files d'attente ont été intégrés dans la modélisation de divers domaines d'activité. On assista alors à une évolution rapide de la théorie des files d'attente qu'on appliqua à l'évaluation des performances des systèmes informatiques et aux réseaux de communication. Les chercheurs oeuvrant dans cette branche d'activité ont élaboré plusieurs nouvelles méthodes qui ont été ensuite appliquées avec succés dans d'autres domaines, notamment dans le secteur de la fabrication. On a aussi constaté une résurgence des applications pratiques de la théorie des files d'attente dans des secteurs plus traditionnels de la recherche opérationnelle, un mouvement mené par Peter Kolesar et Richard Larson. Grâce à tous ces développements, la théorie des files d'attente est aujourd'hui largement utilisée et ses applications sont multiples.

Dans cette thèse on s'intéresse aux systèmes de files d'attente avec clients impatients. Cette impatience est due à plusieurs facteurs, soit par insuffisance de nombre de serveurs, soit par une qualité de service médiocre, ou encore une mauvaise gestion du système.

Le dérobade (balking) dans une file d'attente 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.

Les modèles de files d'attente avec dérobade, abandon, ou les deux, étaient traités par plusieurs chercheurs. Haight, [46] est le premier qui a étudié le modèle d'attente $M/M/1$ avec dérobade. Ensuite, Haight, [48] a traité ce même modèle $M/M/1$ mais avec abandon. L'effet de la combinaison entre le dérobade et l'abandon sur une file $M/M/1/N$ a été le but d'un travail proposé par Ancker et Gafarian [7, 8]. Abou-EI-Ata et Hariri [1] ont considéré le modèle $M/M/c/N$ avec dérobade, abandon et plusieurs serveurs. Ensuite Wang et Chang [75] ont fait l'extension de ce travail pour étudier $M/M/c/N$ avec dérobade, abandon et serveurs en panne. Kumar et Sharma [61] ont étudié un système d'un seul serveur avec abandon et dérobade des clients. Kumar et Sharma [6] ont traité le cas d'un seul serveur avec rappels des clients, dérobade et abandon et Mahdy El-Paoumy et Hossam Nabwey [68] ont étudié le modèle $M/M/2/N$ avec dérobade générale.

Le retour (feedback) d'un client dans une file d'attente a fait l'objet de plusieurs travaux. Ce feedback est du à l'insatisfaction de ce client en raison de la qualité inappropriée de service. Après avoir obtenu un service partiel ou incomplet, le client réessaie sa demande pour le service. En communication avec l'ordinateur, le protocole de transmission des données est répété parfois à cause de l'apparition d'une erreur. C'est le cas d'une insatisfaction de la qualité de service. Takacs [74] a étudié la file d'attente avec feedback pour déterminer la taille de la file d'attente et les deux premiers moments de la fonction de distribution

du temps total passé par un client dans le système pour un processus stationnaire. Dans [15] Avignon et Disney ont étudié les files d'attente à un seul serveur avec retour. Santhakumaran et Thangaraj [72] ont considéré le modèle de file $M/M/1$ à un serveur avec feedback. ils ont calculé la longueur, la moyenne et la variance de la de la file d'attente à des instants d'arrivée aléatoires pour la distribution stationnaire. Thangaraj et Vanitha [77] ont obtenu les solutions du modèle $M/M/1$ avec feedback pour un état transitoire. Ayyapan et. al [16] ont étudié $M/M/1$ un système de file d'attente avec rappels et feedback dans un service avec priorité en utilisant la méthode des matrices géométriques. Sharma et Kumar [19] ont considéré un système avec un serveur unique, la capacité de la file étant finie, le processus de feedback est Markovien.

La théorie des files d'attente avec rappel a été surtout utilisée pour modéliser les systèmes téléphoniques, centres d'appels et les réseaux informatiques. Elle sert à résoudre des problèmes pratiques, tels que l'analyse du temps d'attente des abonnés dans les réseaux téléphoniques, l'évitement de collision dans les réseaux locaux, l'analyse du temps d'attente pour accéder à la mémoire sur les disques magnétiques, $\cdots$ [10], [90].

Une file d'attente avec rappels est un modèle où un client qui arrive de l'extérieur et trouve tous les serveurs occupés, quitte le service pour rappeler ultérieurement à des instants aléatoires jusqu'à satisfaction de sa demande. Entre deux rappels successifs le client rejoint un espace appelé orbite. Ces systèmes sont en particulier utilisés dans la modélisation des réseaux de télécommunication et des systèmes informatique.

Parmi les premiers travaux sur les modèles d'attente avec rappels, on trouve ceux de Cohen [33], de Eldin [37], de Hashida et Kawashima [49] et de Lubacz et Roberts [67], Kulkani et Liang [60], Templeton [86] et dans les monographies de Falin et Templeton [39], Artalejo et Gòmez-Corral [11], Gòmez-Corral et Ramalhoto [47], Rodrigo [78] et dans les travaux bibliographiques de Artalejo [12], [13].

Dans notre thèse nous analysons des modèles de systèmes de files d'attente avec dérobade, abandon, rappel et feedback. La thèse est composée de cinq chapitres.

Dans le premier chapitre nous abordons les processus à la base de l'étude de tels systèmes d'attente qui sont les chaînes de Markov. Nous présentons une introduction aux concepts de base de la théorie des chaînes de Markov en temps discret et en temps continu. Nous présentons également les propriétés fondamentales des chaînes de Markov (l'homogénéité, la périodicité, irréductibilité,...). Ensuite nous donnons certaines relations entre les chaînes de Markov des états finis et la théorie des matrices. Enfin nous classons les différents états des chaînes de Markov (états communiquants, états périodiques, ...

Dans le deuxième chapitre, nous introduisons la terminologie de la théorie des files d'attente. Certaines définitions et notations qui sont nécessaires dans l'étude des systèmes de files d'attente (la notation de KANDELL, la formule de LITTLE $\cdots$) sont nottamment données. Et nous étudions quelque modèles de files d'attente markoviennes ($M/M/1$, $M/M/1/K$, $M/M/c$, $M/M/\infty$,) et non markovienne ($M/G/1$) et l'évaluation de leurs paramètres de performance.

Le troisième chapitre représente une étude de certains modèles d'attente avec clients impatients. Nous traitons le cas de files d'attente avec dérobade ($M/M/1$, $M/M/s$), ensuite un système d'attente avec abandon et enfin un modèle avec rappels ($M/G/1$).

Dans le quatrième chapitre, nous traitons le cas d'un système de files d'attente de deux serveurs hétérogènes $M/M/2/N$ (capacité N finie) avec dérobade, abandon et feedback. En utilisant le processus de Markov, nous développons d'abord les équations des probabilités d'état stable. Ensuite, nous donnons quelques mesures de performance du système. Enfin, nous présentons quelques exemples numériques pour démontrer comment les différents paramètres du modèle influencent sur le comportement du système.

Enfin dans le cinqième chapitre, nous étudions un nouveau modèle de file d'attente avec rappel et dérobade, ou avec une certaine probabilité, les clients résidant dans l'orbite peuvent abandonner le système après un certain temps aléatoire. Pour ce système nous dérivons la formule analytique la fonction génératrice de la probabilité conjointe du nombre de clients dans l'orbite et du temps de service restant pour un client entrain d'être servi. De plus nous obtenons la fonction génératrice de la probabilité du nombre total de clients dans le système. Notons que ces fonctions génératrices pourraient être utilisées pour obtenir des mesures de performance importantes telles que la taille moyenne du système et le temps d'attente moyen.

Les résultats issus du quatrième ont fait l'objet d'une publication dans le journal Mathematical Sciences And Applications E-Notes, (2013) Volume 2 No. 2 pp. 10-21 [24], et d'une communication internationale [69]. Ainsi que ceux du cinquième chapitre, qui font l'objet d'un travail soumis dans le journal Applied and Computational Mathematics (Septembre 2014).

Chapitre 1

Chaînes de Markov

1.1 Introduction

Les chaînes de Markov fournissent des moyens très puissants et efficaces pour la description et l'analyse des propriétés des systèmes dynamiques (files d'attente, réseaux informatiques et téléphoniques, physique, biologiques ou économiques ... etc). Leurs utilisations pour la modélisation des phénomènes aléatoires évoluant dans le temps, donnent aux chaînes de Markov une place importante. La structure simple des chaînes de Markov permet de dire beaucoup de choses sur le comportement et l'évolution de ses phénomènes. En effet, l'ensemble des études mathématiques des processus aléatoires peut être considéré comme une généralisation d'une manière ou d'une autre de la théorie des chaînes de Markov.
Le processus de Markov fournit un outil simple de modélisation des systèmes de files d'attente.
Dans ce chapitre, nous donnons d'abord une introduction aux concepts de base de la théorie des chaînes de Markov en temps discret et en temps continu. Nous présentons également les propriétés fondamentales des chaînes de Markov. Ensuite nous donnons certaines relations entre les chaînes de Markov des états finis et la théorie des matrices.

1.1.1 Définitions et propriétés de base

Définition 1.1.1. *Un processus stochastique est une famille de variables aléatoires $X(t), t \in T$ où chaque variable aléatoire $X(t)$ est indexée par le paramètre $t \in T$. Si T est un ensemble de $\mathbb{R}_+$, alors t signifie temps.*
Généralement $X(t)$ représente l'état du processus stochastique au temps t.

- *Si T est dénombrable, i.e $T \subseteq \mathbb{N}$, alors nous disons que $X(t), t \in T$ est un processus à temps discret.*
- *Si T est un intervalle de $[0; \infty)$, alors le processus stochastique est dit un processus à temps continu.*

L'ensemble des valeurs de $X(t)$ est appelé l'espace d'état, qui peut également être soit discret (fini ou infini dénombrable) ou continu (un sous-ensemble de $\mathbb{R}$ ou $\mathbb{R}^n$), donc nous écrivons $(X_n)_{n\geq 0}$ pour le processus à temps discret et $(X_t)_{t\geq 0}$ pour le processus à temps continu.

1.1.2 Matrices stochastiques

Définition 1.1.2. *Une matrice stochastique est une matrice (a_{ij}) avec $i, j \in \{1, 2, \cdots r\}$ telle que :*

$$\begin{aligned} a_{ij} &\geq 0 \\ \sum_{j=1}^{r} a_{ij} &= 1 \quad \forall i \in \{1, 2, \cdots, r\} \end{aligned}$$

Théorème 1.1.2.1. *[53] Le produit de deux matrices stochastiques est une matrice stochastique*

Démonstration. Supposons que les matrices A et B sont stochastiques et posons $C = AB$. On a

$$\forall i, j \in \{1, 2, \cdots, r\}, c_{ij} = \sum_{k=1}^{r} a_{ik} b_{kj} \geq 0$$

et

$$\begin{aligned} \forall i, j \in \{1, 2, \cdots, r\}, \sum_{k=1}^{r} c_{ij} &= \sum_{j=1}^{r} \sum_{k=1}^{r} a_{ik} b_{kj} \\ &= \sum_{k=1}^{r} a_{ik} \underbrace{\sum_{j=1}^{r} b_{kj}}_{=1, \forall k} \\ &= \sum_{k=1}^{r} a_{ik} = 1 \end{aligned}$$

En général la matrice A^n de transition en n étapes est stochastique.

1.2 Chaînes de Markov à temps discret

Les Chaînes de Markov à temps discret jouent un rôle central dans la théorie des processus aléatoires. Leur vastes domaines d'applications concernent la modélisation dans les secteurs économiques, le traitement des signaux et des systèmes informatiques, phénomènes météorologiques ou sismiques, $\cdots$. elles sont définies par des familles discrètes de variables ou de vecteurs aléatoires $(X_n)_{n \in T}$, où T est l'ensemble des temps des observations des états du processus.

Pour cette section nous donnons les définitions de base sur les chaînes de Markov homogènes ainsi que certaines propriétés.

Définition 1.2.1. *(**Chaînes de Markov.**) Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires prenant leurs valeurs dans un espace dénombrable E (dans notre cas, nous prendrons $\mathbb{N}$ ou un sous-ensemble de $\mathbb{N}$). $(X_n)_{n \in \mathbb{N}}$ est une chaîne de Markov si et seulement si*

$$\mathbb{P}\left(X_{n+1} = j | X_0 = i_0, \cdots, X_{n-1} = i_{n-1}, X_n = i\right) = \mathbb{P}\left(X_{n+1} = j | X_n = i\right) \tag{1.1}$$

Ainsi, pour une chaîne de Markov, la probabilité d'un comportement futur particulier du processus, lorsque son présent actuel est bien connu, n'est pas modifié par toute connaissance supplémentaire du passé. Le processus est dit sans mémoire ou non héréditaire.

Soit la notation :

$$p_{ij} = \mathbb{P}\left(X_n = j | X_{n-1} = i\right) \tag{1.2}$$

p_{ij} est appelé la probabilité de transition de l'état i vers l'état j. Lorsque cette probabilité ne dépend pas de n, nous disons que la chaîne de Markov est homogène.

Remarque 1.2.1. *En général les probabilités de transition sont des fonctions non seulement des états initiaux et finaux, mais aussi du temps de transition. Lorsque les probabilités de transition en une seule étape sont indépendantes de la variable temps n, nous disons que la chaîne de Markov est stationnaire.*

Définition 1.2.2. *La matrice*

$$P = \begin{pmatrix} P_{00} & P_{01} & P_{02} & P_{03} & \cdots \\ P_{10} & P_{11} & P_{12} & P_{13} & \cdots \\ P_{20} & P_{21} & P_{22} & P_{23} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \\ P_{i0} & P_{i1} & P_{i2} & P_{i3} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \cdots \end{pmatrix} \tag{1.3}$$

dont les coefficients sont les probabilités de transition P_{ij} est appelée à la fois matrice de Markov et matrice des probabilités de transitions du processus.

La $i$ème ligne de P, pour $i = 0, 1, \cdots$ est la distribution de probabilité des valeurs X_{n+1} avec la condition que $X_n = i$. Si le nombre d'états est fini, alors P est une matrice carrée finie dont l'ordre (le nombre de lignes) est égal au nombre d'états. Les quantités P_{ij} satisfont la conditions :

$$P_{ij} \geq 0 \quad \text{pour} \quad i, j = 0, 1, 2, \cdots \tag{1.4}$$

$$\sum_{j=0}^{\infty} P_{ij} = 1 \quad \text{pour tout} \quad i = 0, 1, 2, \cdots \tag{1.5}$$

les nombres P_{ij} sont des probabilités donc des nombres positifs.
Un processus de Markov est complètement défini une fois que sa probabilité de transition et X_0 son état initial sont donnés.

1.3 Classification des états

1.3.1 Relation de communication entre états

Définition 1.3.1.1. *Soient i et j deux états. L'état j est accessible à partir de l'état i s'il existe $m > 0$ tel que $p_{ij}^m > 0$.*

Résultat 1.3.1.1. *Un état i est toujours accessible à partir de lui-même, car $p_{ii}(0) = 1$. on dit que les états i et j communiquent si i est accessible à partir de j et j est accessible à partir de i, ce qui est désigné par $i \leftrightarrow j$.*
Il est clair que,
$i \leftrightarrow i$ (réflexivité),
$i \leftrightarrow j \Rightarrow j \leftrightarrow i$(symétrie),

$i \leftrightarrow j, j \leftrightarrow k \Rightarrow i \leftrightarrow k$ *(transitivité).*

Démonstration. Les deux premières sont évidentes d'après la définition 1.3.1.1. Pour la transitivité, on suppose que :
$i \leftrightarrow j, j \leftrightarrow k$ donc il existe m et n tel que $p_{ij}^m > 0$ $p_{jk}^n > 0$ d'où

$$\begin{aligned} p_{ik}^{m+n} &= \mathbb{P}\left(X_{m+n} = k \mid X_0 = i\right) \\ &\geq \mathbb{P}\left(X_{m+n} = k, X_m = j \mid X_0 = i\right) \\ &= \mathbb{P}\left(X_m = j \mid X_0 = i\right) \mathbb{P}\left(X_{m+n} = k \mid X_m = j\right) \\ &= p_{ij}^m p_{jk}^n > 0. \end{aligned}$$

Par conséquent, la relation de communication ($\leftrightarrow$) est une relation d'équivalence, et elle génère une partition de l'espace des états E en classes d'équivalence disjointes appelées classes de communication.

Définition 1.3.1.2. *(***Irréductibilité.***)* *Une chaîne de Markov est* **irréductible** *s'il existe une seule classe de communication.*

1.3.2 Etats périodiques

Définition 1.3.2.1. La périodicité. *La période $d(i)$ d'un état i est définie par :*

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

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

Théorème 1.3.2.1. *[63] La période est une propriété de classe. Si les états i et j communiquent entre eux alors ils ont la même période.*

Démonstration. Comme i et j communiquent, alors il existe des entiers N et M tel que $p_{ij}(M) > 0$ et $p_{ji}(N) > 0$. Pour tout $k > 1$.

$$p_{ii}(M + nk + N) \geq p_{ij}(M)(p_{jj}(k))^n p_{ji}(N)$$

(En effet, le chemin $X_0 = i; X_M = j; X_{M+k} = j; X_{M+nk} = j; X_{M+nk+N} = i$ est une possibilité d'aller de i à i dans $M + nk + N$ étapes).

Par conséquent, pour tout $k > 1$ tel que $p_{jj}(k) > 0$; on a $(M + nk + N) > 0$ pour tous $n > 1$. Et par conséquent, d_i divise $M + nk + N$ pour tout $n > 1$, et en particulier, d_i divise k. Nous avons donc démontré que d_i divise tous les k de telle sorte que $p_{jj}(k) > 0$, et en particulier, d_i divise d_j. Par symétrie, d_j divise d_i, et donc, finalement, $d_i = d_j$

Théorème 1.3.2.2. *[81] Soit P une matrice stochastique irréductible avec la période d, alors pour tous les états i, j il existent $m > 0$ et $n_0 > 0$ (m et n peuvent être dépendant de i, j) de telle sorte que*

$$p_{ij}(m + nd) > 0 \quad \forall n \geq 0 \tag{1.6}$$

Démonstration. Il suffit de démontrer pour $i = j$. En effet, il existe m tel que $p_{ij}(m) > 0$, parce que j est accessible à partir de i, la chaîne étant irréductible, et donc, si pour une $n_0 > 0$ nous avons $p_{jj}(nd) > 0$ pour tout $n > n_o$, alors $p_{ij}(m+nd) > p_{ij}(m)p_{jj}(nd) > 0$ pour tout $n > n_0$. Le PGCD de l'ensemble $A = k \geq 1; p_{jj}(k) > 0$ est d et A est fermé pour addition. L'ensemble A contient donc tout sauf un nombre fini des multiples positifs de d. En d'autres termes, il existe n_0 tel que $n > n_0$ implique $p_{jj}(nd) > 0$

Propriété 1.3.2.1. *[31] Tous les états d'une même classe de communication ont la même période.*

Démonstration. Si x et $x^{'}$ communiquent, il existe k et l tels que $p_{xx'}(k) > 0$ et $p_{x'x}(l) > 0$. Alors $p_{xx}(k + l) > 0$. Soit m tel que $p_{x'x'}(m) > 0$. Alors $p_{xx}(k + m + l) > 0$. Donc $d(x)$ divise $k + l$ et $d(x)$ divise $k + m + l$; donc $d(x)$ divise m et ceci, pour tout m tel que $p_{x'x'}(m) > 0$, donc $d(x)$ divise $d(x)$. Comme x et $x^{'}$ jouent le même rôle, alors $d(x) = d(x^{'})$

Remarque 1.3.2.1. *Une classe dont un élément admet une boucle, (c'est-à-dire $p_{xx} > 0$), est obligatoirement apériodique (c'est-à-dire de période 1), mais ce n'est pas une condition nécessaire.*

Théorème 1.3.2.3. *[31] Si x et y sont dans une même classe de communication de période d, si $p_{xy}(n) > 0$ et si $p_{xy}(m) > 0$, alors d divise $m - n$.*

Démonstration. Comme y mène à x, il existe k tel que $p_{yx}(k) > 0$. On a donc $p_{xx}(m + k) > 0$ et $p_{xx}(n + k) > 0$. Donc d divise $m + k$ et $n + k$, il divise donc la différence.
Ainsi, on peut partitionner les classes périodiques de la façon suivante.

Définition 1.3.2.2. *Soit C une classe de communication de période d, et soit x_0 fixé dans C. On définit, pour $k \in 0, 1, \cdots, d - 1$,*

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

Les $C_k^{'}$ sont appelées sous-classes cycliques de C.

Théorème 1.3.2.4. *[63] Soit X une chaîne de Markov homogène dans un espace d'états E, et soit $E^{'} \subseteq E$ une classe irréductible non vide. Alors pour tout $i, j \in E^{'}$, les périodes de i et j sont les même, i.e., $d(i) = d(j)$.*

Démonstration. Soient $i, j \in E^{'}$, $i \neq j$ deux états. Comme $E^{'}$ est une classe irréductible, alors ils existent deux entiers $t, s \geq 1$ tel que $p_{ij}(t) > 0$ et $p_{ij}(s) > 0$. L'équation de Chapman Kolmogorov donnent :

$$p_{ii}(t + s) \geq p_{ij}(t)p_{ji}(s) > 0 \quad \text{et} \quad p_{jj}(t + s) \geq p_{ji}(s)p_{ij}(t) > 0;$$

Par conséquent, les nombres $d(i)$ et $d(j)$ sont différents de 0. En choisissant arbitrairement un entier $m \geq 1$,

tel que $p_{ii}(m) > 0$, appliquons plusieurs fois l'équation de Chapman Kolmogorov, nous aurons pour tout $k \geq 1$

$$p_{jj}(t + s + km) \geq p_{ji}(s)p_{ii}(km)p_{ij}(t) \geq p_{ji}(t) \left(p_{ii}(m)\right)^k p_{ij}(s) > 0.$$

Ainsi, par la définition de la période de l'état j, $d(j)$ est un diviseur à la fois de $(t + s + m)$ et de $(t + s + 2m)$, et donc c'est aussi un diviseur de leur différence $(t + s + 2m) - (t + s + m) = m$. Il en

découle immédiatement que $d(j)$ est un diviseur de tout m pour lequel $p_{ii}(t+s) > 0$, et donc un diviseur de $d(i)$; Donc, $d(j) \leq d(i)$. En changeant le rôle de i et j on obtient l'inégalité inverse $d(j) \geq d(i)$, et par conséquent $d(j) = d(i)$.

1.3.3 Etats récurrents et états transients

Définition 1.3.3.1. Les probabilités de transition. *Pour une chaîne de Markov homogène,*

$$p_{ji} = \mathbb{P}\left(X_n = i | X_{n-1} = j\right)$$

est appelé la probabilité de transition *de l'état j à l'état i.*

Définition 1.3.3.2. *Si au temps 0, la chaîne est en l'état e_i, la variable aléatoire $T_i = Min\{n \geq 1; X_n = e_i\}$ définit le temps de premier retour en e_i.*

Définition 1.3.3.3. *Un état e_i est récurrent si partant de e_i on y revient sûrement, ce qui s'exprime par :*

$$\mathbb{P}((\quad \textit{il existe} \quad n \geq 1 \quad \textit{tel que} \quad X_n = e_i) | X_0 = e_i) = 1,$$

soit encore :

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

Un état récurrent est donc visité un nombre infini de fois.

Définition 1.3.3.4. *Un état e_i est dit transient s'il n'est pas récurrent :*

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

Un état est transient s'il n'est visité qu'un nombre fini de fois.
Par définition, une chaîne irréductible ne contient aucun état transient ou absorbant.

Définition 1.3.3.5. *Une chaîne dont tous les états sont récurrents est dite récurrente, une chaîne dont tous les états sont transients est dite transiente.*

Définition 1.3.3.6. *La probabilité de première transition en n unités de temps, de l'état e_i à l'état e_j, notée $f_{i,j}^{(n)}$ est définie par :*

$$f_{i,j}^{(n)} = \mathbb{P}\left(T_j = n | X_0 = e_i\right) = \mathbb{P}\left(X_n = e_j, X_{n-1} \neq e_j, \cdots, X_1 \neq e_j | X_0 = e_i\right),$$

$f_{i,j}^{(n)}$ est la probabilité de premier retour à l'état e_i en n unités de temps.

Théorème 1.3.3.1. *[99]*

$$e_i \textit{ est récurrent} \Leftrightarrow \sum_{n=1}^{+\infty} f_{i,i}^{(n)} = 1,$$

$$e_i \textit{ est transitoire} \Leftrightarrow \sum_{n=1}^{+\infty} f_{i,i}^{(n)} < 1.$$

Définition 1.3.3.7. *On appelle temps moyen de retour en e_j :*

$$\mu_j = \mathbb{E}\left[T_j = n | X_0 = e_i\right] = \sum_n n P_{j,j}^{(n)}.$$

Il existe deux sortes d'états récurrents

- *Les états récurrents positifs*
- *Les états récurrents nuls.*

Définition 1.3.3.8. *Un état e_j est récurrent positif (ou non nul) s'il est récurrent et si le temps moyen μ_j de retour est fini. Dans le cas où le temps moyen de retour est infini, l'état est dit récurrent nul.*

Théorème 1.3.3.2. *[98] (Caractérisation des états)*

$$(e_i \quad \textit{transient} \quad) \Leftrightarrow \sum_{n\geq 1} P_{i,i}^{(n)} < +\infty,$$

$$(e_i \quad \textit{récurrent nul} \quad) \Leftrightarrow \sum_{n\geq 1} P_{i,i}^{(n)} = +\infty \quad et \quad \lim_{n\to+\infty} P_{i,i}^{(n)} = 0,$$

$$(e_i \quad \textit{récurrent positif} \quad) \Leftrightarrow \sum_{n\geq 1} P_{i,i}^{(n)} = +\infty \quad et \quad \lim_{n\to+\infty} P_{i,i}^{(n)} > 0.$$

Théorème 1.3.3.3. *[98] (Propriétés des chaînes finies)*

1. *Aucun état n'est récurrent nul.*
2. *Aucune chaîne finie n'est transiente ; en revanche, certains de leurs états peuvent être transients.*
3. *Il existe au moins un état récurrent positif.*
4. *Si en outre la chaîne est irréductible, elle est récurrente positive.*

Démonstration. (1) Supposons qu'il existe un état récurrent nul e_j, et C_j sa classe de communication, alors

$$\forall n \in \sum_{k\in C_j} P_{i,i}^{(k)} = 1 \tag{1.7}$$

Mais $\lim_{n\to+\infty} p_{j,k}^{(n)} = 0$, qui est en contradiction avec (1.7).

(2) Supposons tous les états transients, on montre alors que $\forall e_i, e_j$ transients, la série $\sum p_{j,k}^{(n)}$ converge, donc

$$\lim_{n\to+\infty} p_{j,k}^{(n)} = 0. \tag{1.8}$$

Or, $\sum_{j=1}^{card(E)} p_{ij}^{(n)} = 1 (\forall n \in \mathbb{N})$, donc, par passage à la limite quand $n \to +\infty$, on arrive à une contradiction de (1.8).

(3)Est conséquence de (1) et (2).

(4) La chaîne ayant au moins un état récurrent d'après (c) et une seule classe est récurrente positive.

1.4 Distributions stationnaires et distributions limites

Définition 1.4.1. *Une distribution $\pi_k = (\phi_1, \phi_1, \cdots, \phi_i, \cdots)$ sur l'ensemble des états est une distribution stationnaire ou invariante de la chaîne $(X_n)_n$ de matrice de transition P, si :*

$$\pi = \left\{ \pi.P \quad avec \quad \pi \geq (\phi_1, \phi_1, \cdots, \phi_i, \cdots) \quad et \quad \sum_i \pi_i = 1 \right\} \tag{1.9}$$

L'équation (1.9) est équivalente à :

$$\left(\pi_i \sum_{j \in E} \pi_j p_{ji} \right)_{i=1,2,\cdots}$$

Interprétation : π_i désigne la proportion de temps passé par la chaîne dans l'état e_i lorsque le temps d'évolution de la chaîne devient infiniment long.

Théorème 1.4.1. *[98] Soit π la distribution stationnaire d'une chaîne de Markov si $\pi(0) = \pi$ alors $\forall k, \pi(k) = \pi$ i.e la loi $\pi(k)$ des états au temps k est invariante pour tout k.*

Démonstration. $\pi(k) = \pi(0)P^k$, donc si $\pi(0) = \pi$, $\pi(1) = \pi P = \pi$ et par récurrence, $\pi(k) = \pi$

Théorème 1.4.2. *[98] Si π est une distribution stationnaire, alors pour tout état récurrent e_i :*

$$\pi_i = \frac{1}{\mu_i},$$

où μ_i est le temps moyen de retour à e_i.

Théorème 1.4.3. *[98] Une chaîne transiente infinie n'admet pas de distribution stationnaire.*

Démonstration. On démontre par l'absurde. Pour cela on suppose qu'il existe une distribution stationnaire π : alors $\pi P = \pi$. et $\pi P^n = \pi$ Pour tout état j et pour tout entier n, $\pi_j = \sum_i \pi_i p_{ij}^{(n)}$; les états étant transitoires, tous les $p_{ij}^{(n)}$ tendent vers 0. $\forall j, \pi_j = \lim_{n\to\infty} \sum_i \pi_i p_{ij}^{(n)} = \sum_i \lim_{n\to\infty} \pi_i p_{ij}^{(n)} = 0$, ce qui est contradictoire avec $\sum_j \pi_j = 1$.

Théorème 1.4.4. *[98] (Conditions nécessaires et suffisantes d'existence de distributions stationnaires)*

Aperçu pour la Démonstration. Cas 1 : Si la chaîne est irréductible, elle possède une distribution stationnaire unique $\pi = \left(\frac{1}{\mu_i}\right)_{e_i \in E}$, si et seulement si tous les états sont récurrents positifs. (μ_i est le temps moyen de retour à e_i).

Cas 2 : Si la chaîne est quelconque, elle possède une distribution stationnaire π, si et seulement si la chaîne possède au moins une classe récurrente positive. Dans le cas de plusieurs classes récurrentes positives, il existe une distribution stationnaire pour chaque classe.

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

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

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

Ainsi, la seule donnée de la matrice de transition P et de $\overrightarrow{\pi}(0)$ suffit à déterminer la loi de X_n pour tout $n \in \mathbb{N}$. Pour mieux visualiser les transitions entre états, il est souvent utile de faire un graphe de la chaîne , équivalent à la donnée de P où :

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

1.5.1 Régime permanent

1.5.1.1 Stationnarité :

Nous introduisons maintenant la notion centrale de la théorie de la stabilité du temps discret.

Définition 1.5.1.1.1. *Une distribution de probabilité stationnaire π satisfaisant*

$$\pi^T = \pi^T P \tag{1.10}$$

est appelée une distribution stationnaire de la matrice de transition P ou de la chaîne de Markov correspondante. L'équation (1.10) est équivalente à

$$\pi(i) = \sum_{j \in E} \pi(j) P_{ji} \tag{1.11}$$

pour tous les états.

1.6 Générateur infinitésimal

Nous considérons un système dynamique dont l'évolution est contrôlée par une chaîne de Markov à temps discret, le pas de temps étant supposé arbitrairement petit, $\epsilon > 0$. La matrice de transition $P(\epsilon)$ est proche de l'identité car on suppose que le système ≪evolue peu≫ pendant le temps $\epsilon > 0$. On suppose donc que l'on a au premier ordre :

$$P(\epsilon) = I + \epsilon A + o(\epsilon)$$

L'équation d'état du système $\pi(t+\epsilon) = \pi(t)P(\epsilon)$ devient

$$\frac{1}{\epsilon}(\pi(t+\epsilon) - \pi(t)) = \pi(t)A + o(\epsilon),$$

et en passant à la limite on obtient l'équation de Chapman-Kolmogorov ≪forward≫ :

$$\frac{d\pi(t)}{dt} = \pi(t)A$$

La matrice A est le générateur infinitésimal de la chaîne. Par construction, il s'agit d'une matrice carrée dont la somme des termes de chaque ligne est nulle, avec des termes positifs ou nuls en dehors de la diagonale.

On vérifie que la solution de l'équation d'état est de la forme $\pi(t) = \pi(0)e^{tA}$, avec par définition on a

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

Réciproquement, lorsque l'on veut discrétiser une chaîne de Markov initialement définie à temps continu, on déduit de la formule $\pi(t) = \pi(0)e^{tA}$, que $\pi(t+1) = \pi(t)e^{A}$ ce qui prouve que e^A est la matrice de transition de la chaîne discrétisée.

Définition 1.6.1. *La loi initiale $\pi(0)$ d'une chaîne $(X_n)_n$ définie sur l'ensemble des états est la distribution de probabilités d'occupation des états en $t = 0$:*

$$\pi(0) = (\mathbb{P}(X_0 = e_1), \mathbb{P}(X_0 = e_2), \cdots, \mathbb{P}(X_0 = e_n), \cdots).$$

La loi d'occupation des états ou loi de la chaîne au temps k, notée $\pi(k)$, est la distribution définie par :

$$\pi(k) = (\mathbb{P}(X_k = e_1), \mathbb{P}(X_k = e_2), \cdots, \mathbb{P}(X_k = e_n), \cdots).$$

.

Un état e_i est absorbant si $p_{i,i} = 1$.

1.7 Matrices des probabilités de transition d'une chaîne de Markov

L'analyse d'une chaîne de Markov concerne principalement le calcul des probabilités des réalisations possibles du processus. Ces calculs concernent les matrices des n-étapes de transitions $P^n = P_{ij}^n$. ici P_{ij}^n désigne la probabilité que le processus passe de l'état i à l'état j en n transitions. Nous disons qu'une matrice $P = (p_{ij} : i, j \in T)$ est stochastique si chaque ligne $(p_{ij} : j \in T)$ est une distribution. Si l'espace d'états E est fini la matrice (1.3) s'écrit :

$$A = \begin{pmatrix} P_{00} & P_{01} & \cdots & P_{0r} \\ P_{10} & P_{11} & \cdots & P_{1r} \\ \vdots & \vdots & \ddots & \vdots \\ P_{r0} & P_{r1} & \cdots & P_{rr} \end{pmatrix}$$

Définition 1.7.1. *Une chaîne de Markov est* **irréductible** *si et seulement si chaque état peut être atteint à partir de tout autre état.*

Définition 1.7.2. *Pour une chaîne de Markov homogène,*

$$p_{ji} = \mathbb{P}\left(X_n = i | X_{n-1} = j\right)$$

est appelé la **probabilité de transition** *de l'état j à l'état i.*

Définition 1.7.3. *Un état i est* **périodique** *si il existe un entier $\delta > 1$ de telle sorte que $P\left(X_{n+k} = i | X_n = i\right) = 0$ moins k est divisible par δ ; autrement l'état est* **apériodique.**. *Si tous les états d'une chaîne de Markov sont périodiques (respectivement apériodique), on dit que la chaîne est périodique (respectivement apériodique).*

1.8 Processus de Markov à temps continus

Dans cette section, nous introduisons la définition et les principaux Théorèmes pour les processus de Markov continus.

Un processus de Markov continu est un processus aléatoire $(X_t)_{t\in T}$ indexé par un un espace continu. Dans notre cas, il sera indexé par $\mathbb{R}^+(T=\mathbb{R}^+)$

Définition 1.8.1. *Un processus stochastique $(X_t)_{t\in\mathbb{R}^+}$ est un processus de Markov si est seulement si, pour tout $(t_1,t_2,...,t_n)\in(\mathbb{R}^+)^n$ avec $(t_1<t_2<t_3...<t_{n-1}<t_n)$ et pour tout $(i_1,i_2,...,i_n)\in E^n$,*

$$\mathbb{P}\left(X_{t_n}=i_n\left|X_{t_{n-1}}=i_{n-1},X_{t_{n-2}}=i_{n-2},...,X_{t_1}=i_1\right.\right)=\mathbb{P}\left(X_{t_n}=i_n\right|X_{t_{n-1}}=i_{n-1}\left.\right)$$

Un processus de Markov en temps continu est homogène si ne dépend pas de t. Dans ce qui suit, nous allons considérer que les processus de Markov homogènes.
Notations : On pose $p_{xy}(t)=\mathbb{P}\left([X_{t+s}=y]\,|\,[X_s=x]\right)=\mathbb{P}\left([X_t=y]\,|\,[X_0=x]\right)$ et $P(t)=(p_{xy}(t))_{x,y\in E}$
On pose aussi $\vec{\pi}(t)=P_{X_t}$(vecteur ligne de composantes $\pi_x(t)=\mathbb{P}\left([X_t=x]\right)$).

Propriété 1.8.1. *1. $P(t)$ est une matrice stochastique (c.à.d $p_{xy}(t)\geq 0$ et $\sum_y p_{xy}(t)=1$ pour tout x)*

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

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

Démonstration.

1. $p_{xy}(t)\in[0,1]$ car c'est une probabilité. De plus, la ligne x correspond à la loi $\mathbb{P}_{X_t}^{[X_0=x]}$ et on a bien
$$\sum_y p_{xy}(t)=\sum_y \mathbb{P}^{[X_0=x]}\left([X_t=y]\right)=1$$
2. Pour calculer $p_{xy}(t+s)=\mathbb{P}^{[X_0=x]}\left([X_{t+s}=y]\right)$ on va faire intervenir les différents états pouvant être occupés à l'instant t. Ainsi, on a
$$p_{xy}(t+s)=\sum_{z\in E} p_{x,z}(t)p_{z,y}(s)$$
qui est le coefficient (x,y) de la matrice produit $P(t)P(s)$.
3. $$\pi_y(t+s)=\mathbb{P}\left([X_{t+s}=y]\right)=\sum_{x\in E}\mathbb{P}\left([X_{t+s}=y]\,/\,[X_t=x]\right)\mathbb{P}\left([X_t=x]\right)$$
$$\pi_y(t+s)=\sum_{x\in E}\pi_x(t)p_{xy}(s)$$

Définition 1.8.2. *Le vecteur (ligne) V_n de X_n des probabilités d'état du système est défini par*

$$V_n=\left(\mathbb{P}\left(X_n=0\right),\mathbb{P}\left(X_n=1\right),\mathbb{P}\left(X_n=2\right),..\right)$$

Définition 1.8.3. *Si X_n a le même vecteur de distribution X_k pour tous $(k;n)\in\mathbb{N}^2$, alors la chaîne de Markov est stationnaire.*

Théorème 1.8.1. *Soit $(X_n)_{n\in\mathbb{N}}$ une chaîne de Markov. Nous supposons que X_n est irréductible, apériodique et homogène. La chaîne de Markov peut posséder* ***une distribution d'équilibre*** *également appelée* ***distribution stationnaire****, soit un vecteur de distribution π (avec $\pi = (\pi_0, \pi_1, \cdots)$) satisfaisant*

$$\pi P = \pi, \tag{1.12}$$

et

$$\sum_{k\in E} \pi_k = 1. \tag{1.13}$$

Si nous pouvons trouver un vecteur π équations satisfaisant (1.12) et (1.13), la distribution est unique et

$$\lim_{n\to+\infty} \mathbb{P}\left(X_n = i | X_0 = j\right) = \pi_i$$

de sorte que π est la distribution limite de la chaîne de Markov.

Théorème 1.8.2. ***(Ergodicité)****. Soit $(X_n)_{n\in\mathbb{N}}$ une chaîne de Markov. Nous supposons que X_n est irréductible, apériodique et homogène. Si une chaîne de Markov possédant une distribution d'équilibre, alors la proportion de temps que la chaîne de Markov passe dans l'état i pendant la période $[0, n]$ converge vers π_i avec $n \to +\infty$, qui est :*

$$\lim_{n\to+\infty} \frac{1}{n} \sum_{k=0}^{n} \mathbb{1}_{X_k=i} = \pi_i$$

1.9 Processus de Poisson

Les arrivées des clients à un système de file d'attente sont caractérisées par l'ensemble des instants ou dates d'arrivée de chaque client. La collection de ces dates d'arrivée peut être modéliser par le processus de Poisson.

1.9.1 Définitions et propriétés élémentaires

Il y a trois façons naturelles de décrire un tel processus

- On peut, tout d'abord, encoder une réalisation d'un tel processus par une collection $0 < T_1(\omega) < T_2(\omega) < \ldots$ de nombres réels positifs, correspondant à la position des points sur $\mathbb{R}_+$. Il est pratique de poser également $T_0 = 0$.
- Une seconde façon de coder une réalisation revient à donner, pour chaque intervalle de la forme $I = (t, t + s]$, le nombre de points $N_I(\omega)$ contenus dans l'intervalle. Si l'on utilise la notation simplifiée $N_t = N_{(0;t]}$, on aura alors $N_{(t;t+s]} = N_{t+s} - N_t$. La relation entre les variables aléatoires T_n et N_t est donc simplement

$$N_t(\omega) = sup\{n \geq 0 : T_n(\omega) \leq t\}, \quad T_n(\omega) = inf\{t \geq 0 : N_t(\omega) \leq n\}.$$

- Une troisième façon naturelle d'encoder cette information est de considérer la suite $X_1(\omega); X_2(\omega); X_3(\omega), \ldots$ de nombres réels positifs correspondant aux distances successives entre deux points. La relation entre ces variables et les T_n est donnée par

$$X_k = T_k - T_{k-1}; \quad T_k = \sum_{i=1}^{k} X_i.$$

Définition 1.9.1.1. *Soient $X_1, X_2, \ldots$ une suite de variables aléatoires satisfaisant $\mathbb{P}(X_k > 0) = 1$ pour tout $k \geq 1$. Soit $T_0 = 0$ et $T_n = \sum_{i=1}^{n} X_i$. Finalement, posons $N_t = sup\{n \geq 0 : T_n \leq t\}$. Le processus $(N_t)_{t\geq 0}$ est appelé processus de comptage.*

Remarque 1.9.1.1. *On supposera toujours par la suite qu'un processus de comptage satisfait presque sûrement $T_n \to \infty$ lorsque $n \to \infty$.*

On appelle souvent les variables aléatoires X_n les temps d'attente ou durées de vie du processus $(N_t)_{t\geq 1}$.
Le cas le plus simple, mais très important, est celui où les temps d'attente forment un processus i.i.d.

Définition 1.9.1.2. *Un processus de comptage pour lequel les temps d'attente sont i.i.d. est appelé processus de renouvellement.*

Le processus de Poisson est l'exemple le plus important de processus de renouvellement.

Définition 1.9.1.3. *Un processus de Poisson d'intensité λ est un processus de renouvellement dont les durées de vie suivent une loi $exp(\lambda)$. On note $\mathbb{P}\lambda$ la loi du processus de Poisson d'intensité λ.*

Théorème 1.9.1.1. *[97] Soit $(N_t)_{t\geq 1}$ un processus de Poisson d'intensité λ. Alors, pour tout $t, s \geq 0$,*

1. *$N_{t+s} - N_t$ suit la même loi que N_s.*
2. *$\forall 0 < t_1 < t_2 < \ldots < t_n$, les variables aléatoires $(N_{t_{i+1}} - N_{t_i})_{i=1,\ldots,n-1}$ sont indépendantes.*

Démonstration. Soit $t > 0$ fixé. Notons $X_1^t = T_{N_{t+1}} - t$ le temps restant Après t jusqu'au point suivant du processus, $X_k^t = X_{N_{t+k}}, k \geq 2$, et $T_k^t = X_1^t + \ldots + X_k^t, k \geq 1$. Évidemment,

$$N_{t+s} - N_t = n \Leftrightarrow T_n^t \leq s < T_{n+1}^t.$$

Observons à présent que l'indépendance de X_{n+1} et de T_n implique que, pour tout $x > 0$,

$$\begin{aligned}
\mathbb{P}_\lambda\left(X_1^t > x | N_t = n\right)\mathbb{P}_\lambda\left(N_t = n\right) &= \mathbb{P}_\lambda\left(X_1^t > x, T_n \leq t < T_{n+1}\right) \\
&= \mathbb{P}_\lambda\left(T_n \leq t, X_{n+1} > t + x - T_n\right) \\
&= \int_0^t dy \int_{t+x-y}^{\infty} dz f_{(T_n, X_{n+1})}(y, z) \\
&= \int_0^t dy f_{T_n}(y) \int_{t+x-y}^{\infty} dz f_{X_{n+1}}(z) \\
&= \int_0^t \mathbb{P}_\lambda\left(X_{n+1} > t + x - y\right) f_{T_n}(y) dy \\
&= e^{-\lambda x} \int_0^t \mathbb{P}_\lambda\left(X_{n+1} > t - y\right) f_{T_n}(y) dy \\
&= e^{-\lambda x} \mathbb{P}_\lambda\left(T_n \leq t, X_{n+1} > t - T_n\right) \\
&= e^{-\lambda x} \mathbb{P}_\lambda\left(N_t = n\right). \qquad (1.14)
\end{aligned}$$

En procédant de la même façon, on voit que

$$\begin{aligned}
\mathbb{P}_\lambda\left(X_1^t > x_2, \ldots, X_k^t > x_k | N_t = n\right) &= \mathbb{P}_\lambda\left(T_n \leq t, X_{n+1} > t + x_1 - T_n, X_{n+2} > x_2, \ldots, X_{n+k} > x_k\right) / \mathbb{P}_\lambda\left(N_t = n\right) \\
&= \prod_{l=2}^{k} \mathbb{P}_\lambda\left(T_n \leq t, X_{n+1} > t + x_1 - T_n\right) / \mathbb{P}_\lambda\left(N_t = n\right) \\
&= e^{-\lambda(x_1 + \ldots + x_k)}.
\end{aligned}$$

La seconde identité suivant de l'indépendance de T_n, $X_{n+1}, \ldots, X_{n+k}$, et la dernière identité de (1.14). On en déduit que, conditionnellement à $N_t = n$, les variables aléatoires $X_t^k, k \geq 1$, sont des variables aléatoires i.i.d. de loi $exp(\lambda)$. Par conséquent, la loi conjointe des variables aléatoires $T_t^k, k \geq 1$, sous $\mathbb{P}_\lambda(.|Nt = n)$ coïncide avec celle des variables aléatoires $T_k, k \geq 1$, sous $\mathbb{P}_\lambda$. On a donc

$$\begin{aligned}
\mathbb{P}_\lambda\left(N_{t+s} - N_t = n\right) &= \sum_{n\geq 0} \mathbb{P}_\lambda\left(N_{t+s} = n+k | N_t = n\right) \mathbb{P}_\lambda\left(N_t = n\right) \\
&= \sum_{n\geq 0} \mathbb{P}_\lambda\left(T_k^t \leq s < T_{k+1}^t | N_t = n\right) \mathbb{P}_\lambda\left(N_t = n\right) \\
&= \sum_{n\geq 0} \mathbb{P}_\lambda\left(T_k \leq s < T_{k+1}\right) \mathbb{P}_\lambda\left(N_t = n\right) \\
&= \mathbb{P}_\lambda\left(T_k \leq s < T_{k+1}\right) \\
&= \mathbb{P}_\lambda\left(N_s = k\right).
\end{aligned}$$

Passons à la seconde affirmation. Les arguments ci-dessus montrent que la loi conjointe des variables aléatoires $N_s - N_t = \max\{n \geq 0 : T_n^t \leq s - t\}$

sous $\mathbb{P}_\lambda\left(.|N_t = l\right)$ coïncide avec celle des variables aléatoires $N_{s-t} = \max\{n \geq 0 : T_n \leq s - t\}$ sous $\mathbb{P}_\lambda$. Posons $m_i = k_1 + \cdots + k_i, i \geq 1$. On a alors

$$\begin{aligned}
\mathbb{P}_\lambda\left(N_{t_{i+1}} - N_{t_i} = k_i, i = 1, \ldots, n-1\right) &= \sum_{l\geq 0} \mathbb{P}_\lambda\left(N_{t_{i+1}} - N_{t_i} = k_i, i = 1, \ldots, n-1 | N_t = l\right) \mathbb{P}_\lambda\left(N_{t_1} = l\right) \\
&= \sum_{l\geq 0} \mathbb{P}_\lambda\left(N_{t_{i+1}} - N_{t_i} = m_i, i = 1, \ldots, n-1 | N_t = l\right) \mathbb{P}_\lambda\left(N_{t_1} = l\right) \\
&= \sum_{l\geq 0} \mathbb{P}_\lambda\left(N_{t_{i+1}-t_1} = m_i, i = 1, \ldots, n-1\right) \mathbb{P}_\lambda\left(N_{t_1} = l\right) \\
&= \mathbb{P}_\lambda\left(N_{t_{i+1}-t_1} = m_i, i = 1, \ldots, n-1\right) \\
&= \mathbb{P}_\lambda\left(N_{t_2-t_1} = k_1, N_{t_{i+1}-t_1} - N_{t_2-t_1} = m_i - m_1, i = 2, \ldots, n-1\right).
\end{aligned}$$

De la même façon, puisque $t_{i+1} - t_1 - (t_2 - t_1) = t_{i+1} - t_2$,
$\mathbb{P}_\lambda\left(N_{t_2-t_1} = k_1, N_{t_{i+1}-t_1} - N_{t_2-t_1} = m_i - m_1, i = 2, \ldots, n-1\right)$

$$\begin{aligned}
&= \mathbb{P}_\lambda\left(N_{t_{i+1}} - N_{t_1} = m_i - m_1, i = 2, \ldots, n-1 | N_{t_2-t_1} = k_1\right) \times \mathbb{P}_\lambda\left(N_{t_2-t_1} = k_1\right). \\
&= \mathbb{P}_\lambda\left(N_{t_{i+1}-t_2} = m_i - m_1, i = 2, \ldots, n-1\right) \mathbb{P}_\lambda\left(N_{t_2-t_1} = k_1\right). \qquad (1.15)
\end{aligned}$$

Mais
$$\begin{aligned}
&\mathbb{P}_\lambda\left(N_{t_{i+1}-t_2} = m_i - m_1, i = 2, \ldots, n-1\right) \\
&= \mathbb{P}_\lambda\left(N_{t_3-t_2} = k_2, N_{t_{i+1}-t_2} - N_{t_3-t_2} = m_i - m_2, i = 2, \ldots, n-1\right),
\end{aligned}$$
et l'on peut donc répéter la procédure (1.15), pour obtenir finalement

$$\begin{aligned}
\mathbb{P}_\lambda\left(N_{t_{i+1}} - N_{t_i} = k_i, i = 1, \ldots, n-1\right) &= \prod_{i=1}^{n-1} \mathbb{P}_\lambda\left(N_{t_{i+1}-t_i} = k_i\right) \\
&= \prod_{i=1}^{n-1} \mathbb{P}_\lambda\left(N_{t_{i+1}} - N_{t_i} = k_i\right),
\end{aligned}$$

la dernière identité résulte de la première partie du Théorème.

Proposition 1.9.1.1. *[97] Soit $(N_t)_{t\geq 0}$ un processus de Poisson d'intensité λ. Alors, T_n suit une loi gamma (λ, n),*

$$f_{T_n}(x) = \frac{1}{(n-1)!} \lambda^n x^{n-1} e^{-\lambda x} \mathbb{1}[0;\infty)(x).$$

Démonstration. T_n est une somme de n variables aléatoires i.i.d. de loi $exp(\lambda)$. Manifestement, T_1 suit une loi $exp(\lambda)$, et celle-ci coïncide avec la loi $gamma(\lambda, 1)$. On procède par récurrence. Supposons l'énoncé vrai pour T_n. On a alors,

$$\begin{aligned} f_{T_{n+1}}(x) &= f_{T_n+X_{n+1}}(x) \\ &= \int_{-\infty}^{+\infty} f_{T_n}(u) f_{X_{n+1}}(x-u)du \\ &= \int_0^x \frac{1}{(n-1)!}\lambda^n u^{n-1} e^{-\lambda u}\lambda e^{-\lambda(x-u)}du \\ &= \frac{\lambda^{n+1}}{(n-1)!}e^{-\lambda x}\int_0^x u^{n-1}du \\ &= \frac{\lambda^{n+1}}{n!}e^{-\lambda x}x^n. \end{aligned}$$

Théorème 1.9.1.2. *[97] Soit $(N_t)_{t\geq 0}$ un processus de Poisson d'intensité λ. Alors, pour tout $t \geq s \geq 0$, $N_t - N_s$ suit une loi poisson $(\lambda(t-s))$.*

Démonstration. Il suit du Théorème 1.9.1.1 qu'il suffit de considérer le cas $s = 0$. Puisque $N_t = n \Leftrightarrow T_n \leq t < T_{n+1}$, on a immédiatement

$$\begin{aligned} \mathbb{P}_\lambda(N_t = n) &= \mathbb{P}_\lambda(T_n \leq t < T_{n+1}) \\ &= \mathbb{P}_\lambda(T_{n+1} > t) - (T_n > t) \\ &= \frac{\lambda^n}{n!}\int_t^{+\infty}\left(x^n\lambda e^{-\lambda} - nx^{n-1}e^{-\lambda x}\right)dx \\ &= \frac{\lambda^n}{n!}\int_t^{+\infty}\frac{d}{dx}\left(-x^n e^{-\lambda x}\right)dx \\ &= \frac{\lambda^n}{n!}t^n e^{-\lambda x}. \end{aligned}$$

Les Théorèmes 1.9.1.1 et 1.9.1.2 montrent que les accroissements $N_{t+s} - N_t$ d'un processus de Poisson d'intensité λ sont stationnaires, indépendants et suivent une loi de Poisson de paramètre λs. Nous allons montrer que ces propriétés caractérisent ce processus. Ceci fournit donc une définition alternative du processus de Poisson.

Théorème 1.9.1.3. *[97] Un processus de comptage $(N_t)_{t\geq 0}$ est un processus de Poisson d'intensité λ si et seulement si ses accroissements $N_{t+s} - N_t$ sont stationnaires et indépendants, et suivent une loi poisson (λs).*

Remarque 1.9.1.2. *[97] En fait, on peut montrer assez facilement qu'un processus de comptage $(N_t)_{t\geq 0}$ est un processus de Poisson (d'intensité non spécifiée) si et seulement si ses accroissements $N_{t+s} - N_t$ sont stationnaires et indépendants. Cela montre que ce processus va correctement modéliser toutes les situations où ces deux hypothèses sont approximativement vérifiées.*

Démonstration. On a déjà montré que le processus de Poisson possède les propriétés énoncées. Montrons donc que ces propriétés caractérisent ce processus. Fixons $0 = t_0 \leq s_1 \leq t_1 \leq s_2 \leq t_2 \leq \cdots \leq s_n \leq t_n$. En observant que $T_1 \in (s_1, t_1], \ldots, T_n \in (s_n, t_n]$ si et seulement si

- $N_{s_i} - N_{t_{i-1}} = 0, 1 \leq i \leq n$, (avec $t_0 = 0$),
- $N_{t_i} - N_{s_i} = 1, 1 \leq i < n$,
- $N_{t_n} - N_{s_n} \geq 1$,

et en utilisant les hypothèses sur les accroissements, on obtient

$$\begin{aligned}
\mathbb{P}\left(T_1 \in (s_1,t_1], \ldots, T_n \in (s_n,t_n]\right) &= \prod_{i=1}^{n} \mathbb{P}\left(N_{s_i} - N_{t_{i-1}} = 0\right) \prod_{i=1}^{n-1} \mathbb{P}\left(N_{t_i} - N_{s_i} = 1\right) \mathbb{P}\left(N_{t_n} - N_{s_n} \geq 1\right) \\
&= \prod_{i=1}^{n} e^{-\lambda}(s_i - t_{i-1}) \prod_{i=1}^{n-1} \lambda(t_i - s_i) e^{-\lambda(t_i - s_i)} \left(1 - e^{-\lambda(t_n - s_n)}\right) \\
&= \lambda^{n-1} \left(e^{-\lambda s_n} - e^{-\lambda t_n}\right) \prod_{i=1}^{n-1} (t_i - s_i) \\
&= \int_{s_1}^{t_1} \ldots \int_{s_n}^{t_n} \lambda^n e^{-\lambda u_n} du_n \ldots du_1.
\end{aligned}$$

La loi conjointe de $(T_1, \ldots, T_n)$ possède donc la densité

$$f_{(T_1,\ldots,T_n)}(u_1, \ldots, u_n) = \begin{cases} \lambda^n e^{-\lambda u_n} & \text{si} \quad 0 < u_1 < \cdots < u_n, \\ 0 & \text{sinon.} \end{cases}$$

Déterminons à présent la densité de la loi conjointe de $(X_1, \ldots, X_n)$. La fonction de répartition conjointe est donnée par

$$\begin{aligned}
\mathbb{P}\left(X_1 \leq x_1, \ldots, X_n \leq x_n\right) &= \mathbb{P}\left(T_1 \leq x_1, T_2 - T_1 \leq x_2 \ldots, T_n - T_{n-1} \leq x_n\right) \\
&= \int_0^{x_1} \int_{u_1}^{u_1 + x_2} \cdots \int_{u_{n-1}}^{u_{n-1} + x_n} f_{(T_1,\ldots,T_n)}(u_1, \ldots, u_n) du_n \ldots du_1,
\end{aligned}$$

et la densité conjointe est donc donnée par

$$\begin{aligned}
\frac{\partial^n}{\partial x_1 \cdots \partial x_n} \mathbb{P}\left(X_1 \leq x_1, \ldots, X_n \leq x_n\right) &= f_{(T_1,\ldots,T_n)}(x_1, x_1 + x_2, \ldots, x_1 + \cdots + x_n) \\
&= \lambda^n e^{-\lambda(x_1 + \cdots + x_n)}.
\end{aligned}$$

On reconnait la densité conjointe de n variables aléatoires i.i.d. de loi $exp(\lambda)$.

1.9.2 La distribution de Poisson

La distribution de Poisson de paramètre $\lambda > 0$ est donnée par :

$$p_k = \frac{e^{-\lambda} \lambda^k}{k!} \quad \text{pour} \quad k = 0, 1, 2, \cdots \tag{1.16}$$

Soit X une variable aléatoire ayant la distribution de Poisson dans (1.16) nous évaluons la moyenne, où premier moment, par :

$$\begin{aligned}
E[X] &= \sum_{k=0}^{+\infty} k p_k \\
&= \lambda
\end{aligned}$$

Pour trouver la variance, on détermine d'abord

$$\begin{aligned}
\mathbb{E}[X(X-1)] &= \sum_{k=2}^{+\infty} k(k-1) p_k \\
&= \lambda^2,
\end{aligned}$$

Alors

$$\begin{aligned}\mathbb{E}\left[X^2\right] &= \mathbb{E}[X(X-1)] + \mathbb{E}[X] \\ &= \lambda^2 + \lambda\end{aligned}$$

tandis que

$$\begin{aligned}\sigma_X^2 &= Var[X] \\ &= \lambda^2 + \lambda - \lambda^2 = \lambda.\end{aligned}$$

1.9.3 Distribution exponentielle

Soit t une variable aléatoire continue avec $t \geq 0$ qui suit une distribution exponentielle. La densité de probabilité de t est $f(t) = \alpha e^{-\alpha t}$ et la distribution cumulée correspondante est $F(t) = 1 - e^{-\alpha t}$. L'espérance et la variance de t sont $\mathbb{E}(t) = 1/\alpha$, et $Var(t) = 1/\alpha^2$, respectivement.

1.9.4 Relation entre la distribution Exponentielle et la distribution de Poisson

La densité de probabilité d'une distribution exponentielle est $f(t) = \alpha e^{-\alpha t}$ Supposons τ est exponentielle avec une espérance $1/\alpha$, et n est de Poisson de moyenne α. On a :

$$\begin{aligned}\mathbb{P}(\tau > t) &= 1 - F(t) \\ &= e^{-\alpha t} \\ &= P(n = 0 \quad \text{en} \quad t) \\ &= P(0, t)\end{aligned}$$

Notons $P(n, t)$ la probabilité d'avoir n unités dans le temps t.
$P(0, t) = e^{-\alpha t}$
$P(1, t) = \int_{\tau=0}^{t} P(0, \tau) f(1 - \tau) d\tau = \alpha t e^{-\alpha t}$
$P(2, t) = \int_{\tau=0}^{t} P(1, \tau) f(1 - \tau) d\tau = (\alpha t)^2 e^{-\alpha t}/2!$
...
$P(n, t) = \int_{\tau=0}^{t} P(n - 1, \tau) f(1 - \tau) d\tau = (\alpha t)^n e^{-\alpha t}/n!$

1.9.5 Propriété sans mémoire de la distribution exponentielle

Quand une variable aléatoire t est exponentielle, la densité de probabilité est $f(t) = \alpha e^{-\alpha t}$, et la distribution cumulée correspondante est $F(t) = 1 - e^{-\alpha t}$.
Pour un accroissement de temps h, la probabilité que t est supérieur à h devient $P(t > h) = e^{-\alpha h}$.
De plus pour $t = (t' + h)$, la probabilité de t est plus grande que $(t' + h)$ est $P(t > (t' + h)) = e^{-\alpha(t' + h)}$
La probabilité conditionnelle de $t > (t' + h)$ sachant $t > t'$ est
$P(t > t' + h | t > t') = e^{-\alpha(t' + h)}/e^{-\alpha t'} = e^{-\alpha h}$
Du fait que les deux probabilités sont les mêmes, la distribution exponentielle est appelée une distribution de probabilité sans mémoire.

1.9.6 Distribution cumulative pour un petit accroissement h

Nous considérons le temps t qui suit la distribution exponentielle, et observons, pour un accroissement de temps particulier h, la distribution cumulative de h devient $F(h) = 1 - e^{-\alpha h}$.
Notez l'expression de $F(h)$ peuvent être convertis en utilisant l'équation. (2.9) ci-dessus dans ce qui suit

Définition 1.9.6.0.2. *Un processus de Poisson est une chaîne de Markov en temps continu $(Nt)_{t\geq 0}$ a valeurs dans $\mathbb{N}$ de générateur infinitésimal A :*

$$\begin{cases} a_{i,i} = -\lambda \\ a_{i,i+1} = \lambda \\ a_{i,j} = 0 \quad sinon \end{cases}$$

On dit alors que le processus est de densité, ou de taux λ, avec sa matrice de transition :

$$A = \begin{pmatrix} -\lambda & \lambda & & & (0) \\ & -\lambda & \lambda & & \\ & & -\lambda & \lambda & \\ & & & \ddots & \ddots \\ (0) & & & & \ddots \end{pmatrix}$$

On représente schématiquement ce processus par le graphe de transition de la figure 2.1. Ce graphe de transition se déduit directement du générateur infinitésimal A. C'est une notion très classique pour l'étude des processus de sauts, nous y reviendrons plus loin.

Avec des notations évidentes, on a donc : $A = \lambda U - \lambda I$. Puisque I et U commutent, on a :

$$e^{At} = e^{-\lambda t} e^{\lambda t U}$$

mais il est clair que $e^{-\lambda t I} = e^{-\lambda t} I$, donc :

$$e^{At} = e^{-\lambda t} e^{\lambda t U} = e^{-\lambda t} \sum_{n=0}^{+\infty} \frac{(\lambda t)^n}{n!} U^n$$

D'où les probabilités de transition sont :

$$p_{ij}(t) = e^{-\lambda t} \sum_{n=0}^{+\infty} \frac{(\lambda t)^n}{n!} (U)_{ij}^n$$

Or U a des puissances successives très simples : les coefficients non nuls sont tous au-dessus de la diagonale, donc :

$$p_{ij}(t) = 0 \quad \forall j < i$$

et pour $j \geq i$, la seule matrice U^n pour laquelle le terme (i, j) est non nul est U^{j-i}. Ceci donne :

$$p_{ij}(t) = e^{-\lambda t} \frac{(\lambda t)^{j-i}}{(j-i)!}, \quad \forall j \geq i$$

1.9.7 Processus non homogènes

Le taux λ dans un processus de Poisson $X(t)$ est la constante de proportionnalité de la probabilité qu'un évènement se produise au cours d'un petit intervalle arbitraire, plus précisément,

$$\begin{aligned}\mathbb{P}\{X(t+h)-X(t)=1\} &= \frac{(\lambda h)e^{-\lambda h}}{1!}\\ &= (\lambda h)\left(1-\lambda h+\frac{1}{2}\lambda^2 h^2-\cdots\right)\\ &= \lambda h+o(h)\end{aligned}$$

où $o(h)$ désigne un terme général et sans précision d'ordre inférieur à h.

1.10 Processus de naissance et de mort

Les processus Markoviens de saut les plus simples sont les processus de naissance et de mort. Ils servent à modéliser l'évolution d'une population au cours du temps.

Définition 1.10.1. (Processus de naissance et de mort) On appelle processus de naissance et de mort toute chaîne de Markov en temps continu a valeurs dans $\mathbb{N}$ et dont le générateur infinitésimal A est de la forme :

$$a_{ij} \neq 0 \Longleftrightarrow j \in \{i-1, i, i+1\}$$

Dans de tels processus les seules transitions possibles à partir de i sont, soient vers $i+1$ en cas de naissance ou vers $i-1$ en cas de décès. On dénote les taux de transition à droite par $\lambda_i, i \in \mathbb{N}$, et les taux de transition à gauche par $\mu_i, i \in \mathbb{N}^*$ (comme le processus est à valeurs dans $\mathbb{N}$, on a forcément $\mu_0 = 0$). On appelle les λ_i taux de naissance et les μ_i taux de mort. Autrement dit, le générateur d'un processus de naissance et mort de taux de naissance $\lambda_i, i \in \mathbb{N}$ et de taux de mort $\mu_i, i \in \mathbb{N}^*$ est donné par une matrice tridiagonale :

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

avec comme graphe :

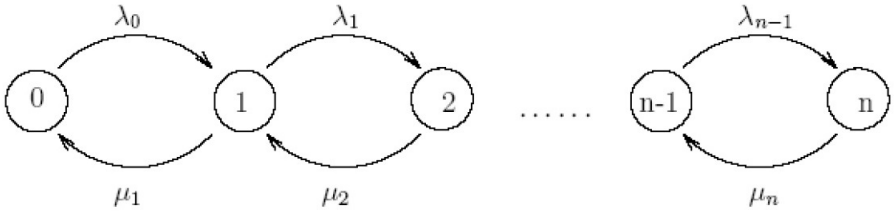

FIGURE 1.1 – Graphe de transition d'un processus de naissance et de mort.

Ce graphe représente les transitions d'un état à un autre. La transition vers la droite représente une naissance et celle vers la gauche représente une mort.

- Si tous les λ_i sont nuls, on parle de processus de mort.
- Si tous les μ_i sont nuls, on parle de processus de naissance.

Comme les λ_i et les μ_i sont tous strictement positifs, tous les états communiquent entre eux et la chaîne est irréductible, d'où le processus admet une distribution stationnaire π. Pour la trouver, on résout le système d'équation $\pi A = 0$ qui donne :

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

D'où il vient :

$$\pi_n = \frac{\lambda_0\lambda_1\cdots\lambda_{n-1}}{\mu_1\mu_2\cdots\mu_n}\pi_0,$$

or

$$\sum_{i=0}^{+\infty}\pi_i = 1,$$

l'existence d'une loi stationnaire dépend de :

$$\begin{aligned} \pi(0) &= \left[1 + \sum_{i=0}^{+\infty}\prod_{j=1}^{i}\frac{\lambda_{j-1}}{\mu_j}\right]^{-1} \\ &= \left[\sum_{i=0}^{+\infty}\prod_{j=1}^{i}\frac{\lambda_{j-1}}{\mu_j}\right]^{-1}, \end{aligned}$$

avec la convention usuelle qu'un produit vide

$$\prod_{j=1}^{0} = .1$$

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

Chapitre 2

Files d'attente

La théorie des files d'attente fournit un outil très puissant et efficace pour la modélisation des systèmes admettant un phénomène d'attente. Cette théorie datent du début du XXème siècle par les travaux de l'ingénieur danois Agner Krarup Erlang (1878, 1929). Ses études sur le trafic téléphonique de Copenhague pour le mieux gérer afin de déterminer le nombre de circuits nécessaires pour fournir un service téléphonique acceptable, sont considérées comme la première brique dans cette théorie. Ensuite, les files d'attente ont été intégrés dans la modélisation des systèmes informatiques et aux réseaux de communication. Cette intégration dans ces domaines et d'autres a permit une évolution de cette théorie surtout dans l'évaluation des paramètres de performances des systèmes informatiques et aux réseaux de communication. Actuellement ce sont les applications dans le domaine de l'analyse de performance des réseaux (téléphone mobile, Internet, multimédia,$\cdots$) qui suscitent le plus de travaux.

L'étude d'un système de files d'attente consiste à calculer ces paramètres de performance afin d'évaluer son rendement, et améliorer son fonctionnement (minimiser le temps d'attente et le temps d'inactivité de l'installation).

Depuis les travaux d'Erlang [38] Un grand nombre d'applications dans tous les domaines ont été mis en oeuvre et publiées. En 1953, Kendall [58] a introduit la notation de Kendall pour décrire les caractéristiques d'un système de file d'attente. En 1957 d'une manière particulièrement élégante et efficace Jackson a traité certains réseaux de files d'attente. En 1961, Saaty [79], auteur de l'un des premiers livres complets sur la théorie des files d'attente. Ensuite c'est les contributions des mathématiciens Khintchine [59], Palm, Pollaczek et Kolmogorov qui ont vraiment poussés la théorie des files d'attente.

Dans ce chapitre, nous introduisons la terminologie de la théorie des files d'attente et de certaines définitions et notations qui sont nécessaires dans l'étude des systèmes de files d'attente (la notation de KANDELL, la formule de LITTLE...). En suite nous étudions quelque modèles de files d'attente markoviennes ($M/M/1$, $M/M/1/K$, $M/M/c$, $M/M/\infty$,) et non markovienne ($M/G/1$) et l'évaluation des paramètres de performance.

Une file d'attente est associé à l'attente, qui est une partie inévitable de la vie moderne. Ces actions vont de l'attente d'être servis dans un guichet de poste ou banque, dans une station d'essence, un bus, dans une agence, jusqu'à l'attente d'un opérateur sur un appel téléphonique, les feux de circulation dans le trafic routier$\cdots$. Même les situations de stocks et ateliers de réparation peuvent également être considérées comme des files d'attente.

La caractéristique dominante dans tous ces exemples précédents est que certaines unités appelées clients cherchent à accéder à un système, afin de recevoir un service d'une manière aléatoire dont les durées de service sont encore aléatoires. Mais parfois la demande de ce service par plusieurs clients engendre une file d'attente et les clients se dirigent vers l'espace d'attente du système.

2.1 Description d'une file simple

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.

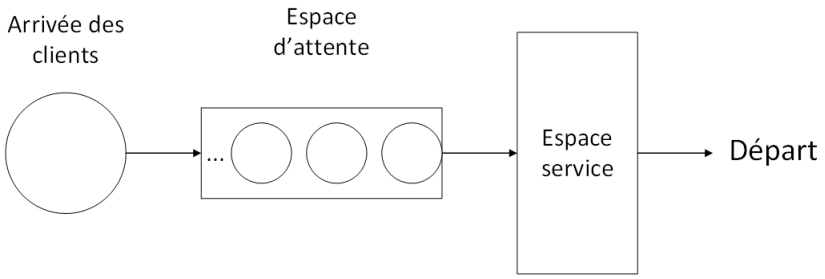

FIGURE 2.1 – Système de file d'attente simple

Afin de spécifier un système de file d'attente, on se base sur trois éléments :

- Le processus d'arrivée : On s'intéresse aux instants d'arrivés des clients dans la file. Ils sont en général aléatoires. Certaines hypothèses sont faites en se basant sur leurs lois. Tout d'abord, il n'arrive qu'un client à la fois. La deuxième hypothèse est l'homogénéité dans le temps. Cela se traduit par le fait que les temps d'interarrivée des clients sont des v.a. de même loi. Ils sont également supposés indépendants. Enfin, la loi des temps d'interarrivée est supposée connue. Le cas le plus courant est celui où cette loi est exponentielle. Dans ce cas, le modèle des temps d'arrivée des clients est un processus de Poisson. Evidemment d'autres cas peuvent se présenter : temps d'interarrivés constants, de loi uniforme, loi normale ou encore gamma.
- Le processus de service : Il comprit le nombre de serveurs et la loi probabiliste décrivant la durée des services.
- Structure et discipline de la file : La discipline de service détermine l'ordre dans lequel les clients sont rangés dans la file et y sont retirés pour recevoir un service.

2.1.0.1 Notation de Kendall

La notation suivante, appelée la notation de Kendall, est largement utilisée pour classer les différents systèmes de files d'attente :

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

avec

1. T : indique le processus d'arrivée des clients. Les codes utilisés sont :
   - M (Markov) : Interarrivées des clients sont indépendamment, identiquement distribuées selon une loi exponentielle. Il correspond à un processus de Poisson ponctuel (propriété sans mémoire).
   - D (Répartition déterministe) : les temps interarrivées des clients ou les temps de service sont constants et toujours les mêmes.
   - GI (général indépendant) : Interarrivées des clients ont une distribution générale (il n'y a aucune hypothèse sur la distribution mais les interarrivées sont indépendantes et identiquement distribuées).
   - G (général) : Interarrivées de clients ont une distribution générale et peuvent être dépendantes .
   - E_k : Ce symbole désigne un processus où les intervalles de temps entre deux arrivées successives sont des variables aléatoires indépendantes et identiquement distribuées suivant une loi d'Erlang d'ordre k.
2. Y : décrit la distribution des temps de service d'un client. les codes sont les mêmes que T.
3. C : nombre de serveurs
4. K : capacité de la file c'est le nombre de places dans le système en d'autre terme c'est le nombre maximal de clients dans le système y compris ceux en service.
5. m : population des usagers
6. Z : discipline de service c'est la façon dont les clients sont ordonnés pour être servi. Les codes utilisés sont les suivants :
   - FIFO (first in, first out) ou FCFS (first come first served) : c'est la file standard dans laquelle les clients sont servis dans leur ordre d'arrivée. Notons que les disciplines FIFO et FCFS ne sont pas équivalentes lorsque la file contient plusieurs serveurs. Dans la première, le premier client arrivé sera le premier à quitter la file alors que dans la deuxième, il sera le premier à commencer son service. Rien n'empêche alors qu'un client qui commence son service après lui, dans un autre serveur, termine avant lui.
   - LIFO (last in, first out) ou LCFS (last come, first served). Cela correspond à une pile, dans laquelle le dernier client arrivé (donc posé sur la pile) sera le premier traité (retiré de la pile). A nouveau, les disciplines LIFO et LCFS ne sont équivalentes que pour une file monoserveur.
   - SIRO (Served In Random Order), les clients sont servis aléatoirement.
   - PNPN (Priority service), les clients sont servis selon leur priorité. Tous les clients de la plus haute priorité sont servis premiers, puis les clients de priorité inférieur sont servis, et ainsi de suite.
   - PS (Processor Sharing), les clients sont servis de manière égale. La capacité du système est partagée entre les clients.

Remarque : Dans sa version courte, seuls les trois premiers symboles $T/Y/C$ sont utilisés. Dans un tel cas, on suppose que la file est régie par une discipline FIFO et que le nombre de places d'attente ainsi que celui des clients susceptibles d'accéder au système sont illimités.

2.2 Modélisation d'une file d'attente

La modélisation mathématique d'un système de files d'attente se fait le plus souvent pour déterminer les paramètres de performance à partir de la description du système. Le paramètre de performance le plus important dans une files d'attente est : le processus aléatoire $(X_t)_{t\geq o}$ enregistrant le nombre de clients

dans la file d'attente à l'instant t. Ceci est toujours prise pour inclure à la fois ceux qui sont servis et ceux qui attendent d'être servis. Et à partir de $(X_t)_{t\geq o}$ toutes les valeurs moyennes de la plupart des autres paramètres peuvent être déduites comme suit :

1. Les probabilités d'état $\pi_n(t) = \mathbb{P}([X_t = n])$ qui définissent le régime transitoire du processus $(X_t)_{t\geq o}$;
2. Le régime stationnaire du processus, défini par

$$\pi_n = \lim_{t\to\infty} \pi_n(t) = \lim_{t\to\infty} \mathbb{P}([X_t = n])$$

A son tour la distribution stationnaire du processus $(X_t)_{t\geq o}$, permet de calculer d'autres paramètres de performances du système telles que :

- Le nombre moyen L de clients dans le système ;
- Le nombre moyen L_q de clients dans la file d'attente ;
- La durée d'attente moyenne W_q d'un client ;
- La durée de séjour moyenne W dans le système (attente + service) ;
- Le taux d'occupation des postes de service ;
- Le pourcentage de clients n'ayant pu être servis ;
- La durée d'une période d'activité, c'est-à-dire de l'intervalle de temps pendant lequel il y a toujours au moins un client dans le système ...

2.2.1 Files d'attente non markoviennes

Le processus est non markovien si :

- L'une des deux quantités stochastiques : le temps des inter-arrivées ou la durée de service n'est pas exponentielle,
- Certains paramètres supplémentaires sur les problèmes de files d'attente, rendent le modèle non markovien.

Ces facteurs rendent l'étude mathématique du modèle très délicate, voire impossible, pour cela on essaye de transformer ce modèle à un processus de Markov à l'aide de l'une des méthodes d'analyse suivantes :

1. **Méthode des étapes d'Erlang :** Elle consiste à approximer la loi de probabilité qui possède une transformation de Laplace rationnelle par une loi de Cox (mélange de lois exponentielles), cette dernière possède la propriété d'absence de mémoire par étape.
2. **Méthode de la chaîne de Markov induite :** Son principe est de choisir une suite d'instants $1, 2, 3, \cdots, n$ tels que la chaîne induite $N_n, n \geq 0$, où $N_n = N(n)$, soit markovienne et homogène.

2.2.2 Files d'attente markoviennes

Ce modèle est caractérisé par des interarrivées exponentielles ainsi que les durées de service. Leur notation de Kendall est de la forme $M/M/\cdots$

2.2.3 La propriété PASTA

Avant de spécifier davantage le type de files sur lequel nous allons travailler, nous commençons par un résultat assez général.

Théorème 2.2.3.1. *[43] Supposons que les arrivées se font selon un processus de Poisson. Alors chaque client entrant verra une distribution de clients déjà présents dans le système égale à celle vue en régime permanent. Avec le théorème ergodique, cela signifie qu'en régime permanent un nouveau client verra n clients devant lui avec une probabilité égale à la proportion du temps pendant laquelle le système est occupé par n clients.*

C'est ce qui est appelé la propriété PASTA (Wolff 1982) : Poisson Arrivals See Time Averages. Notons qu'on ne fait aucune hypothèse sur la loi des services.

Démonstration. Soit N_t le nombre de clients dans le système à l'instant $t > 0$, et Δt un intervalle de temps petit. Notons aussi $A_{t-\Delta t,t}$ le nombre d'arrivées entre $t - \Delta t$ et t. Par la formule de Bayès :

$$\mathbb{P}\left(N_{t-\Delta t} = n | A_{t-\Delta t,t} \geq 1\right) = \mathbb{P}\left(A_{t-\Delta t,t} \geq 1 | N_{t-\Delta t} = n\right) \mathbb{P}\left(N_{t-\Delta t} = n\right) / \mathbb{P}\left(A_{\Delta t,t} \geq 1\right)$$

Comme un Processus de Poisson est sans mémoire (indépendance des nombres d'arrivées sur des intervalles disjoints), $\mathbb{P}\left(A_{t-\Delta t,t} \geq 1 | N_{t-\Delta t} = n\right) = \mathbb{P}\left(A_{t-\Delta t,t} \geq 1\right)$.

Maintenant lorsque Δt tend vers 0, $\mathbb{P}\left(N_{t-\Delta t} = n | A_{t-\Delta t,t}\right)$ est la probabilité que le n clients soient dans le système lorsqu'un nouveau client arrive (on ne compte pas ce nouveau client). D'après ce qui précède, cette probabilité est égale à

$$\lim_{t \to +\infty} \mathbb{P}\left(N_{t-\Delta t} = n\right) = \mathbb{P}\left(N_t = n\right)$$

Cela signifie que la probabilité pour qu'il y ait n clients dans la file à l'instant t est indépendante du fait que t soit un moment d'arrivée. En passant à la limite ($t \to +\infty$), en régime permanent, cela signifie qu'un nouveau client voit n clients devant lui avec une probabilité égale à la proportion de temps pendant laquelle le système est occupé par n clients. En effet, la première probabilité est :

$$\lim_{t \to +\infty} \lim_{\Delta t \to 0} \mathbb{P}\left(N_{t-\Delta t} = n | A_{t-\Delta t,t} \geq 1\right)$$

et la seconde :

$$\lim_{t \to +\infty} \mathbb{P}\left(N_t = n\right)$$

2.2.4 Caractéristiques d'un système de files d'attente markoviennes

La plupart du temps, l'arrivée des clients à une file simple est supposée décrite par un processus de renouvellement. Le processus d'arrivée le plus simple et le plus couramment employé est le processus de Poisson. C'est un processus de renouvellement, tel que les interarrivées sont distribuées selon une loi exponentielle. On note λ le taux des arrivées : $\frac{1}{\lambda}$ est l'intervalle moyen entre deux arrivées consécutives. Nous prenons quelques exemples sur les files d'attente markoviennes.

2.2.5 Modèle d'attente $M/M/1$

2.2.5.1 Description du modèle

La file d'attente $M/M/1$ est un modèle caractérisé par des arrivées suivants un processus de Poisson de taux λ, temps de service exponentiel de paramètre μ et un seul serveur. Les clients arrivent à la station selon un processus de Poisson de taux λ, si le serveur est vide le client est pris en charge immédiatement sinon il rejoint la file d'attente (de capacité illimitée et discipline FIFO), les temps des interarrivées sont indépendants. Ces clients reçoivent le service selon une distribution exponentielle de paramètre μ. Ces temps de service sont également supposés indépendants. En outre, toutes les variables aléatoires concernés sont censés être indépendantes les unes des autres. Les clients sont servis dans leur ordre d'arrivée. Cette

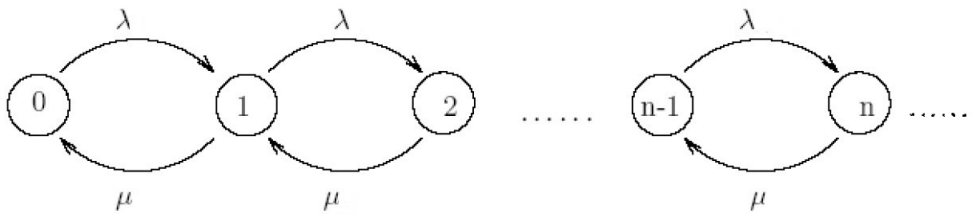

FIGURE 2.2 – Graphe de transition d'une file $M/M/1$

file est un cas particulier du processus de naissance et de mort, dans lequel les probabilités de naissance et de mort sont constantes, voir figure (2.2).

Soit $N(t)$ le nombre de clients dans le système à l'instant t.

Le processus $(N(t); t \geq 0)$ est un processus de naissance et de mort avec un taux de naissance $\lambda_i = \lambda$, pour tout $i \geq 0$ et le taux de mortalité $\mu_i = \mu$ pour tout $i \geq 1$.

En raison de la distribution exponentielle des temps interarrivée et des temps de service, le processus $(N(t); t \geq 0)$ est un processus de Markov. D'autre part, étant donné que la probabilité d'avoir deux évènements (départs, arrivées) dans l'intervalle de temps $(t; t+h)$ est $o(h)$, nous avons

$$\begin{aligned}
\mathbb{P}(N(t+h) = i+1 | N(t) = i) &= \lambda h + o(h) \quad \forall i \geq 0, \\
\mathbb{P}(N(t+h) = i-1 | N(t) = i) &= \mu h + o(h) \quad \forall i \geq 1, \\
\mathbb{P}(N(t+h) = i | N(t) = i) &= 1 - (\lambda + \mu)h + o(h) \quad \forall i \geq 1, \\
\mathbb{P}(N(t+h) = i | N(t) = i) &= 1 - \lambda h + o(h) \quad \text{for} \quad i = 0, \\
\mathbb{P}(N(t+h) = j | N(t) = i) &= o(h) \quad \text{si} \quad |j - i| \geq 2.
\end{aligned}$$

Cela montre que $(N(t); t \geq 0)$ est un processus de naissance et de la mort.
Soit $\pi(i), i \geq 0$, la fonction de distribution du nombre de clients dans le système à l'état d'équilibre.
Les équations d'équilibre pour ce processus sont :

$$\begin{aligned}
\lambda \pi(0) &= \mu \pi(1) \\
(\lambda + \mu)\pi(i) &= \lambda \pi(i-1) + \mu \pi(i+1) \quad \forall i \geq 1
\end{aligned}$$

Nous supposons dans tout le reste sauf mention contraire que,

$$\rho = \frac{\lambda}{\mu} \tag{2.1}$$

La quantité ρ mentionne l'intensité du trafic.

Les probabilités d'état pour un régime stationnaire du processus est donnée par :

$$\begin{cases} \pi_i = \pi_0 \rho^i, \\ \pi_0 = \dfrac{1}{\sum\limits_{n=0}^{+\infty} \rho^i} = 1 - \rho. \end{cases}$$

Donc

$$\pi(i) = (1 - \rho)\rho^i \quad \forall i \geq 0.$$

Tous les paramètres de performance sont calculés dans le cas où la file d'attente est stable ($\rho < 1$).

Débit *d*

Le débit du système (noté d) représente la probabilité pour que la file ne soit pas vide donc

$$d = \mathbb{P}([\text{file non vide}])\mu = \sum_{i=1}^{\infty} \pi(i)\mu = [1 - \pi_0]\mu = \lambda.$$

Ici $d = \lambda$ car $\lambda_i = \lambda$ pour tout $i \geq 0$. Une autre façon de voir les choses est de remarquer que le service s'effectue avec un taux μ dans chaque état où le système contient au moins un client.
On retrouve bien que si la file est stable, le débit moyen de sortie est égal au débit moyen d'entrée.

Taux d'utilisation du serveur U

Le taux d'utilisation du serveur (noté U) est la probabilité pour que le serveur de la file soit occupée

$$U = \sum_{i=1}^{+\infty} \pi_i = 1 - \pi_0 = \rho = \frac{\lambda}{\mu}.$$

Nombre moyen de clients L

Le nombre moyen de clients (noté L) est calculé à partir des probabilités stationnaires de la façon suivante :

$$\begin{aligned} L &= \sum_{i=1}^{+\infty} i\pi_i, \\ &= \sum_{i=1}^{+\infty} i(1-\rho)\rho^i, \\ &= \rho(1-\rho)\frac{d}{d\rho}\left(\frac{1}{1-\rho} - 1\right). \end{aligned}$$

Soit

$$L = \frac{\rho}{1-\rho}.$$

Temps moyen de séjour W

Le temps moyen de séjour W se calcule en utilisant la loi de Little :

$$W = \frac{L}{d} = \frac{1}{\mu(1-\rho)},$$

qui peut se décomposer en :

$$W = \frac{1}{\mu} + \frac{\rho}{\mu(1-\rho)}.$$

On en déduit le temps moyen passé dans la file d'attente W_q :

$$W_q = \frac{\rho}{\mu(1-\rho)}.$$

Nombre moyen de clients dans la file d'attente L_q :

$$L_q = \lambda W_q = \frac{\rho^2}{1-\rho}.$$

2.2.6 Modèle d'attente $M/M/1/K$

2.2.6.1 Description du modèle

Dans la pratique, les files d'attente sont toujours finie. Dans ce cas, quand un client arrive alors qu'il y a déjà K clients présents devant lui dans le système, il est perdu, (par exemple, les appels téléphoniques).

Ce système est connu sous le nom de file $M/M/1/K$.
Ce modèle de files d'attente est identique à la file $M/M/1$ sauf que la capacité de la file est k, est finie. Soient, comme dans le cas $M/M/1$, λ le taux du processus de Poisson pour les arrivées et μ le paramètre de la distribution exponentielle pour les temps de service. L'espace d'états E est maintenant fini : $E = \{0, 1, 2, \cdots, K\}$.
Donc ce processus est considéré comme un processus de naissance et de mort avec :
un taux de naissance $\lambda_i = \lambda$, pour tout $i < K$ et le taux de mortalité $\mu_i = \mu$ pour tout $i \neq 0$.

Soit $\pi_{(i)}, i = 0, 1, \cdots, K$, la probabilité pour qu'il ait i clients dans le système à l'instant t.

$$\begin{aligned} \lambda\pi_0 &= \mu\pi_1 \\ (\lambda+\mu)\pi_{(i)} &= \lambda\pi_{(i-1)} + \mu\pi_{(i+1)} \quad \text{pour} \quad i = 1, 2, \cdots, K-1, \\ \lambda\pi(k-1) &= \mu\pi(k). \end{aligned}$$

Le calcule de $\pi_{(i)}$ se fait de la manière suivante :

$$\begin{aligned} \pi_i &= \frac{(1-\rho)\rho^i}{1-\rho^{K+1}} \quad \text{pour} \quad i \leq K, \\ \pi_{(i)} &= 0 \quad \text{pour} \quad i > K. \end{aligned}$$

Cas particulier $\rho = 1$
Dans ce cas

$$\begin{aligned} \pi_i &= \frac{1}{K+1} \quad \text{pour} \quad i \leq K, \\ \pi_{(i)} &= 0 \quad \text{pour} \quad i > K. \end{aligned}$$

Débit *d.* On peut calculer le débit du système de deux manières équivalentes : soit par le calcul du taux de départ des clients en sortie du serveur, d_s, soit par le calcul du taux d'arrivée effectif des clients acceptés dans le système d_e.

1. Le débit en sortie du système est égal à μ (le débit du sortie du service) donc :

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

2. Le débit d'entrée dans la file est égal au taux d'arrivée à la station λ :

$$\begin{aligned} d_e &= \mathbb{P}([\text{file non pleine aux instants d'arrivée}])\lambda \\ &= \sum_{i=0}^{K-1} \pi_i\lambda = (1-\pi_K)\lambda = \frac{1-\rho^K}{1-\rho^{K+1}}\lambda \end{aligned}$$

comme $\rho = \frac{\lambda}{\mu}$ donc $d_e = d_s = d$

Taux d'utilisation du serveur $U(K)$:

Le taux d'utilisation du serveur (noté U) est la probabilité pour que le serveur de la file soit occupé donc au moins il y a un client dans la file

$$\begin{aligned} U(K) &= \sum_{i=1}^{K} \pi_i = 1 - \pi_0 \\ &= \rho\frac{1-\rho^K}{1-\rho^{K+1}}. \end{aligned}$$

Remarque :

$$\lim_{K\to+\infty} U(K) = \begin{cases} \rho & \text{pour} \quad \rho < 1 \\ 1 & \text{pour} \quad \rho > 1 \end{cases}$$

qui représente le taux d'utilisation du serveur dans le cas de la file $M/M/1$

Nombre moyen de clients L :

$$\begin{aligned} L &= \sum_{i=0}^{K} i\pi_i = \frac{1-\rho}{1-\rho^{K+1}} \sum_{i=0}^{K} i\rho^i \\ &= \frac{\rho(1-\rho)}{1-\rho^{K+1}} \sum_{i=1}^{K} i\rho^{i-1} \\ &= \frac{\rho(1-\rho)}{1-\rho^{K+1}} \frac{d}{dp}\left(\frac{1-\rho^{K+1}}{1-\rho} - 1\right) \\ &= \frac{\rho(1-\rho)}{1-\rho^{K+1}} \frac{1-(K+1)\rho^K + K\rho^{K+1}}{(1-\rho)^2} \\ &= \frac{\rho}{1-\rho} \frac{1-(K+1)\rho^K + K\rho^{K+1}}{1-\rho^{K+1}}. \end{aligned}$$

Pareil que le taux d'utilisation du serveur, lorsque $K \to +\infty$ on obtient

$$L = \frac{\rho}{1-\rho}.$$

qui représente le nombre moyen de clients pour la file $M/M/1$.

Temps moyen de séjour W :

Le calcul du temps moyen de séjour W dans la file se fait en appliquant la formule de Little qui relie Nombre moyen de clients L et le débit d de la file :

$$W = \frac{L}{d}.$$

2.2.7 Modèle d'attente $M/M/s$

2.2.7.1 Description du modèle

Pour ce modèle de file d'attente, il existe s serveurs identiques et indépendants les uns des autres, et une salle d'attente de capacité infinie. Le processus d'arrivée des clients dans la file est un processus de Poisson de taux λ et le temps de service d'un client est une variable aléatoire exponentielle de paramètre μ.

Dans ce cas aussi, le processus modélisant le nombre de clients dans le système est un processus de naissance et de la mort avec :

$$\lambda_i = \lambda$$

$$\mu_i = \begin{cases} i\mu & \text{pour} \quad i = 1, 2, \cdots, s-1 \\ s\mu & \text{pour} \quad i \geq s \end{cases}$$

Pour calculer $\pi_{(i)}$, le système doit être stabile ($\lambda < s\mu$ donc $\rho < s$) qui exprime le fait que le nombre moyen de clients qui arrivent à la file par unité de temps doit être inférieur au nombre moyen de clients

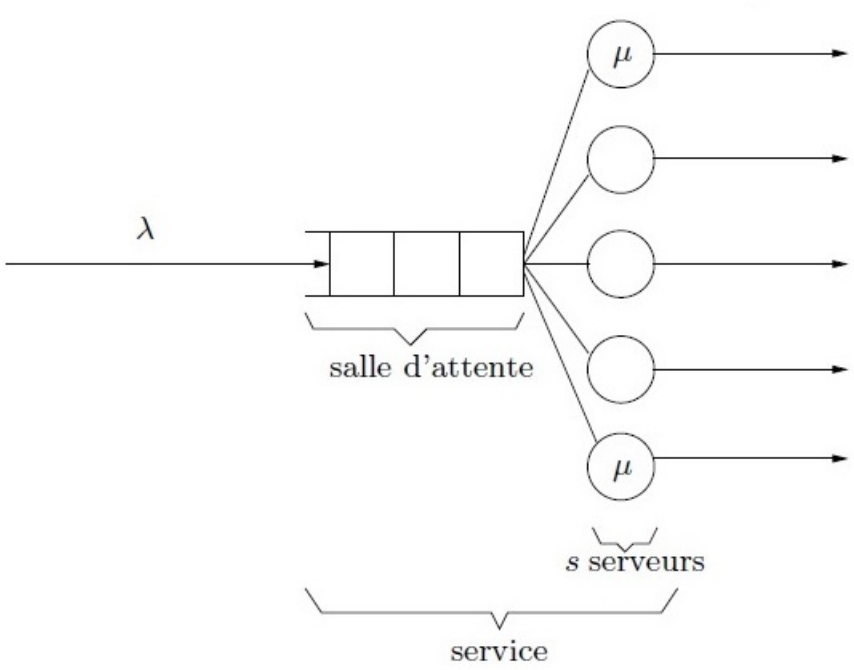

FIGURE 2.3 – Représentation schématique d'une file $M/M/s$

que les serveurs de la file sont capables de traiter par unité de temps. Alors

$$\pi_i = \begin{cases} \pi_0 \frac{\rho^i}{i!} & \text{pour} \quad i = 1, 2, \cdots, s-1 \\ \pi_0 \rho^i \frac{s^{s-i}}{s!} & \text{pour} \quad i \geq s \end{cases}$$

avec

$$\pi_0 = \left[\sum_{i=0}^{s-1} \frac{\rho^i}{i!} + \left(\frac{\rho^s}{s!} \right) \left(\frac{1}{1 - \frac{\varrho}{s}} \right) \right]^{-1}$$

Remarque :

pour la valeur de $s = 1$ on obtient :

$$\pi_i = (1 - \rho)\rho^i$$

qui représente le cas de la file $M/M/1$

Débit d Quand le système contient moins de s clients, ces derniers sont pris en charge par les s serveurs à un taux $i\mu$ et avec $s\mu$ un taux dans chaque état où le système contient plus de s clients :

$$d = \sum_{n=1}^{s-1} \pi_n n\mu + \sum_{n=s}^{+\infty} \pi_n s\mu.$$

En remplaçant les expressions obtenues pour les probabilités π_i et π_0, on trouve bien que la file est stable :

$$d = \lambda.$$

Pour des raison de calcul pour la file $M/M/s$, il est plus simple de calculer d'abord le temps moyen de séjour et en suite déduire le nombre moyen de clients.

Temps moyen de séjour W **:** Le temps moyen de séjour d'un client est composé du temps moyen dans la file d'attente, et le temps moyen de service. Il suffit alors d'appliquer la loi de Little.

$$W = W_q + W_S = \frac{L_q}{d} + \frac{1}{\mu} = \frac{L_q}{\lambda} + \frac{1}{\mu}.$$

Il reste alors à calculer le nombre moyen de clients en attente dans la file, L_q :

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

On en déduit l'expression du temps moyen de séjour

$$W = \frac{\rho^s}{\mu(S-1)!(S-\rho)^2}\pi_0 + \frac{1}{\mu}.$$

Nombre moyen de clients L : La loi de Little permet de trouver le nombre moyen de clients dans la file :

$$L = W \times d = W \times \lambda = \frac{\rho^{s+1}}{(s-1)!(s-\rho)^2}\pi_0 + \rho.$$

2.2.8 Modèle d'attente $M/M/\infty$

2.2.8.1 Description du modèle

Pour ce modèle de file d'attente, le système est composé d'un nombre illimité de serveurs identiques et indépendants les uns des autres. Dès qu'un client arrive, il est immédiatement servi (c'est le cas ou il n'y a pas d'attente). Dans cette file les clients arrivent à des instants $0 < t_1 < t_2 < \cdots$ formant un processus de poisson de taux λ et les temps de service sont exponentiels de taux μ (pour tous les serveurs). Le taux de transition d'un état n quelconque vers l'état $n-1$ est égal à $n\mu$ et correspond au taux de sortie d'un client parmi les n clients en service. De même, le taux de transition d'un état n vers l'état $n+1$ est égal à λ qui correspond au taux d'arrivée d'un client, donc c'est un processus de naissance et de mort avec :

$$\lambda_k = \lambda \quad \text{et} \quad \mu_k = k\mu \quad \text{et} \quad k = 0, 1, 2, \cdots$$

Soit π_n la probabilité stationnaire d'être dans l'état n. Les équations d'équilibre nous donnent :

$$\pi_{n-1} = \pi_n n\mu \quad \text{et} \quad n = 1, 2, \cdots$$

avec

$$\pi_0 = \frac{1}{\sum_{n=0}^{+\infty}\frac{1}{k!}(\frac{\lambda}{\mu})^k} = e^{-\frac{\lambda}{\mu}}$$

et on obtient finalement :

$$\pi_n = \frac{1}{n!}\left(\frac{\lambda}{\mu}\right)^n e^{-\frac{\lambda}{\mu}} \quad \text{pour} \quad n = 1, 2, \cdots$$

par le calcul de π_n on trouve les différents paramètres de performances comme suit :

Débit d le service s'effectue avec un taux $n\mu$ dans chaque état où le système contient n clients :

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

Nombre moyen de clients L :

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

Temps moyen de séjour W :

La loi de Little permet d'écrire :

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

2.2.9 Modèle d'attente $M/G/1$

2.2.9.1 Description du modèle

Pour ce modèle, on étudie une file d'attente composée d'un seul serveur. Le processus d'arrivée de clients est un processus de Poisson d'intensité λ, mais le temps de service indépendants identiquement distribués de loi générale. On notera Y la fonction de répartition de cette loi et, lorsqu'elle est bien définie, y sa fonction de densité. On supposera que la discipline de service est FIFO.
Nous allons commencer par établir l'expression de la probabilité p_0 pour que le serveur soit inoccupé. Le nombre moyen d'arrivées pendant la durée d'un service doit être strictement inférieur à 1, i.e. $\lambda\mathbb{E}(Y) < 1$.

Proposition 2.2.9.1.1. *[43] Dans une file $M/G/1$ avec arrivée Poissonnienne de taux λ et services indépendants de durée moyenne $\mathbb{E}(Y)$, si p_0 est la probabilité pour que le serveur soit inoccupé, alors :*

$$1 - p_0 = \lambda\mathbb{E}(Y)$$

Démonstration. Notons X_n le nombre de clients présents dans le système après le $n$ème départ, A_n le nombre d'arrivées pendant le service du $n$ème client.
Alors :

$$X_{n+1} = \lambda_n + A_n + u(X_n)$$

où $u(X_n) = 1$ si $X_n > 1$ (il y a un départ car dans ce cas le serveur est effectivement occupé)
et $u(X_n) = 0$ si $X_n = 0$ (serveur inactif)
Or d'après la propriété PASTA le nombre moyen de clients vu par un nouvel arrivant ne dépend pas de l'instant d'arrivée, donc $\mathbb{E}(X_n) = \mathbb{E}(X_{n+1})$.
Par conséquent : $\mathbb{E}(A_n) = \mathbb{E}(u(X_n))$.
Or $\mathbb{E}(u(X_n)) = 0.\mathbb{P}(X_n = 0) + 1.\mathbb{P}(X_n \geq 1) = 1 - p_0$, et $\mathbb{E}(A_n)$ est le nombre moyen d'arrivées pendant un service, que l'on calcule par :

$$\begin{aligned}
\mathbb{E}(A_n) &= \sum_{k=0}^{n} k \int_0^{+\infty} \mathbb{P}(k \text{ arrivées pendant } t) y(t) dt \\
&= \sum_{k=0}^{n} k \int_0^{+\infty} e^{-\lambda t} \frac{(\lambda t)^k}{k!} y(t) dt \\
&= \int_0^{+\infty} e^{-\lambda t} y(t) \lambda t \left(\sum_{k=0}^{+\infty} \frac{(\lambda t)^k}{k!} y(t) \right) dt.
\end{aligned}$$

Or $\sum_{k=0}^{+\infty} \frac{(\lambda t)^k}{k!} = e^{\lambda t}$,
donc

$$\mathbb{E}(A_n) = \int_{t=0}^{+\infty} \lambda t y(t) dt = \lambda\mathbb{E}(Y)$$

On a bien $1 - p_0 = \lambda\mathbb{E}(Y)$ A partir de là on peut déterminer le nombre moyen de clients dans le système et le temps moyen d'attente.

Proposition 2.2.9.1.2. *[43] Dans une file $M/G/1$, le temps moyen résiduel de service (définie comme la durée restant à un serveur pour terminer son service lorsqu'un nouveau client arrive dans la file) $\overline{t_r}$ vérifie :*

$$\overline{t_r} = \frac{1}{2}\mathbb{E}(Y) + \frac{Var(y)}{2\mathbb{E}(Y)}$$

où Y est la durée d'un service.

Nous allons établir des formules établissant le temps d'attente moyen et le nombre moyen de clients dans une file $M/G/1$. Notons $\tau = \lambda\mathbb{E}(Y)$. On suppose que $\tau < 1$ pour que la file d'attente ne s'allonge pas indéfiniment. D'après la propriété PASTA, un nouveau clients voit $\overline{N_f}$ clients dans la file, qui seront servis avant lui sous la discipline de service FIFO. Chacun de ces clients aura un temps de service moyen de $\mathbb{E}(Y)$.

Donc le temps moyen d'attente d'un nouveau client est : $\overline{N_f}E(Y)$, augmenté de son temps résiduel moyen d'attente, le temps résiduel d'attente étant la différence entre l'instant de la fin du service courant et l'instant d'arrivée.
Le temps résiduel d'attente est 0 si le serveur est inoccupé lorsque le nouveau client arrive (avec une certaine probabilité $1 - \tau$ d'après la proposition et la propriété PASTA) et égal au temps résiduel de service sinon (avec une probabilité complémentaire τ).

Donc : $\overline{T_f} = \overline{N_f}.E(Y) + \tau.\overline{t_r}$. D'autre part, avec la loi de Little, $\overline{N_f} = \lambda\overline{T_f}$, d'ou

$$\overline{T_f} = \frac{\tau\overline{t_r}}{1 - \lambda E(y)}$$

et

$$\overline{N_f} = \frac{\lambda\overline{t_r}}{1 - \lambda E(y)}$$

Comme le temps moyen de séjour $\overline{T}$ (temps d'attente + temps de service) vérifie $\overline{T} = \overline{T_f} + E(Y)$ et le nombre moyen de client dans le système vérifie $\overline{N} = \lambda\overline{T}$, on conclut :

Proposition 2.2.9.1.3. *[43] (Formules de Pollaczek Khinchine). Le nombre moyen de clients dans une file $M/G/1$ avec des arrivées Poissonniennes de taux λ et une durée de service Y de loi quelconque (admettant au moins une espérance et une variance.) est :*

$$\overline{N} = \tau + \frac{\tau^2(1 + Var(y)/\mathbb{E}(y)^2)}{2(1 - \tau)}$$

où $\tau = \lambda\mathbb{E}(y)$

Le temps moyen de séjour dans la file est :

$$\overline{T} = \mathbb{E}(y)(1 + \frac{\tau(1 + Var(y)/\mathbb{E}(y)^2)}{2(1 - \tau)}$$

Remarque : Pour le cas particulier $M/M/1$, on a $Var(y)/\mathbb{E}(y)^2 = 1$ et on retrouve les formules déjà établies.

Chapitre 3

Systèmes de files d'attente avec dérobade, abandon et rappels

Dans divers domaines, 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. Ces systèmes qui contiennent des clients impatiens ont fait des pertes considérables à l'économie de plusieurs firmes.

Dans ce chapitre on s'intéresse aux files d'attente :

- Avec dérobade Un client impatient qui decide ne pas joindre le système.
- Avec abandon Après un temps passé dans la file, le client impatient décide de quitter le système sans avoir le service.
- Avec feedback Le client insatisfait de la qualité du service, décide de quitter la file pour redemander ou compléter son service après un temps aléatoire.

Les files d'attente avec dérobade, abandon ou les deux peuvent être considérées comme un phénomène caractéristique de la vie contemporaine. Plusieurs mathématiciens se sont intéréssés par l'étude de ces modèles, tels que Haight, 1957, [46] qui a étudié le modèle d'attente $M/M/1$ avec dérobade, puis, en 1959, [48] il a traité ce même modèle $M/M/1$ avec abandon. Ce sont ensuite Ancker et Gafarian [7, 8] qui ont étudié l'éffet de la combinaison entre le dérobade et l'abondon sur une file $M/M/1/N$. Abou-EI-Ata et Hariri [1] ont considéré le modèle $M/M/c/N$ avec dérobade, abandon et plusieurs serveurs. Enfin Wang et Chang [75] ont fait l'extention de ce travail pour étudier $M/M/c/N$ avec dérobade, abandon et serveurs en panne.
Actuellement ce sont les applications dans le domaine de l'analyse de performance des réseaux (téléphone mobile, Internet, multimédia,....) qui suscitent le plus de travaux.

3.1 Système de files d'attente avec dérobade

Les systèmes de files 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.

3.1.1 Description du modèle

Supposons qu'un client qui arrive à un système et trouve n clients devant lui, entre avec la probabilité p_n ou part avec la probabilité $q_n = 1 - p_n$. Si la longueur de la file d'attente décourage les clients, alors p_n serait une fonction décroissante de n. Comme un cas particulier, s'il y a une salle d'attente de capacité C, nous pourrions supposer que

$$p_n = \begin{cases} 1 & \text{pour} \quad n < C, \\ 0 & \text{pour} \quad n \geq C. \end{cases}$$

indiquant que, une fois la salle d'attente est remplie, les clients n'ont plus la possibilité d'entrer dans le système.
Soit $X(t)$ le nombre de clients dans le système à l'instant t, et le processus d'arrivée des clients dans le système est Poissonien de taux λ. Un client qui arrive de l'extérieur et trouve n clients dans le système devant lui entre avec la probabilité p_n alors le processus de naissance appropriés est

$$\lambda_n = \lambda p_n \quad \text{pour} \quad n = 0, 1, \cdots .$$

Et dans le cas d'un seul serveur, $\mu_n = \mu$ pour $n = 1, 2, \cdots$ et nous pouvons évaluer la distribution stationnaire π_k de la longueur de la file d'attente. Dans les systèmes avec abandon, tous les clients qui arrivent entrent dans le système, et certains sont perdus. Le taux d'entrée des clients dans le système à l'état stationnaire est donné par

$$\lambda_i = \lambda \sum_{n=0}^{\infty} \pi_n p_n.$$

La vitesse à laquelle les clients sont perdus est

$$\lambda \sum_{n=0}^{\infty} \pi_n q_n.$$

et la fraction des clients perdus dans le long terme est donnée par

$$\sum_{n=0}^{\infty} \pi_n q_n.$$

Maintenant nous examinons en détail le cas d'un système $M/M/s$. A son arrivée un client entre dans le système, si et seulement si un serveur est libre. Pour celà nous avons

$$\lambda_k = \begin{cases} \lambda & \text{pour} \quad k = 0, 1, \cdots, s-1, \\ 0 & \text{pour} \quad k = s, \end{cases}$$

et

$$\mu_k = k\mu \quad \text{pour} \quad k = 0, 1, \cdots, s$$

Afin de déterminer la distribution, nous avons

$$\pi_k = \frac{1}{k!} \left(\frac{\lambda}{\mu} \right)^k \quad \text{pour} \quad k = 0, 1, \cdots, s$$

et par conséquent

$$\pi_k = \frac{\frac{1}{k!} \left(\frac{\lambda}{\mu} \right)^k}{\sum_{j=0}^{s} \frac{1}{j!} \left(\frac{\lambda}{\mu} \right)^j} \quad \text{pour} \quad k = 0, 1, \cdots, s.$$

La proportion des clients perdus est $\pi_s q_s = \pi_s$; d'où $q_s = 1$.

3.1.2 Taux de service

Dans la même veine, on peut envisager un système dont le taux de service dépend du nombre de clients dans le système. Par exemple, un second serveur peut être ajouté à un système à un seul serveur à chaque fois que la longueur de la file d'attente dépasse un point critique. Si les arrivées sont de Poisson et les taux de service sont sans mémoire, les paramètres de naissance et de mort appropriés sont

$$\lambda_k = \lambda \quad \text{pour} \quad k = 0, 1, \cdots$$

et

$$\mu_k = \begin{cases} \mu & \text{pour} \quad n \leq \eta \\ 2\mu & \text{pour} \quad n > \eta \end{cases}$$

Plus généralement, nous considérons des arrivées Poissonniennes $\lambda_k = \lambda \quad \text{pour} \quad k = 0, 1, \cdots$ et les taux de service arbitraires μ_k pour $k = 0, 1, \cdots$. La distribution stationnaire dans ce cas, est donnée par

$$\pi_k = \frac{\pi_0 \lambda^k}{\mu_1 \mu_2 \cdots \mu_k} \quad \text{pour} \quad k \geq 1, \tag{3.1}$$

où

$$\pi_0 = \left[1 + \sum_{k=1}^{\infty} \frac{\lambda^k}{\mu_1 \mu_2 \cdots \mu_k} \right]^{-1}. \tag{3.2}$$

3.1.3 Modèle de files d'attente $M/M/1$ avec dérobade

Dans ce modèle, un client arrivant à un 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. Ce modèle a été traité dans [52]
Donc le nombre de clients dans le système est un processus de naissance et de mort avec :
des taux de naissances

$$\lambda_k = \lambda b_k \quad k = 0, 1, \cdots$$

tel que

$$b_k = \frac{1}{k+1} \quad k = 0, 1, \cdots$$

ainsi

$$\mathbb{P}_k = \frac{\rho^k}{k!} \mathbb{P}_0 \quad k = 0, 1, \cdots,$$

puis en utilisant la condition de normalisation, nous obtenons

$$\mathbb{P}_k = \frac{\rho^k}{k!} e^{-\rho} \quad k = 0, 1, \cdots$$

La condition de stabilité ici est $\mathbb{E}(\rho) < \infty$, on n'a pas besoin de la condition $\rho < 1$ comme dans la file $M/M/1$.
Les paramètres de performance sont comme suit :
Taux d'utilisation du serveur U :

$$\begin{aligned} U_s &= 1 - \mathbb{P}_0 = 1 - e^{-\rho}, \\ &= \frac{\mathbb{E}(\delta)}{\frac{1}{\lambda} + \mathbb{E}(\delta)}, \end{aligned}$$

avec $\mathbb{E}(\delta)$ est la période moyenne de l'occupation du serveur.
Par conséquent

$$\mathbb{E}(\delta) = \frac{1}{\lambda}.\frac{1-e^{-\rho}}{e^{-\rho}}.$$

Nombre moyen de clients dans le système L :

$$L = \rho.$$

$$Var(L) = \rho.$$

Nombre moyen de clients dans la file d'attente L_q :

$$L_q = L - U_s = \rho - \left(1 - e^{-\rho}\right) = \rho + e^{-\rho} - 1.$$

Temps moyen de séjour W :

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

et on déduit le temps moyen passé dans la file d'attente

$$W_q = W - \frac{1}{\mu} = \frac{1}{\mu}\left(\frac{\rho + e^{-\rho} - 1}{1-e^{-\rho}}\right).$$

Débit d : on a aussi

$$d = \sum_{k=0}^{\infty} \lambda_k P_k = \sum_{k=1}^{\infty} \mu_k P_k = \mu(1-e^{-\rho}),$$

d'où

$$d.W = \mu(1-e^{-\rho}).\frac{\rho}{\mu(1-e^{-\rho})} = \rho = L.$$

Ce dernier résultat confirme la formule de LITTLE pour ce système.

3.1.4 Modèle de files d'attente $M/M/s$ avec dérobade

Nous considérons une file d'attente $M/M/s$ avec dérobade. Le processus d'arrivée est de Poisson de taux λ et le service a une distribution exponentielle de moyenne $1/\mu$. Un client arrivant de l'extérieur, rejoint la file d'attente si et seulement s'il observe que le temps d'attente ne dépasse pas un temps fixe b. Ce modèle a été traité dans [66].

Soit $N(t)$ le nombre de clients dans le système à l'instant t. La définition de $W(t)$ implique que $W(t) = 0$ si et seulement si $N(t) \leq s-1$. Si $N = \{N(t), t \geq 0\}$ le processus subit une transition de l'état $s-1$ à l'état s à l'instant t, alors il y a un saut dans le processus $W = \{W(t), t \geq 0\}$ en même temps. Ceci est illustré par la Figure 3.1 montrant un système à 2 serveurs.

On note que, dans la trajectoire de l'échantillon représenté, le client qui arrive à l'instant T_3 dérobade. A l'instant T_1 (et T_4), le nombre de clients dans le système augmente de 1 à 2. Dans le même temps, le temps d'attente passe de 0 à un nombre positif. Il est clair que le processus de W termine un cycle de régénération du T_1 au T_4. Soit I une variable aléatoire représentant la période d'inactivité définie comme l'intervalle de temps pendant lequel $W(t) = 0$. En d'autres termes, soit t_1 l'instant de la fin du service tel que $N(t_1-) = s$. Alors,

$$I = \min_{t>0}\{t : W(t_1 + t) > 0\}$$

et

$$I = \min_{t>0}\{t : N(t_1 + t) = s\}.$$

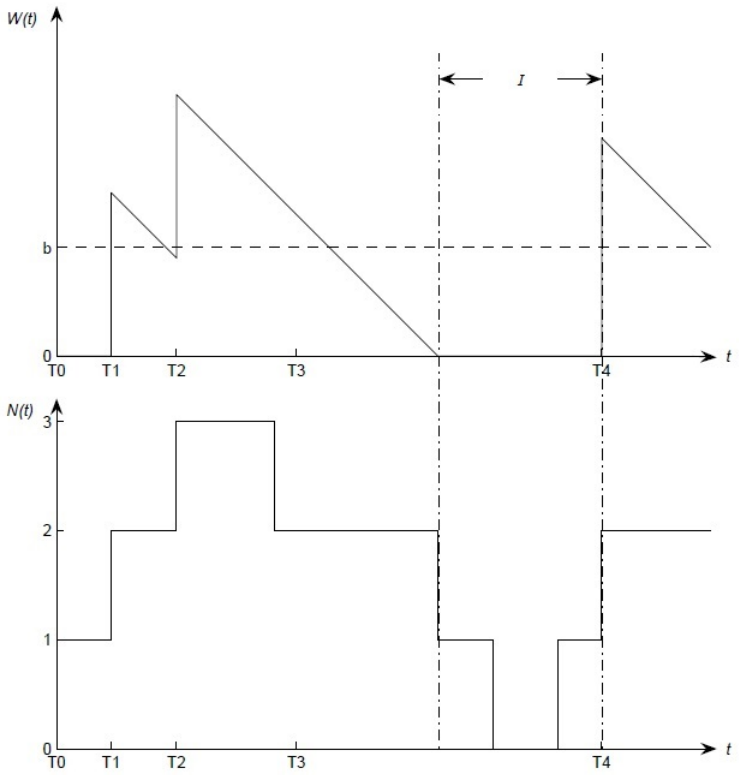

FIGURE 3.1 – Un chemin d'échantillon de $W(t)$ et $N(t)$ [66]

Théorème 3.1.4.1. *[66] La durée prévue de I est donnée par*

$$\mathbb{E}(I) = \frac{1 - p_s}{s\mu p_s}. \tag{3.3}$$

avec

$$p_s = \frac{(\lambda/\mu)^s}{s!} \sum_{i=0}^{s} \frac{(\lambda/\mu)^i}{i!}$$

Démonstration.

Considérons un système $M/M/s$ avec le taux d'arrivée λ et la moyenne des temps de service $1/\mu$. De toute évidence, le temps entre deux périodes consécutives lorsque le système est plein a la même distribution que I. La durée moyenne de chaque période du système complet est clairement $1/(s\mu)$. De la théorie du processus de renouvellement, nous obtenons :

$$\frac{1/(s\mu)}{1/(s\mu) + E(I)} = p_s$$

où p_s est la probabilité que le système $M/M/s/s$ est plein.

Tenant compte du cycle de régénération de T_1 à T_4, (illustré sur la figure 3.1). Le cycle se compose d'une période d'activité intense (où $W(t) > 0$) et une période de repos (où $W(t) = 0$). Il est facile de voir que l'évolution du processus de W pendant la période occupée est stochastiquement identique à celui d'un système avec dérobade à serveur unique avec un taux d'arrivée λ et des temps de service $exp(s\mu)$. Soit $\tilde{W}(t)$ le temps d'attente et $\tilde{N}(t)$ le nombre de clients à l'instant t dans ce système. La même règle de dérobade appliquée, à savoir, un client arrivant à l'instant t pénètre si et seulement si $\tilde{W}(t-) \leq b$. Les clients arrivent selon un processus de Poisson de taux d'arrivée $\tilde{\lambda}(t)$ en fonction de $\tilde{N}(t)$ comme suit :

$$\tilde{\lambda}(t) = \begin{cases} \gamma & \text{si} \quad \tilde{N}(t) = 0 \\ \lambda & \text{sinon} \end{cases}$$

avec $\gamma = 1/\mathbb{E}(I)$. Désignons la fonction de distribution cumulative (cdf) limitant du processus de W et le processus $\tilde{W}$ comme suit.

$$F(x) = \lim_{t \to +\infty} \mathbb{P}\{W(t) \leq x\}, x \geq 0;$$

$$\tilde{F}(x) = \lim_{t \to +\infty} \mathbb{P}\{\tilde{W}(t) \leq x\}, x \geq 0$$

soit $F(0) = c$ et $\tilde{F}(0) = \tilde{c}$
et soit $f(x) = \frac{dF(x)}{dx}$; $\tilde{f}(x) = \frac{d\tilde{F}(x)}{dx}$; $x > 0$ la pdf.

Notons que pour un système à un seul serveur le temps d'attente est égal au temps d'occupation de ce système.

Nous appliquons la même méthode utilisé dans Liu et Kulkarni [66] et les résultats du le Théorème 3.1.4.2 qui donne les équations d'équilibre et de normalisation satisfaite par $f(x)$ et le Théorème 3.1.4.3 qui donne l'expression de $\tilde{f}(x)$ de façon explicite.

Théorème 3.1.4.2. *[66] La loi des probabilités à l'état d'équilibre de $\tilde{f}(x)$ du processus $\widetilde{W}$ satisfait :*

$$\tilde{f}(x) = \lambda \int_0^{x \wedge b} f(u) e^{-s\mu(x-u)} du + \tilde{c}\gamma e^{-s\mu x},$$

$$\int_0^{\infty} \tilde{f}(x) dx + \tilde{c} = 1$$

ou $x \wedge b = min(x, b)$ on pose $\rho = \frac{\lambda}{s\mu}$ l'intensité du traffic.

Théorème 3.1.4.3. *[66]*

La loi des probabilités du processus $\widetilde{W}$ est :

$$\tilde{f}(x) = \begin{cases} \widetilde{c}\gamma e^{-(s\mu-\lambda)x} & si \quad 0 < x < b \\ \widetilde{c}\gamma e^{\lambda b} e^{-s\mu x} & si \quad x \geq b. \end{cases} \tag{3.4}$$

Avec

$$\tilde{c} = \begin{cases} \left[\gamma \left(\frac{1}{s\mu-\lambda} - e^{-(s\mu-\lambda)b} \frac{\lambda}{(s\mu-\lambda)s\mu}\right) + 1\right]^{-1} & si \quad \rho \neq 1 \\ \frac{\lambda}{\lambda+\gamma+\lambda\gamma b} \quad si \quad \rho = 1 \end{cases} \tag{3.5}$$

3.2 Etude d'un modèle de file $M/M/1$ avec abandon

La théorie des files d'attente avec abandon joue un rôle important dans la modélisation de beaucoup de problèmes de la vie réelle. Ces applications sont utilisées dans plusieurs secteurs (informatique, communication, l'industrie, ...) ou encore dans les secteurs de la santé et des sciences médicales, ...etc.

Dans cette section, on s'intéresse à l'étude d'un modèle de systèmes de files d'attente avec abandon. Après un temps passé dans la file, le client impatient décide de quitter le système sans avoir le service. Ce modèle a été traité dans [22]

Le modèle est basé sur les hypothèses suivantes :

- Le système est composé d'un serveur unique, les clients arrivent selon un processus de Poisson de taux qui dépend du nombre de clients présents dans le système i.e. $\frac{\lambda}{(n+1)}$
- Les temps de service sont indépendants les uns des autres, distribués selon une loi exponentielle de paramètre μ.

- Les clients sont servis selon la discipline FIFO.
- La capacité du système est limitée de taille N.
- Le client arrivant au système attend un certain moment son service. Si ce service ne commence pas, alors il devient impatient et il peut quitter le système avec une probabilité p, ou rejoint la file avec une probabilité $q = 1 - p$. Le temps d'abandon suit une distribution exponentielle de paramètre ξ

Soit $P(n)$ la probabilité qu'il y ait n clients dans le système à l'instant t. Le processus est considéré comme un processus de naissance et de mort, donc les équations différentielles du modèle à l'état stable du système sont,

$$\frac{d}{dt}P_0(t) = -\lambda P_0(t) + \mu P_1(t) \tag{3.6}$$

$$\frac{d}{dt}P_n(t) = -\left[\left(\frac{\lambda}{n+1}\right) + \mu + (n-1)\xi p\right] P_n(t) + (\mu + n\xi p)P_{n+1}(t) + \left(\frac{\lambda}{n}\right) P_{n-1}(t); \quad n = 1, \cdots, N-1 \tag{3.7}$$

$$\frac{d}{dt}P_N(t) = -\left[\mu + (N-1)\xi p\right] P_N(t) + \left(\frac{\lambda}{N}\right) P_{N-1}(t) \tag{3.8}$$

En état d'équilibre, $\lim_{t\to+\infty} P_n(t) = P_n$ et, $\frac{d}{dt}P_n(t) = 0$ et donc, la solution des équations (3.6) à (3.8) donne :

$$0 = -\lambda P_0 + \mu P_1 \tag{3.9}$$

$$0 = -\left[\left(\frac{\lambda}{n+1}\right) + \mu + (n-1)\xi p\right] P_n + (\mu + n\xi p)P_{n+1} + \left(\frac{\lambda}{n}\right) P_{n-1}; \quad n = 1, \cdots, N-1 \tag{3.10}$$

$$0 = -\left[\mu + (N-1)\xi p\right] P_N + \left(\frac{\lambda}{N}\right) P_{N-1}. \tag{3.11}$$

En résolvant les équations de (3.9) à (3.11), nous obtenons

$$P_n(t) = -\left[\frac{1}{n!}\prod_{k=1}^{n}\frac{\lambda}{\mu + (k-1)\xi p}\right] P_0; \quad n = 1, \cdots, N \tag{3.12}$$

En utilisant la condition de normalisation, $\sum_{n=0}^{N} P_n = 1$, nous obtenons

$$P_0 = \frac{1}{1 + \sum_{n=1}^{N}\left(\frac{1}{n!}\prod_{k=1}^{n}\frac{\lambda}{\mu+(k-1)\xi p}\right)}. \tag{3.13}$$

Et par conséquent, la distribution stationnaire du processus permet de calculer d'autres paramètres de performances du système telles que :

Nombre moyen de clients dans le système L_s :

$$L_s = \sum_{n=1}^{N} n\left[\frac{1}{n!}\prod_{k=1}^{n}\frac{\lambda}{\mu + (k-1)\xi p}\right] P_0.$$

Nombre moyen de clients dans la file L_q :

$$L_q = \left[\sum_{n=1}^{N} n\left[\frac{1}{n!}\prod_{k=1}^{n}\frac{\lambda}{\mu + (k-1)\xi p}\right] P_0\frac{\lambda}{\mu}\right].$$

Temps moyen de séjour dans le système W_s :

$$W_s = \left[\frac{1}{\lambda}\sum_{n=1}^{N} n\left[\frac{1}{n!}\prod_{k=1}^{n}\frac{\lambda}{\mu + (k-1)\xi p}\right]\right] P_0.$$

Temps moyen de séjour dans la file W_q :

$$W_q = \left[\frac{1}{\lambda}\sum_{n=1}^{N} n\left[\frac{1}{n!}\prod_{k=1}^{n}\frac{\lambda}{\mu+(k-1)\xi p}\right]P_0 - \frac{1}{\mu}\right].$$

Le nombre moyen de clients servis E(client servis) :

$$E(\text{client servis}) = \sum_{n=1}^{N} n\mu P_n$$

$$E(\text{client servis}) = \sum_{n=1}^{N} n\mu\left[\frac{1}{n!}\prod_{k=1}^{n}\frac{\lambda}{\mu+(k-1)\xi p}\right]P_0.$$

Taux d'abandon R_{abond} :

$$R_{\text{abond}} = \lambda\sum_{n=0}^{N} P_n - E(\text{client en attente})$$

$$R_{\text{abond}} = \lambda - \sum_{n=1}^{N} n\mu\left[\frac{1}{n!}\prod_{k=1}^{n}\frac{\lambda}{\mu+(k-1)\xi p}\right]P_0.$$

Le nombre de clients en attente E(client en attente) :

$$E(\text{client en attente}) = \frac{\sum_{n=2}^{N}(n-1)P_n}{\sum_{n=2}^{N}P_n}$$

$$E(\text{client en attente}) = \frac{\sum_{n=2}^{N}(n-1)\left[\frac{1}{n!}\prod_{k=1}^{n}\frac{\lambda}{\mu+(k-1)\xi p}\right]P_0}{\sum_{n=2}^{N}\left[\frac{1}{n!}\prod_{k=1}^{n}\frac{\lambda}{\mu+(k-1)\xi p}\right]P_0}.$$

La probabilité de la période d'occupation $\mathbb{P}$(période d'occupation) :

$$\mathbb{P}(\text{période d'occupation}) = \sum_{n=1}^{N}\left[\frac{1}{n!}\prod_{k=1}^{n}\frac{\lambda}{\mu+(k-1)\xi p}\right]P_0$$

Où P_0 est la valeur calculée en (3.13).

3.3 Système de files d'attente avec rappel

Les systèmes de files d'attente avec rappel sont des systèmes utilisés dans la modélisation des réseaux de télécommunication et dans les systèmes informatiques. Après son arrivée à une station donnée, un client qui trouve tous les serveurs occupés et ne peut pas recevoir le service immédiatement, quitte le système pour être rappeler ultérieurement à des instants aléatoires jusqu'à satisfaction de sa demande. C'est le cas pour les appels téléphoniques par exemple, entre deux rappels successifs, le client en question se trouve en orbite. Un tel système est appelé système de files d'attente avec rappel (retour).
En général un système de files d'attente avec rappels est composé de $c \geq 1$ serveurs et de $m - c$ ($m \geq c$) places d'attente. 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 trouvent 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. Entre deux rappels successifs, le client est dit "en orbite". La capacité de l'orbite peut être finie ou infinie. Le client rappelé de l'orbite, est traité de la même manière qu'un client

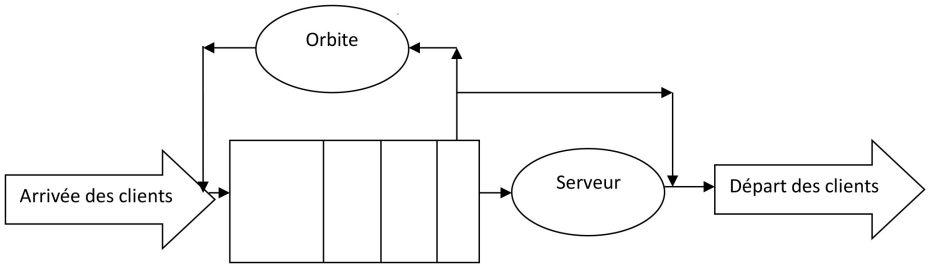

FIGURE 3.2 – système de files d'attente avec rappel

venant de l'extérieur.

Ce phénomène de files d'attente avec rappel, a fait inspirer plusieurs chercheurs, Cohen [32], Eldin [37], Hashida et Kawashima [49] et Lubacz et Roberts [54]. les articles de synthèse de nombreux travaux résument les progrès de la théorie des files d'attente, a cité : Falin et Templeton [40] ont souligné que les modèles standards de file d'attente ne prennent pas le phénomène de rappel en compte et donc ils ne peuvent pas être appliqués à la résolution d'un nombre importants de problèmes pratiques qui contient le phénomène de rappel. Après ils sont venus les travaux de Kulkarni et Liang [60],et Templeton [86]. Ce progrès est aussi signalé dans les monographies de Falin et Templeton [87] et Artalejo et Gomez [11].
Récemment un peu partout dans le monde un grand intérêt est donné aux files d'attente avec rappel par l'organisation de plusieurs conférences et workshops : Madrid (Espagne) (1998), puis Minsk (Biélorussie) (1999), Amsterdam (Pays-Bas) (2000), Cochin (Inde) (2002), Seoul (Corée du sud) (2004), Miraflores de la Sierra (Espagne) (2006), Athens (Grèce) (2008), Beijing (Chine) (2010).

3.3.1 Etude du modèle de file d'attente $M/G/1$ avec rappel

Cette partie est consacrée à l'étude du système de files d'attente $M/G/1$ avec rappel exponentiel. Ce système est composé d'un seul serveur avec les arrivées des clients Poissonniennes, et les durées de service suivent une loi générale. Nous décrivons d'abord le modèle associé à ce type de files d'attente puis nous déduisons les équations de Chapmann Kolmogorov et nous calculerons les mesures de performance de la file $M/G/1$ avec rappel exponentiel. Ce modèle a été traité dans [70]

3.3.1.1 Description du modèle

Pour ce modèle, qui est composé d'un seul serveur. le processus d'arrivée des clients dans le système est de Poisson de taux $\lambda > 0$ et la durée de service μ est de loi générale $B(x)$ et de transformée de Laplace-Stieltjes $\widetilde{B}(s)$, $Re(s) > 0$. Soient les moments $\beta_k = (-1)^k \widetilde{B}^k(0)$, l'intensité du trafic $\rho = \lambda\beta_1$. Le taux de rappel de l'orbite est exponentiellement distribuée de paramètre $\nu > 0$. La capacité de la file d'attente étant infinie.
On pose :
$C(t)$: La variable aléatoire indiquant l'état du serveur à l'instant t.
Donc

$$C(t) = \begin{cases} 0 & \text{si le serveur est libre à l'instant } t. \\ 1 & \text{si le serveur est occupé à l'instant } t. \end{cases}$$

$N(t)$: Le nombre de clients dans l'orbite à l'instant t.
$\beta(x)$ est l'intensité instantané du service donné quand le temps de service écoulé est $\eta(t) = x$.

3.3.1.2 Les équations de Chapmann Kolmogorov en régime stationnaire de la file d'attente $M/G/1$ avec rappel

Nous présentons les trois équations de Chapman Kolmogorov comme suit :

$$(\lambda + n\nu)P_{0n} = \int_0^\infty P_{1n}(x)\beta(x)dx \quad n \geq 0 \tag{3.14}$$

$$P_{1n}^{'} = (-\lambda + b(x))P_{1n}(x) + \lambda P_{1n-1}(x)dx \quad n \geq 1 \tag{3.15}$$

$$P_{1n}(0) = \lambda P_{0n}(x) + (n+1)\nu P_{0n+1} \quad n \geq 0 \tag{3.16}$$

3.3.1.3 Les caractéristiques du système

Les fonctions génératrices partielles

Nous avons utilisé la méthode des fonctions génératrices pour calculer les probabilités stationnaires P_{0n} et P_{1n} :

$$P_0(z) = \sum_{k=0}^{\infty} z^n \times P_{0n} = (1-\rho)exp\left[\frac{\lambda}{\nu}\int_1^z \frac{1-K(u)}{K(u)-u}du\right] \tag{3.17}$$

$$P_1(z) = \sum_{k=0}^{\infty} z^n \times P_{1n} = \frac{1-k(z)}{k(z)-z} \times P_0(z) \tag{3.18}$$

Le nombre moyen de clients dans le système

Après un calcul à l'aide de fonctions génératrices ; nous obtenons le nombre moyen de clients dans le système :

$$N = \rho + \frac{\lambda^2\beta}{2(1-\rho)} \tag{3.19}$$

La probabilité que le serveur est occupé

La probabilité que le serveur est occupé est ainsi :

$$P\left[C(t) = 1\right] = \rho \tag{3.20}$$

Chapitre 4

Analyse d'un modèle de files d'attente $M/M/2/N$ avec dérobade, abandon, feedback et deux serveurs hétérogènes.

Ce chapitre a fait l'objet d'une publication dans le journal Mathematical Sciences And Applications E-Notes, (2013) Volume 2 No. 2 pp. 10-21, [24].

Analysis of two heterogeneous server queueing model with balking, reneging and feedback

AMINA ANGELIKA BOUCHENTOUF[1], MOKHTAR KADI[2], ABBES RABHI[3]
[1,3] Department of Mathematics, Djillali Liabes university of Sidi Bel Abbes,
B. P. 89, Sidi Bel Abbes 22000, Algeria
E-mail :bouchentouf_amina@yahoo.fr, rabhi_abbes@yahoo.fr
[2] Department of Mathematics, Dr. Moulay Taher University, Saida 20000, Algeria
E-mail :kadi1969@yahoo.fr

Abstract.

This paper presents an analysis of an $M/M/2/N$ (capacity N) queueing system with balking, reneging, feedback and two heterogeneous servers. By using the Markov process method, we first develop the equations of the steady state probabilities. Then, we give some performance measures of the system. Finally, we present some numerical examples to demonstrate how the various parameters of the model influence the behavior of the system.

Key words and phrases. Balking, reneging, feedback, heterogeneous servers, steady state solution, performance measures.

4.1 Introduction

In real life, many queueing situations arise in which there may be a tendency for customers to be discouraged by a long queue. As a result, the customers either decide not to join the queue (i.e. balk) or depart after joining the queue without getting service due to impatience (i.e. renege). Balking and reneging are not only common phenomena in queues arising in daily activities, but also in various machine repair models.

Many practical queueing systems especially those with balking and reneging have been widely applied to many real-life problems, such as the situations involving impatient telephone switchboard customers, the hospital emergency rooms handling critical patients, and the inventory systems with storage of perishable goods [11]. In this paper, we consider an $M/M/2/N$ queueing system with balking, reneging and feedback. Queueing systems with balking, reneging, or both have been studied by many researchers. Haight [6] first considered an $M/M/1$ queue with balking. An $M/M/1$ queue with customers reneging was also proposed by Haight [7]. The combined effects of balking and reneging in an $M/M/1/N$ queue have been investigated by Ancker and Gafarian [2, 3]. Abou-EI-Ata and Hariri [1] considered the multiple servers queueing system $M/M/c/N$ with balking and reneging. Wang and Chang [16] extended this work to study an $M/M/c/N$ queue with balking, reneging and server breakdowns.

Feedback in queueing literature represents customer dissatisfaction because of inappropriate quality of service. In case of feedback, after getting partial or in- complete service, customer retries for service. In computer communication, the transmission of protocol data unit is sometimes repeated due to occurrence of an error. This usually happens because of non-satisfactory quality of service. Rework in industrial operations is also an example of a queue with feedback. Takacs [14] studied queue with feedback to determine the stationary process for the queue size and the first two moments of the distribution function of the total time spent in the system by a customer. In [5] D'Avignon and Disney studied single server queues

with state dependent feedback. Santhakumaran and Thangaraj [12] considered a single server feedback queue with impatient and feedback customers, they studied $M/M/1$ queueing model for queue length at arrival epochs and obtained result for stationary distribution, mean and variance of queue length. Thangaraj, and Vanitha [15] obtained transient solution of $M/M/1$ feedback queue with catastrophes using continued fractions, the steady-state solution, moments under steady state and busy period analysis were calculated. Ayyapan et. al [4] studied $M/M/1$ retrial queueing system with loss and feedback under non preemptive priority service by matrix geometric method. Kumar and Sharma [8] studied a single server queueing system with retention of reneged customers. Kumar and Sharma [9] studied a single server queueing system with retention of reneged customers and balking. Sharma and Kumar [13] considered a single server, finite capacity Markovian feedback queue with reneging, balking and retention of reneged customers in which the inter-arrival and service times follow exponential distribution. Mahdy El-Paoumy and Hossam Nabwey [10] studied the $M/M/2/N$ queue with general balk.

In this paper, we consider an $M/M/2/N$ queueing system with two heterogeneous servers, finite capacity Markovian feedback queueing system with reneging, balking and retention of reneged customers in which the inter-arrival and service times follow exponential distribution. The reneging times are assumed to be exponentially distributed. After the completion of service, each customer may rejoin the system as a feedback customer for receiving another regular service with probability α or he can leave the system with probability β where $\beta + \alpha = 1$. A reneged customer can be retained in many cases by employing certain convincing mechanisms to stay in the queue for completion of his service. Thus, a reneged customer can be retained in the queueing system with some probability γ and may leave the queue without receiving service with probability $\delta = 1 - \gamma$

This paper is organized as follows. At first we give a description of the queueing model. Then, we derive the steady-state equations by the Markov process method, and present a procedure for calculating the steady-state probabilities. After that, we give some performance measures of the system. Finally, some numerical examples are presented to demonstrate how the various parameters of the model influence the behavior of the system.

4.2 Model Description and Notations

Consider an M/M/2/N queue with instantaneous Bernoulli feedback with reneged customers and retention of reneged customers. Capacity of the system is taken as finite. Customers arrive at the service station one by one according to Poisson stream with arrival rate λ. There is a two servers which provide service to all arriving customers. Service times are independently and identically distributed exponential random variables with rates $\mu_i,\ \ i = 1, 2$. After completion of each service, the customer can either join at the end of the queue with probability α or he can leave the system with probability β where $\alpha + \beta = 1$. The customers both newly arrived and those that are fed back are served in order in which they join the tail of original queue. We do not distinguish between the regular arrival and feedback arrival. The customers are served according to first come, first served rule. The customer in queue (regular arrival or feedback arrival) may become impatient when the service is not available for a long time. In fact, each customer, upon arrival, activates an individual timer, which follows an exponential distribution with parameter η. This time is reneging time of an individual customer after which customer either decide to abandon the queue with probability $\delta(= 1 - \gamma)$ or retained with complimentary probabilityγ where $\delta + \gamma = 1$.We suppose that there are n customers in the system $(n > 2)$, the arriving customer joins the system with certain balking probability. It is assumed that an arriving customer balks with probability

1-$\frac{1}{n-1}$, where n is the number of customers in system and therefore joins the system with probability $\frac{1}{n-1}$.

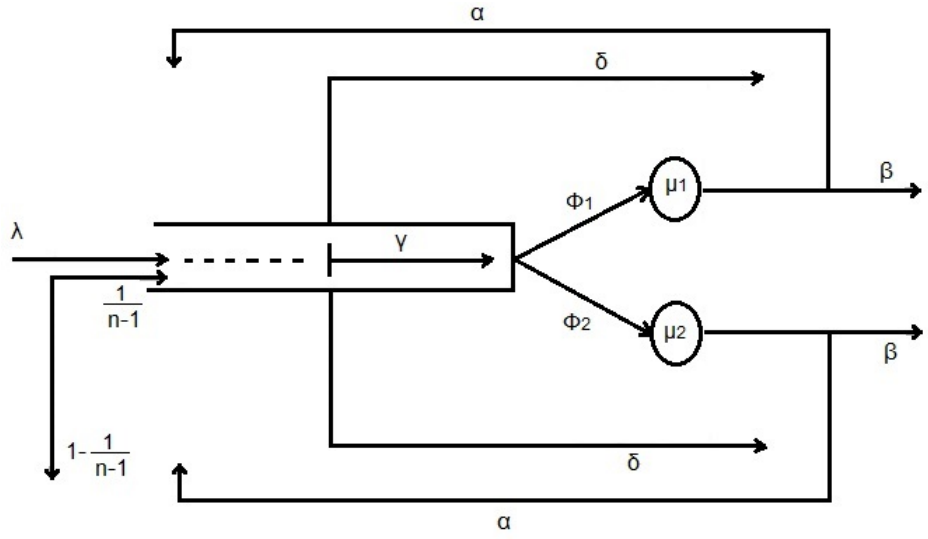

FIGURE 4.1 – The model

The customers are served according to the following discipline : i) If the two servers are busy the customers wait in the queue.
ii) If one server is free , the head customer in the queue goes to it, and ;
iii) If both servers are free, the head customer of the queue chooses server 1 with probability ϕ_1 and server 2 with probability ϕ_2, $\phi_1 + \phi_2 = 1$

4.3 The steady-state equations and their solution

In this part of paper, we derive the steady-state probabilities by the Markov process method. Let P_n be the probability that there are n customers in the system. Applying the Markov process theory, we obtain the following set of steady- state equations We define the equilibrium probabilities : $P_{00} = \mathbb{P}$(there is no customer in the system),

$P_{10} = \mathbb{P}$(there is one customer being served by server 1)

$P_{01} = \mathbb{P}$(there is one customer being served by server 2)

$P_n =$ (there are n customers in the system), $n = 2, 3, \cdots, N$.

Also, $P_0 = P_{00}$, $P_1 = P_{10} + P_{01}$ and $\delta = P_{11}$
A system of difference differential equations satisfied by the $M/M/2/N$ feedback queue with reneged customers, balking and retention of reneged customers are modeled as a continuous time Markov chain

$$\lambda P_0 = \beta\mu_1 P_{10} + \beta\mu_2 P_{01}, \quad n = 0 \tag{4.1}$$

$$n=1,\begin{cases}(\lambda+\beta\mu_1)P_{10} &= \beta\mu_2 P_{11}+\lambda\phi_1 P_0\\ (\lambda+\beta\mu_2)P_{01} &= \beta\mu_1 P_{11}+\lambda\phi_2 P_0\end{cases} \tag{4.2}$$

$$(\lambda+\beta(\mu_1+\mu_2)\delta=(\beta(\mu_1+\mu_2)+\eta\delta)P_3+\lambda P_1, \qquad n=2 \tag{4.3}$$

Instead of equation (4.3), we have

$$\left(\tfrac{\lambda}{n-1}+\beta(\mu_1+\mu_2)+(n-2)\eta\delta\right)P_n=(\beta(\mu_1+\mu_2)+(n-1)\eta\delta)\,P_{n+1}+\tfrac{\lambda}{n-2}P_{n-1},\ \ 3\le n\le N-1 \tag{4.4}$$

$$\tfrac{\lambda}{N-1}P_{N-1}=(\beta(\mu_1+\mu_2)+(N-2)\eta\delta)\,P_N,\ \ n=N \tag{4.5}$$

Solving relations (4.2) and using (4.1) one can easily deduce that :

$$P_{10}=\tfrac{\lambda}{\beta\mu_1}\left(\tfrac{\lambda+\beta(\mu_1+\mu_2)\phi_1}{2\lambda+\beta(\mu_1+\mu_2)}\right)P_0 \tag{4.6}$$

$$P_{01}=\tfrac{\lambda}{\beta\mu_2}\left(\tfrac{\lambda+\beta(\mu_1+\mu_2)\phi_2}{2\lambda+\beta(\mu_1+\mu_2)}\right)P_0 \tag{4.7}$$

Therefore

$$P_1=\tfrac{\lambda}{\beta\mu_1\mu_2}\left(\tfrac{\lambda(\mu_1+\mu_2)+\beta(\mu_1+\mu_2)(\mu_1\phi_2+\mu_2\phi_1))}{(2\lambda+\beta(\mu_1+\mu_2))}\right)P_0 \tag{4.8}$$

Solving iteratively equations (4.2)-(4.4), we get

$$P_n=\tfrac{1}{(n-1)!}\prod_{k=2}^{n}\left(\frac{\lambda}{\beta(\mu_1+\mu_2)+(k-2)\eta\delta}\right)P_1,\ \ 2\le n\le N. \tag{4.9}$$

Consequently

$$P_n=\tfrac{\lambda}{\beta\mu_1\mu_2}\left(\tfrac{\lambda(\mu_1+\mu_2)+\beta(\mu_1+\mu_2)(\mu_1\phi_2+\mu_2\phi_1))}{(2\lambda+\beta(\mu_1+\mu_2))}\right)\tfrac{1}{(n-1)!}\prod_{k=2}^{n}\left(\frac{\lambda}{\beta(\mu_1+\mu_2)+(k-2)\eta\delta}\right)P_0,\ \ 2\le n\le N. \tag{4.10}$$

Using the normalization condition, $\sum_{n=0}^{N}P_n=1$ we get

$$P_0=\left[1+\left(\tfrac{\lambda}{\beta\mu_1\mu_2}\left(\tfrac{\lambda(\mu_1+\mu_2)+\beta(\mu_1+\mu_2)(\mu_1\phi_2+\mu_2\phi_1))}{(2\lambda+\beta(\mu_1+\mu_2))}\right)\right)\left(1+\sum_{n=2}^{N}\frac{1}{(n-1)!}\prod_{k=2}^{n}\left(\frac{\lambda}{\beta(\mu_1+\mu_2)+(k-2)\eta\delta}\right)\right)\right]^{-1} \tag{4.11}$$

4.4 Performance Measures

Some important performance measures are derived.

(i) The Expected System Size

$$\begin{aligned}L_s&=\sum_{n=1}^{N}nP_n\\ &=\left[\tfrac{\lambda}{\beta\mu_1\mu_2}\left(\tfrac{\lambda(\mu_1+\mu_2)+\beta(\mu_1+\mu_2)(\mu_1\phi_2+\mu_2\phi_1))}{(2\lambda+\beta(\mu_1+\mu_2))}\right)\left(1+\sum_{n=2}^{N}\frac{n}{(n-1)!}\prod_{k=2}^{n}\left(\frac{\lambda}{\beta(\mu_1+\mu_2)+(k-2)\eta\delta}\right)\right)\right]P_0\end{aligned}$$

(ii) The Expected Queue Length

$$L_q = \sum_{n=3}^{N}(n-2)P_n$$
$$= \left[\frac{\lambda}{\beta\mu_1\mu_2}\left(\frac{\lambda(\mu_1+\mu_2)+\beta(\mu_1+\mu_2)(\mu_1\phi_2+\mu_2\phi_1))}{(2\lambda+\beta(\mu_1+\mu_2))}\right)\left(\sum_{n=3}^{N}\frac{n-2}{(n-1)!}\prod_{k=2}^{n}\left(\frac{\lambda}{\beta(\mu_1+\mu_2)+(k-2)\eta\delta}\right)\right)\right]P_0$$

The Expected Waiting Time in the System
$W_s = L_s/\lambda$

The Expected Waiting Time in the queue
$W_q = L_q/\lambda$

The Expected Number of customers Served,
$\mathbb{E}(CustomerServed) = \sum_{n=1}^{N} n\beta(\mu_1+\mu_2)P_n$

4.5 Numerical Results

In this part of this paper, we present some numerical examples to show the impact of different parameters and its relationship with the expected system size, the expected queue length, the expected waiting time in the system ,the expected waiting time in the queue and the expected number of customers served.
Firstly, Let us present the evolution of the system when λ varies from 0.1 to 1,
$N = 6$, $\delta = 0.2$ $\mu_1 = 7$, $\mu_2 = 4$, $\beta = 0.01$, $\eta = 0.1$, $\phi_1 = 0.3$, $\phi_2 = 0.7$

λ	L_s	L_q	W_s	W_q	$\mathbb{E}(CustomerServed)$
0.1	0.933977962	0.431967748	9.33977962	4.319677481	0.146387405
0.2	1.213010966	0.34877451	6.065054832	1.743872551	0.181817771
0.3	1.240585123	0.309719988	4.135283745	1.032399959	0.210161695
0.4	1.180298512	0.285826868	2.950746281	0.714567169	0.243151735
0.5	1.086840648	0.265819453	2.173681296	0.531638905	0.278733228
0.6	0.983377705	0.24701997	1.638962842	0.411699951	0.314391502
0.7	0.881352598	0.229088382	1.25907514	0.327269118	0.34835838
0.8	0.786437572	0.212222905	0.983046965	0.265278631	0.379619449
0.9	0.701064439	0.196642663	0.778960487	0.218491847	0.407736074
1	0.625821636	0.18246406	0.625821636	0.18246406	0.432651425

Figure 4.2, shows that along the increase of λ the expected number of customers in the system increases, then from the value of $\lambda = 0.4$; it starts to decrease,while the queue length, the expected waiting time in the queue and in the system decrease along the increase of λ, figure 4.3, because of balking and reneging discipline and finally the expected number of customers served increases with the increase of λ, figure 4. All these results agree absolutely with our intuition.

Now, let us present the relationship between δ (the probability to leave the queue without receiving service) and different performance measures of the system.

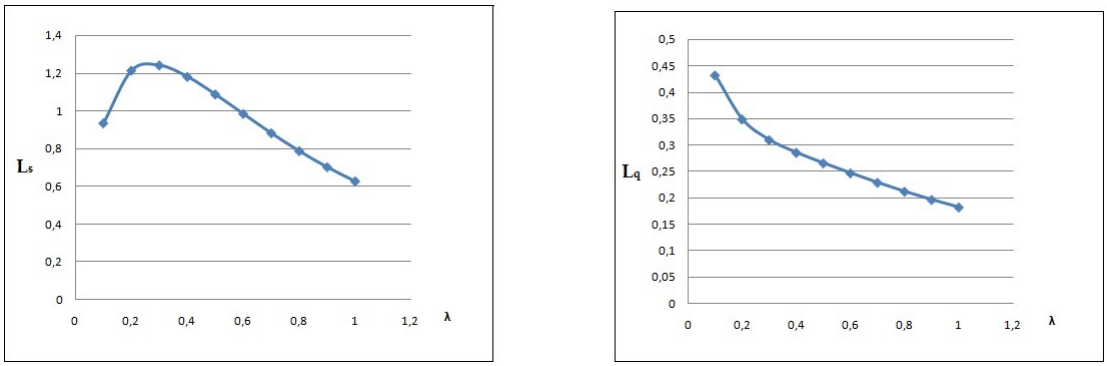

FIGURE 4.2 – λ versus L_s and λ versus L_q

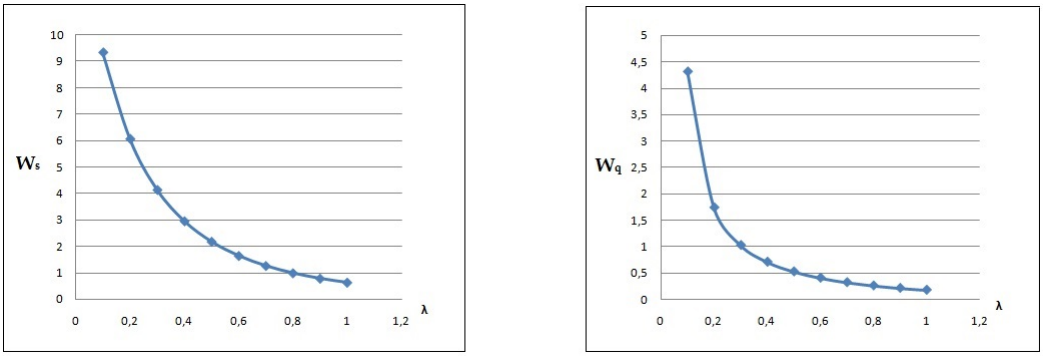

FIGURE 4.3 – λ versus W_s and λ versus W_q

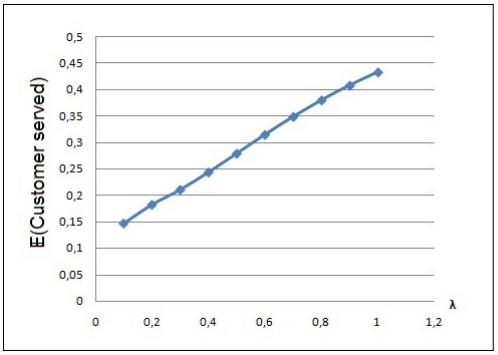

FIGURE 4.4 – λ versus E Number of customers Served

At first, we fix the maximum number of customers in the system $N = 5$, the arrival rate $\lambda = 0.5$ the service rate at server 1, $\mu_1 = 7$, the service rate at server 2, $\mu_2 = 4$ the probability that the customer leave the system without having a feedback $\beta = 0.7$, reneging time rate $\eta = 0.1$, the probability that customer chooses server 1, $\phi_1 = 0.3$, the probability that customer chooses server 2, $\phi_2 = 0.7$. Then we

vary δ from 0 to 1 in increments of 0.1. The numerical results are summarized in the following table :

δ	L_s	L_q	W_s	W_q	$\mathbb{E}(CustomerServed)$
0	0.135312293	0.132785988	0.270624587	0.265571976	1.158397488
0.1	0.1353097	0.132786683	0.2706194	0.265573366	1.158399427
0.2	0.135307114	0.132787376	0.270614228	0.265574752	1.158401363
0.3	0.135304535	0.132788068	0.270609069	0.265576136	1.158403294
0.4	0.135301962	0.132788758	0.270603924	0.265577516	1.158405222
0.5	0.135299397	0.132789446	0.270598793	0.265578892	1.158407146
0.6	0.135296838	0.132790133	0.270593676	0.265580265	1.158409066
0.7	0.135294286	0.132790818	0.270588572	0.265581635	1.158410982
0.8	0.135291741	0.132791501	0.270583481	0.265583002	1.158412894
0.9	0.135289202	0.132792182	0.270578404	0.265584365	1.158414803
1	0.13528667	0.132792862	0.270573341	0.265585725	1.158416708

For the second result, for each value of $N = 5$, $\lambda = 0.5$, $\mu_1 = 7$, $\mu_2 = 4$, $\beta = 0.25$, $\eta = 0.1$, $\phi_1 = 0.3$, $\phi_2 = 0.7$ selected, we vary δ from 0 to 1 in increments of 0.1.
The numerical results are summarized in the following table :

δ	L_s	L_q	W_s	W_q	$\mathbb{E}(CustomerServed)$
0	0,331603365	0,288830411	0,663206731	0,577660822	1,090550257
0.1	0,331480297	0,288859098	0,662960594	0,577718195	1,090644402
0.2	0,331358181	0,28888763	0,662716363	0,57777526	1,090737894
0.3	0,331237006	0,288916009	0,662474012	0,577832018	1,090830738
0.4	0,331116759	0,288944235	0,662233518	0,57788847	1,090922942
0.5	0,330997428	0,288972308	0,661994856	0,577944616	1,091014511
0.6	0,330879002	0,289000229	0,661758003	0,578000459	1,091105453
0.7	0,330761468	0,289027999	0,661522937	0,578055999	1,091195773
0.8	0,330644817	0,289055618	0,661289634	0,578111237	1,091285477
0,9	0.330529037	0,289083087	0,661058074	0,578166175	1,091374571
1	0.330414117	0,289110407	0,660828234	0,578220813	1,091463062

These two tables show that L_s and W_s decrease with the increase of δ, while L_q and W_q increase along the increase of δ, the expected number of customers served is increasing.

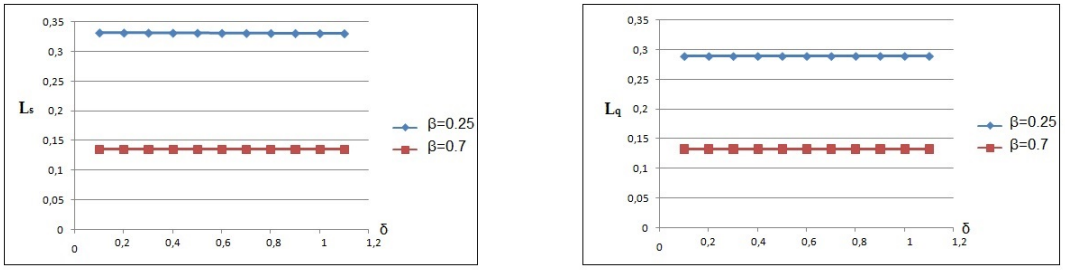

FIGURE 4.5 – δ versus L_s and δ versus L_q

Compared results are shown in figures 4.5, 4.6, 4.7, these later show that when the probability of non-feedback is big, $\beta = 0.7$, the number of customers in the system, in the queue is less than that when

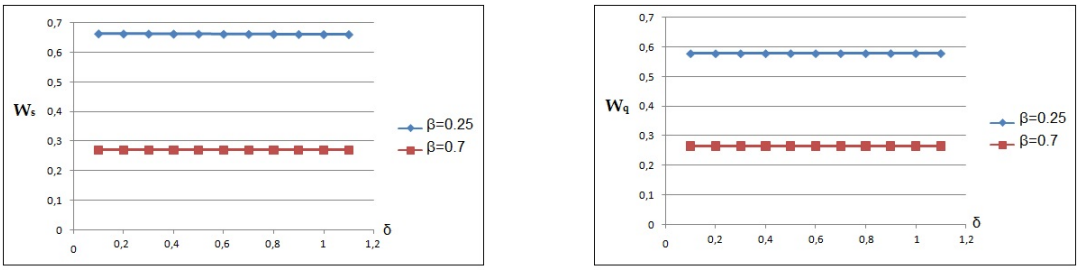

FIGURE 4.6 – δ versus W_s and δ versus W_q

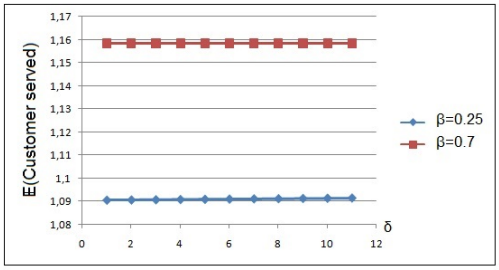

FIGURE 4.7 – δ versus E Number of customers Served

β is small, $\beta = 0.25$, and consequently the expected waiting time in the system, the expected waiting time in the queue, in the first case is less than the second one, the expected number of customers served in the first case is bigger than the second one. These results agree nicely with our intuition.

Finally, we present the effect of of non-feedback probability β, we evaluate different measure performances at its different values while $N = 7$, $\lambda = 2$, $\delta = 0.75$, $\mu_1 = 7$, $\mu_2 = 4$, $\eta = 0.1$, $\phi_1 = 0.3$, $\phi_2 = 0.7$ See the following table :

β	L_s	L_q	W_s	W_q	$\mathbb{E}(CustomerServed)$
0.1	1.181243839	0.334005347	0.59062192	0.167002673	1.881762209
0.2	0.940511222	0.413229209	0.470255611	0.206614604	2.909353183
0.3	0.737499147	0.436879606	0.368749573	0.218439803	3.529375142
0.4	0.604826516	0.426653359	0.302413258	0.21332668	3.871129358
0.5	0.515591814	0.403652243	0.257795907	0.201826121	4.069330069
0.6	0.451791792	0.377567405	0.225895896	0.188783703	4.193837398
0.7	0.403652144	0.352145624	0.201826072	0.176072812	4.278137064
0.8	0.365778891	0.32866691	0.182889446	0.164333455	4.338882923
0.9	0.335022606	0.307435757	0.167511303	0.153717878	4.384887817
1	0.309431653	0.288383991	0.154715827	0.144191995	4.421120692

Now, we evaluate different measure performances at different values of non- feedback probability β, while $N = 7$, $\lambda = 2$, $\delta = 0.75$, $\mu_1 = 14$, $\mu_2 = 10$, $\eta = 0.1$, $\phi_1 = 0.3$, $\phi_2 = 0.7$. The numerical results are

summarized in the following table :

β	L_s	L_q	W_s	W_q	$\mathbb{E}(CustomerServed)$
0.1	0.877125577	0.411818314	0.438562789	0.205909157	2.981668691
0.2	0.547837429	0.403817466	0.273918714	0.201908733	3.811080636
0.3	0.404824653	0.34705907	0.202412327	0.173529535	4.051518476
0.4	0.32522401	0.296927802	0.162612005	0.148463901	4.148853488
0.5	0.273399553	0.257559104	0.136699777	0.128779552	4.199741576
0.6	0.236487835	0.22675818	0.118243918	0.11337909	4.231113209
0.7	0.208663127	0.202268759	0.104331564	0.10113438	4.25262326
0.8	0.186850741	0.182427073	0.093425371	0.091213537	4.268454144
0.9	0.169250844	0.166064974	0.084625422	0.083032487	4.2806916
1	0.154729651	0.152359846	0.077364826	0.076179923	4.29049251

The above tables show the effect of non-feedback probability β for our model with balking, reneging and feedback. When the probability β of non-feedback or returning to the system increases, L_s, L_q, W_s, W_q decrease, while $\mathbb{E}(CustomerServed)$ increases, which agrees perfectly with the intuition.
Compared results of the above two tables are shown in the following figures

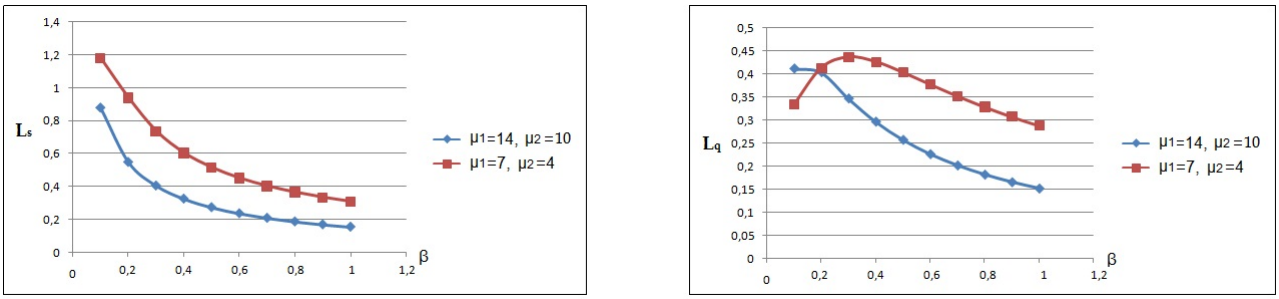

FIGURE 4.8 – β versus L_s and β versus L_q

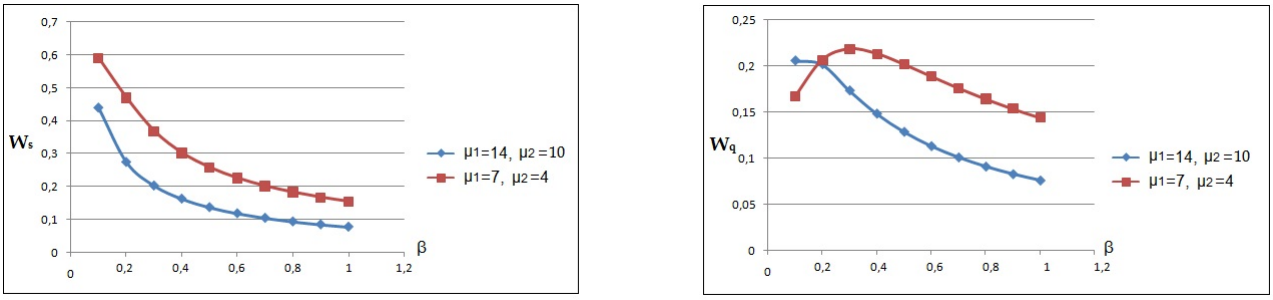

FIGURE 4.9 – β versus W_s and β versus W_q

Figures 4.8, 4.9, 4.10, give a significative result for our system. When the mean service times are small, $\mu_1 = 14$, $\mu_2 = 10$, the expected number of customer in the system, in the queue is less than that when the mean service times is big $\mu_1 = 7$, $\mu_2 = 4$, the expected waiting time in the system, in the queue in the first case is less than in the second one, and finally the expected number of customers served is large

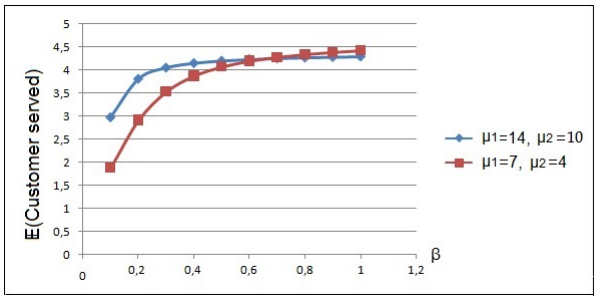

FIGURE 4.10 – β versus E Number of customers Served

in the first case than the second one.

As a conclusion, we conclude that all the figures indicate that the phase-merging algorithm is reasonably effective in approximating L_s, L_q, W_s, W_q, and $E(CustomerServed)$ for all values of λ, μ_1, $\mu_2 = 10$, δ, and β.

4.6 Conclusion

The aim of this paper was to obtain the analytical solution of the $M/M/2/N$ queue (capacity N) with balking, reneging, feedback and two heterogeneous servers. The steady state probabilities and some measures of effectiveness of the system were obtained. Moreover, some numerical values of the analytical results were also obtained

References

[1] Abou El-Ata, M. O. and Hariri, A. M. A. The $M/M/c/N$ queue with balking and reneging. Computers and Operations Research. 19 (1992), No. 13 713-716.

[2] Ancker, Jr., C. J. and Gafarian, A. V. Some queuing problems with balking and reneging : I. Operations Research. 11 (1963), No. 1, 88-100.

[3] Ancker, Jr., C. J. and Gafarian, A. V. Some queuing problems with balking and reneging : II. Operations Research. 11 (1963), No. 6, 928-937.

[4] Ayyapan, G., Muthu Ganapathi Subramanian, A. and Sekar, $GM/M/1$ Retrial Queueing System with Loss and Feedback under Non-pre-emptive Priority Service by Matrix Geometric Method, Applied Mathematical Sciences.4,(2010), No. 48, 2379-2389.

[5] D'Avignon, G.R. and Disney, R.L., Single Server Queue with State Dependent Feed- back,INFOR. 14 (1976), 71-85.

[6] Haight, F. A. Queueing with balking. Biometrika. 44 (1957), 360-369.

[7] Haight, F. A. Queueing with reneging. Metrika. 2 (1959) 186-197.

[8] Kumar, R. and Sharma, S.K. $M/M/1/N$ Queuing System with Retention of Reneged Customers, Pakistan Journal of Statistics and Operation Research. 8(2012), 859-866.

[9] Kumar, R. and Sharma,S.K. An $M/M/1/N$ Queuing Model with Retention of reneged customers and Balking, American Journal of Operational Research. 2(2012), No. 1, 1-5.

[10] Mahdy S. El-Paoumy and Hossam A. Nabwey, The Poissonian Queue with Balking Function,

Reneging and two Heterogeneous Servers. International Journal of Basic & Applied Sciences IJBAS-IJENS. 11 (2011), No. 6, 149-152.

[11] Robert, E. Reneging phenomenon of single channel queues. Mathematics of Operations Re- search. 4 (1979), No. 2, 162-178.

[12] Santhakumaran, A. and Thangaraj, V. A Single Server Queue with Impatient and Feedback Customers,Information and Management Science. 11, (2000), No. 3, 71-79.

[13] Sharma S. K and Kumar. R, A Markovian Feedback Queue with Retention of Reneged Customers and Balking. AMO-Advanced Modeling and Optimization. 14 (2012), No. 3, 681-688.

[14] Takacs, L. A Single Server Queue with Feedback,The Bell System Tech. Journal. 42 (1963), 134-149.

[15] Thangaraj, V. and Vanitha, S. On the Analysis of $M/M/1$ Feedback Queue with Catastrophes using Continued Fractions,International Journal of Pure and Applied Mathematics. 53(2009), No. 1, 131-151.

[16] Wang, K-H. and Chang, Y-C. Cost analysis of a

nite $M/M/R$ queueing system with balking, reneging and server breakdowns. Mathematical Methods of Operations Research 56 (2002), No.2, 169-180.

Chapitre 5

Étude d'un système de files d'attente avec rappels et dérobade

Ce travail a fait l'objet d'un travail soumis dans le journal Applied and Computational Mathematics (Septembre 2014).

Study of retrial queueing model with balking

AMINA ANGELIKA BOUCHENTOUF[1], MOKHTAR KADI[2], ABBES RABHI[3]

[1,3] Department of Mathematics, Djillali Liabes university of Sidi Bel Abbes,
B. P. 89, Sidi Bel Abbes 22000, Algeria
E-mail :bouchentouf_amina@yahoo.fr, rabhi_abbes@yahoo.fr

[2] Department of Mathematics, Dr. Moulay Taher University, Saida 20000, Algeria
E-mail :kadi1969@yahoo.fr

Abstract.

This paper deals a retrial queueing model with balking customers in which an arriving customer that finds the service facility busy will either join the infinite buffer orbit with probability ϑ, or leave the system with probability $1-\vartheta$ The customer in the orbit either attempts service again after a random time or gives up receiving service and leaves the system after a random time. The impatience time of each customer in the orbit is exponentially distributed with mean β^{-1} The impatience times are assumed to be mutually independent. Inter-retrial times are independent and follow a geometric distribution. For this model we derive analytic formula for the joint probability generating function of the remaining service time of the customer currently in the server and the number of customers in the orbit. Furthermore, we obtain the probability generating function of the total number of customers in the system.

Keywords : Queueing models, retrial queue, balking queue, probability generating function. AMS Subject Classification : 60K25, 68M20, 90B22.

5.1 Introduction

Queueing systems constitute a central tool in modelling and performance analysis of e.g. telecommunication systems and computer systems.

Retrial queues are queueing systems in which arriving customers who find all servers occupied may retry for service again after a random amount of time. Retrial queues have been widely used to model many problems in telephone systems, call centers, telecommunication networks, computer networks and computer systems, and in daily life.

A retrial queue is similar to an ordinary queueing system, the fundamental differences are firstly, the entities who enter during a down or busy period of the server or servers may reattempt service at some random time in the future, and secondly a waiting room, which is known as a primary queue in the context of retrial queues, is not mandatory. In place of the ordinary waiting room is a buffer called an orbit to which entities proceed after an unsuccessful attempt at service, and from which they retry service according to a given probabilistic or deterministic policy.

Owing to the utility and interesting mathematical properties of retrial queueing models, a vast literature on the subject has emerged over the past several decades. For a general survey of retrial queues and a summary of many results, the reader is directed to the works Falin [8](1990). For more information about the retrial queueing theory see [12, 9, 7, 17, 21] and references therein.

Two common modes in which customers display their impatience are balking and reneging. A call in customer who cannot be helped immediately by a human server might be told how long a wait or how many people he/she faces before an operator is available. Then the customer might hang up (balk) or decide to hold. This is the balking behavior : a customer refuses to enter the queue if the wait is too

long or the queue is too big. On the other hand, a customer who is waiting for an operator might hang up (renege) before getting served if the wait in line becomes too long. This is the reneging behavior. Of course, there can be a combination of the two. It is acknowledged that customers impatience is significant in practice and modeling call centers.
In general, the economic analysis of customer behavior in a queueing system is based on some reward-cost structure which is imposed on the system and reflects the customer's desire for service and their unwillingness to wait. Customers are allowed to make decisions about their actions in the system, for example they may decide whether to join or balk, to wait or abandon, to retry or not etc. The customers want to maximize their benefit, taking into account that the other customers have the same objective, and so the situation can be considered as a game among them.

Haight [13](1957) was the first who considered an $M/M/1$ queue with balking, then an $M/M/1$ queue with customers reneging was also proposed by Haight [14] (1959). The combined effects of balking and reneging in an $M/M/1/N$ queue have been investigated by Ancker and Gafarian [3, 4] (1963). Abou-EI-Ata and Hariri [2] (1992) considered the multiple servers queueing system $M/M/c/N$ with balking and reneging. Wang and Chang [22] (2002) extended this work to study an $M/M/c/N$ queue with balking, reneging and server breakdowns. Garnett . *al* [10] (2002) gave the study on Designing a call center with impatient customers. Zeltyn and Mandelbaum [24] (2005) studied Call centers with impatient customers ; many server asymptotics of the $M/M/n+G$ queue, where this queue is characterized by Poisson arrivals, exponential service times, n service agents and generally distributed patience times of customers. Ward Whitt [23] (2006) focused on Fluid models for multiserver queues with abandonments.
Kangzhou Wang *.al* [15] (2010) studied a queueing system with impatient customers. (a review paper), queueing systems with impatient customers is surveyed in accordance with various dimensions, the impatient behaviors (balking and reneging) and their various rules proposed in literature were studied. Analytic solutions, numerical solutions and simulation modeling of the queue with impatient customers were investigated. Shuangchi He and Dai [19] (2011) studied Many-server queues with customer abandonment, where numerical analysis of their diffusion models was given. Tkachenko [20] (2013) focused on the multi-channel queueing system with heterogeneous servers, regenerative input flow and balking, where servers times are random variables but not necessary exponential. Kumar and Sharma [16] (2012) studied a single server queueing system with retention of reneged customers.

Studying a discrete time queueing network has a great interest, this interest is motivated by various applications in communication networks, in [26](1999) authors considered a discrete time $PH/Geo/1/1$ retrial queue with geometric retrial times. A discrete time retrial queue with two types of customers was analyzed in [6]. A similar model was studied in [25](1998) where a recursive formula was presented to compute the marginal steady state distribution of the number of customers in the queue and in the orbit. A discrete time $Geo/G/1/1$ retrial queue with general retrial times was analyzed in [5] (2004). In [18](2004) authors considered a discrete time $Geo/G/1/1$ retrial queue with batch arrivals where the number of arrivals during any time slot follows a general distribution and the number of arrivals during different time slots are independent.

Aboul-Hassan *.al* [1] (2005) studied a discrete time $Geo/G/1/1$ retrial queue with balking customers Analytic formula for the joint probability generating function of the remaining service time of the customer currently in the server and the number of customers in the orbit was derived. and the probability generating function of the total number of customers in the system was also obtained.

In this paper, we consider an discrete time $Geo/G/1/1$ retrial queueing model with balking and abandonment from orbit ; The main result in this work consists in deriving the approximate analysis of our special queuing system ; this sort of systems have received a great interest in recent years because they are used in the modeling and analysis of modern communication systems. In this paper we analyze

a queue with balking customers in which an arriving customer that finds the service facility busy will either join the infinite orbit with probability ϑ or leave the system with probability $1-\vartheta$. The customer in the orbit either attempts service again after a random time or gives up receiving service and leaves the system after a random time. The impatience time of each customer in the orbit is exponentially distributed with mean β^{-1}. The impatience times are assumed to be mutually independent. Inter-retrial times are assumed to be independent and geometrically distributed. Early arrival scheme is assumed.
We derived analytic formulas for both the joint probability generating function of the remaining service time of the customer currently in the server and the number of customers in the orbit, and the probability generating function of the total number of customers in the system. These generating functions could be used to obtain important performance measures such as average system size and average waiting time.

5.2 Mathematical model

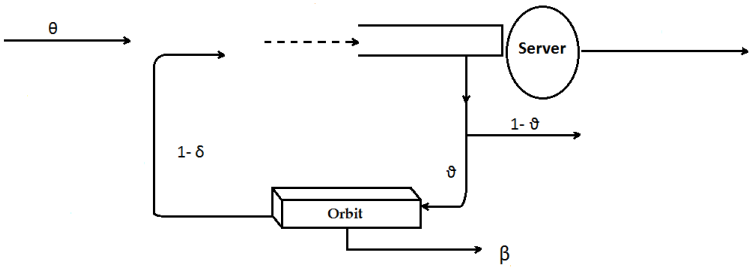

FIGURE 5.1 – Retrial queueing system with balking

Consider a retrial queueing system with a single server, orbit and without a waiting room. We assume that the time axis is divided into intervals of equal length called time slots. All the system events occur at the boundary of these time slots. More specifically, the evolution of the system is controlled by the early arrival scheme [11] (1997). In this scheme, it is assumed that the arrivals (either external or coming from the orbits) during time slot r ($r \geq 0$) occur at the beginning of this time slot. On the other hand, service completion during time slot r occurs by the end of this time slot. During any time slot r, a single external arrival occurs (independently of all the other system events) with probability θ. In other words, it is assumed that the time between consecutive arrivals follows a geometric distribution. If the arriving customer finds the server busy, then he joins one of finite orbits (to retry his demand later) with probability ϑ, and departs completely from the system with probability $1-\vartheta$. Service time (counted in time slots) of any customer is assumed to follow a general distribution γ_j, $j = 1, 2, 3, ...$, where $\gamma_j = \mathbb{P}\{\text{service time} = \text{j time slots }\}$. Customers in the orbit retry (independently of each other) their service during each time slot with probability $1-\delta$, In other words, the time between retrials for each customer follows a geometric distribution with parameter $1-\delta$. If both an external arrival and a retrial occur at the same time and the server is idle, then the external arrival begins his service immediately and the other customer returns to the orbit to retry obtaining his service later. If no external arrival and more than one retrial occur at the same time to an idle server, then one customer is selected at random to start service and the other customers return to the orbits to retry obtaining their service later. The customer in the orbit either attempts service again after a random time or gives up receiving service and leaves the system after a random time. The impatience time of each customer in the orbit is geometric distributed with parameter β. The impatience times are assumed to be mutually independent. Now, let

us define N_r to be the number of customers in the orbit at the beginning of the time slot r ($r \geq 0$). Since a general service time distribution is assumed, we have to keep track of the remaining service time also.

Define X_r to be the number of the remaining time slots for the customer in the server at the beginning of the time slot r ($r \geq 0$). The system state is given by the two dimensional stochastic process $\{(X_r, N_r), r \geq 0\}$. Since both inter-arrival and inter-retrial times follow a geometric distribution, then the stochastic process $\{(X_r, N_r), r \geq 0\}$ is a Markov chain. The next section treats the problem of obtaining the steady state distribution of this process.

5.3 Main result

Our main result consist in obtaining the steady state distribution of $\{(X_r, N_r), r \geq 0\}$. So, let $\{(X_r, N_r), r \geq 0\}$ a two dimensional Markov chain with state space $\{0, 1, 2, ...\} \times \{0, 1, 2, ...\}$. Define $P_k(i)$ as the joint steady state probability of this stochastic process. Precisely $P_k(i) = \mathbb{P}(X_r = i, N_r = k)$. The joint probability generating function of $P_k(i)$ is given in Theorem 1. However, we have to define the following generating functions first :

$$\Gamma(z) = \sum_{j=1}^{\infty} \gamma_j z^j, \;\; H_i(z) = \sum_{k=0}^{\infty} P_k(k) z^k,$$

and

$$P(x, z) = \sum_{i=1}^{\infty} H_i(z) x^i$$

Theorem 5.3.1. *The joint probability generating function of $P_k(i)$ is given by :*

$$\begin{aligned} P(x,z) = \frac{x(\delta z - \beta)}{x(\delta z - \beta) - z(1 - \vartheta\theta(1-z))} \Bigg(& \frac{\theta z(z+\overline{\theta}) - \overline{\theta}(z(1+\beta) - \beta)}{\theta z(\delta z - \beta)} \Gamma(x) \\ & - \frac{\theta(1 - \vartheta\theta(1-z))z + \overline{\theta}((1-\vartheta\theta)\overline{\theta} + (1+z)\vartheta\theta)\Gamma(x)}{\theta(\delta z - \beta)} \\ & \frac{\left(\theta z(z+\overline{\theta}) - \overline{\theta}(z(1+\beta) - \beta)\right) \Gamma\left(\frac{z(1-\vartheta\theta(1-z))}{\delta z - \beta}\right)}{z\left(\theta(1-\vartheta\theta(1-z))z + \overline{\theta}\left((1-\vartheta\theta)\overline{\theta} + (1+z)\vartheta\theta\right)\Gamma\left(\frac{z(1-\vartheta\theta(1-z))}{\delta z - \beta}\right)\right)} \Bigg) H_0(z), \end{aligned} \tag{5.1}$$

and

$$\begin{aligned} H_0(z) = & \frac{1}{1 + \frac{(\delta - \beta)}{(\delta - \beta) - 1} \left(\frac{\theta(2+\overline{\theta}) - 1}{\theta(\delta - \beta)} \Gamma(1) - \frac{\left(\theta + \overline{\theta}\left((1-\vartheta\theta)\overline{\theta} + 2\vartheta\theta\right)\Gamma(1)\right)(\theta(2+\overline{\theta}) - 1)\Gamma\left(\frac{1}{\delta - \beta}\right)}{\theta(\delta - \beta)\left(\theta + \overline{\theta}\left((1-\vartheta\theta)\overline{\theta} + 2\vartheta\theta\right)\Gamma\left(\frac{1}{\delta-\beta}\right)\right)} \right)} \\ & \times \frac{\prod_{k=0}^{\infty} G(\delta^k z)}{\prod_{k=0}^{\infty} G(\delta^k)}, \end{aligned} \tag{5.2}$$

where

$$G(z) = \frac{\left(\theta z(z+\overline{\theta}) - \overline{\theta}(z(1+\beta) - \beta)\right) \Gamma\left(\frac{z(1-\vartheta\theta(1-z))}{\delta z - \beta}\right)}{z\left(\theta(1-\vartheta\theta(1-z))z + \overline{\theta}\left((1-\vartheta\theta)\overline{\theta} + (1+z)\vartheta\theta\right)\Gamma\left(\frac{z(1-\vartheta\theta(1-z))}{\delta z - \beta}\right)\right)}. \tag{5.3}$$

5.4 Proof

To prove our result we have to introduce the balance equations of our system.
The balance equations of the Markov chain $\{(X_r, N_r), r \geq 0\}$ are as follows :

$$(\theta + \beta)P_0(0) = (1 - \vartheta\theta)\bar{\theta}P_0(1) + \beta P_1(0), \tag{5.4}$$

$$(1 - \theta\delta^k + \beta)P_k(0) = (1 - \vartheta\theta)\bar{\theta}P_k(1) + \vartheta\theta P_k(1) + \vartheta\theta P_{k-1}(1) + \beta P_{k+1}(0), \;\; k \geq 1, \tag{5.5}$$

$$P_0(i) = (1 - \vartheta\theta)P_0(i+1) + \lambda\gamma_i P_0(0) + \bar{\theta}(1-\delta)\gamma_i P_1(0) + \beta P_1(i), \;\; i \geq 1, \tag{5.6}$$

$$\delta P_k(i) = (1-\vartheta\theta)P_k(i+1) + \vartheta\theta P_{k-1}(i+1) + \theta\gamma_i P_k(0) + \bar{\theta}(1-\delta^{k+1})\gamma_i P_{k+1}(0) + \beta P_{k+1}(i), \;\; i, k_1 \geq 1. \tag{5.7}$$

With

$$\sum_{k=0}^{\infty}\sum_{k=0}^{\infty} P_k(i) = 1. \tag{5.8}$$

Now, multiplying both sides of equation (5.5) by z^k, making summation from $k = 1$ to ∞, and using the boundary condition in (5.4), we get :

$$H_0(z) = \frac{z\left((1 - \vartheta\theta)\overline{\theta} + \vartheta\theta(1 + z)\right)}{z(1 + \beta) - \beta}H_1(z) + \frac{\theta}{z(1 + \beta) - \beta}zH_0(\delta z). \tag{5.9}$$

Applying a similar procedure to (5.6) and (5.7), we will have :

$$\begin{aligned} H_i(z) = {} & \frac{(1 - \vartheta\theta(1 - z))}{\delta z - \beta}zH_{i+1}(z) + \frac{\gamma_i\left(\overline{\theta} + \theta z\right)}{\delta z - \beta}H_0(z) \\ & - \frac{\overline{\theta}\gamma_i}{\delta z - \beta}H_0(\delta z). \end{aligned} \tag{5.10}$$

Then, substituting for $H_0(\delta z)$ in the above equation from (5.9) gives :

$$\begin{aligned} H_i(z) = {} & \frac{(1 - \vartheta\theta(1 - z))}{\delta z - \beta}zH_{i+1}(z) - \frac{\overline{\theta}\gamma_i\left((1 - \vartheta\theta)\,\overline{\theta} + \vartheta\theta(1 + z)\right)}{\theta(\delta z - \beta)}H_1(z) \\ & + \frac{\gamma_i\left(z\theta(\theta z + \overline{\theta}) - \overline{\theta}(z(1 + \beta) - \beta)\right)}{\theta(\delta z - \beta)z}H_0(z). \end{aligned} \tag{5.11}$$

After that, multiplying both sides by x^i and summing over i we get :

$$\begin{aligned} \frac{x(\delta z - \beta) - z(1 - \vartheta\theta(1 - z))}{x(\delta z - \beta)}P(x, z) = {} & \frac{\theta z(z + \overline{\theta}) - \overline{\theta}(z(1 + \beta) - \beta)}{\theta z(\delta z - \beta)}\Gamma(x)H_0(z) \\ & - \frac{\theta(1 - \vartheta\theta(1 - z))z + \overline{\theta}\left((1 - \vartheta\theta)\overline{\theta} + (1 + z)\vartheta\theta\right)\Gamma(x)}{\theta(\delta z - \beta)}H_1(z). \end{aligned} \tag{5.12}$$

Now, to get a relation between $H_0(z)$ and $H_1(z)$, it is sufficient to set $x = \dfrac{z - \vartheta\theta(1 - z)}{(\delta z - \beta)}$ in equation (5.12), so,

$$H_1(z) = \frac{\left(\theta z(z + \overline{\theta}) - \overline{\theta}(z(1 + \beta) - \beta)\right)\Gamma\left(\frac{z(1-\vartheta\theta(1-z))}{\delta z-\beta}\right)}{z\left(\theta(1 - \vartheta\theta(1 - z))z + \overline{\theta}\left((1 - \vartheta\theta)\overline{\theta} + (1 + z)\vartheta\theta\right)\Gamma\left(\frac{z(1-\vartheta\theta(1-z))}{\delta z-\beta}\right)\right)}H_0(z). \tag{5.13}$$

Later, substituting for $H_1(z)$ in (5.12) using (5.13), we get :

$$\frac{x(\delta z-\beta)-z(1-\vartheta\theta(1-z))}{x(\delta z-\beta)}P(x,z)=\Bigg(\frac{\theta z(z+\overline{\theta})-\overline{\theta}(z(1+\beta)-\beta)}{\theta z(\delta z-\beta)}\Gamma(x) - \frac{\theta(1-\vartheta\theta(1-z))z+\overline{\theta}((1-\vartheta\theta)\overline{\theta}+(1+z)\vartheta\theta)\Gamma(x)}{\theta(\delta z-\beta)} - \frac{\left(\theta z(z+\overline{\theta})-\overline{\theta}(z(1+\beta)-\beta)\right)\Gamma\left(\frac{z(1-\vartheta\theta(1-z))}{\delta z-\beta}\right)}{z\left(\theta(1-\vartheta\theta(1-z))z+\overline{\theta}\left((1-\vartheta\theta)\overline{\theta}+(1+z)\vartheta\theta\right)\Gamma\left(\frac{z(1-\vartheta\theta(1-z))}{\delta z-\beta}\right)\right)}\Bigg)H_0(z). \tag{5.14}$$

Finally, after rearranging the terms of the last equation, we obtain the first equation of the theorem.

Now, in order to find $H_0(z)$, we substitute for $H_1(z)$ in (5.9) using (5.13). Thus

$$\begin{aligned}H_0(z)&=\left(\frac{\frac{\theta z}{z(1+\beta)-\beta}}{1-\frac{z\left((1-\vartheta\theta)\overline{\theta}+\vartheta\theta(1+z)\right)\left(\theta z(z+\overline{\theta})-\overline{\theta}(z(1+\beta)-\beta)\right)\Gamma\left(\frac{z(1-\vartheta\theta(1-z))}{\delta z-\beta}\right)}{(z(1+\beta)-\beta)z\left(\theta(1-\vartheta\theta(1-z))z+\overline{\theta}\left((1-\vartheta\theta)\overline{\theta}+(1+z)\vartheta\theta\right)\Gamma\left(\frac{z(1-\vartheta\theta(1-z))}{\delta z-\beta}\right)\right)}}\right)H_0(\delta z)\\&=G(z)H_0(z).\end{aligned} \tag{5.15}$$

Where $G(z)$ is as defined above. Applying (5.15) recursively we get :

$$H_0(z)=H_0(0)\prod_{k=0}^{\infty}G(\delta^k z) \tag{5.16}$$

Putting $z=1$ in the above equation leads to :

$$H_0(0)=\frac{H_0(1)}{\prod_{k=0}^{\infty}G(\delta^k)}. \tag{5.17}$$

In order to obtain $H_0(1)$ we set $x=1$ and take the limit as z tends to 1 in (5.1) :

$$\begin{aligned}P(1,1)=\frac{(\delta-\beta)}{(\delta-\beta)-1}\Bigg(&\frac{\theta(2+\overline{\theta})-1}{\theta(\delta-\beta)}\Gamma(1)-\frac{\theta+\overline{\theta}\left((1-\vartheta\theta)\overline{\theta}+2\vartheta\theta\right)\Gamma(1)}{\theta(\delta-\beta)}\\&\times\frac{(\theta(2+\overline{\theta})-1)\Gamma\left(\frac{1}{\delta-\beta}\right)}{\theta+\overline{\theta}\left((1-\vartheta\theta)\overline{\theta}+2\vartheta\theta\right)\Gamma\left(\frac{1}{\delta-\beta}\right)}\Bigg)H_0(1).\end{aligned} \tag{5.18}$$

Using the normalizing equation (5.8), it is clear that $H_0(1)+P(1,1)=1$. Hence, $H_0(1)$ has the following form :

$$H_0(1)=\frac{1}{1+\frac{(\delta-\beta)}{(\delta-\beta)-1}\left(\frac{\theta(2+\overline{\theta})-1}{\theta(\delta-\beta)}\Gamma(1)-\frac{\left(\theta+\overline{\theta}\left((1-\vartheta\theta)\overline{\theta}+2\vartheta\theta\right)\Gamma(1)\right)(\theta(2+\overline{\theta})-1)\Gamma\left(\frac{1}{\delta-\beta}\right)}{\theta(\delta-\beta)\left(\theta+\overline{\theta}\left((1-\vartheta\theta)\overline{\theta}+2\vartheta\theta\right)\Gamma\left(\frac{1}{\delta-\beta}\right)\right)}\right)}. \tag{5.19}$$

Substituting (5.19) into (5.17) then substituting with the result into (5.16)we get $H_0(z)$.
Now, let us define $Q(z)$ to be the probability generating function of the number of customers in the system including those in the orbits and the one in the server. That is ;

$$Q(z)=\sum_{k=1}^{\infty}q_k z^k,$$

where q_k is the probability that there are k customers in the system. Hence,

$$Q(z) = H_0(z) + zP(1,z)$$

Using the first equation of theorem 1 we get the following result

Theorem 5.4.1. *The probability generating function of the steady state distribution of the total number of customers in the system is given by :*

$$\begin{aligned} Q(z) = \Bigg(1 + z\Bigg(\frac{\delta z-\beta}{\delta z-\beta-z(1-\vartheta\theta(1-z))}\Bigg(\frac{\theta z(z+\overline{\theta})-\overline{\theta}(z(1+\beta)-\beta)}{\theta z(\delta z-\beta)}\Gamma(1) \\ -\frac{\theta(1-\vartheta\theta(1-z))z+\overline{\theta}((1-\vartheta\theta)\overline{\theta}+(1+z)\vartheta\theta)\Gamma(1)}{\theta(\delta z-\beta)} \\ \frac{\left(\theta z(z+\overline{\theta})-\overline{\theta}(z(1+\beta)-\beta)\right)\Gamma\left(\frac{z(1-\vartheta\theta(1-z))}{\delta z-\beta}\right)}{z\left(\theta(1-\vartheta\theta(1-z))z+\overline{\theta}\left((1-\vartheta\theta)\overline{\theta}+(1+z)\vartheta\theta\right)\Gamma\left(\frac{z(1-\vartheta\theta(1-z))}{\delta z-\beta}\right)\right)}\Bigg)\Bigg)\Bigg)H_0(z) \end{aligned} \tag{5.20}$$

5.5 Conclusion

Motivated by the application to telephone call centers and more general customer contact centers, this work focused on the queueing systems with impatient customers. Recently there has been great interest in queueing networks with impatient customer. An important aspect for modeling of service-oriented call centers is the impatience behavior of the customers. In this paper, we studied an discrete time retrial queueing model with balking and abandonment from orbit ; The main result consisted in deriving the approximate analysis of our special queuing system. The probability generating function of the total number of customers in our system was also obtained.

Références

[1] Abdel-Karim Aboul-Hassan, Sherif I. Rabia and Ahmed Kadry, Analytical study of a discrete time retrial queue with balking customers and early arrival scheme, Alexandria Engineering Journal, 44 (2005), No. 6, 911-917

[2] Abou El-Ata, M. O. and Hariri, A. M. A. The $M/M/c/N$ queue with balking and reneging. Computers and Operations Research, 19 (1992), No. 13 713-716.

[3] Ancker, Jr., C. J. and Gafarian, A. V. Some queuing problems with balking and reneging :I. Operations Research, 11 (1963), No. 1, 88-100.

[4] Ancker, Jr., C. J. and Gafarian, A. V. Some queuing problems with balking and reneging : II. Operations Research, 11 (1963), No. 6, 928-937.

[5] I. Atencia and P. Moreno, $Geo/G/1$ Retrial Queue with General Retrial Times, Queueing Systems, Vol. 48, pp. 5-21 (2004).

[6] B.D. Choi and J.W. Kim, Discrete-Time Geo1 ; $Geo2/G/1$ Retrial Queueing Systems with Two Types of Calls, Computers and Mathematics with Applications, Vol. 33, pp. 79 88 (1997).

[7] Ebrahimi, N. (2006). System reliability based on system wear. Stochastic Models, 22(1), 21-36.

[8] Falin, G. (1990). A survey of retrial queues. Queueing Systems : Theory and Applications, 7(2), 127-167.

[9] Falin, G. I. and Artalejo, J. R. (1998). An infinite source retrial queue. European Journal of Operational Research, 108(2), 409.

[10] O. Garnett, A. Mandelbaum, and M. Reiman. Designing a call center with impatient customers. Manufac- turing & Service Operations Management, 4(3) :208-227, June 2002.

[11] E. Gelenbe and G. Pujolle, Introduction to Queueing Networks. John Wiley, New York, second ed. (1997). 8 APPL. COMPUT. MATH.,

[12] Gharbi, N. and Ioualalen, M. (2006). GSPN analysis of retrial systems with server breakdowns and repairs. Applied Mathematics and Computation, 174(2), 1151-1168.

[13] Haight, F. A. Queueing with balking. Biometrika. 44 (1957), 360-369.

[14] Haight, F. A. Queueing with reneging. Metrika. 2 (1959) 186-197.

[15] Kangzhou Wang, Na Li, and Zhibin Jiang. Queueing system with impatient customers : A review. In 2010 IEEE International Conference on Service Operations and Logistics and Informatics (SOLI), pages 82-87, 2010.

[16] Kumar, R. and Sharma, S.K. $M/M/1/N$ Queuing System with Retention of Reneged Customers, Pakistan Journal of Statistics and Operation Research, 8(2012), 859-866.

[17] Libman, L. and Orda, A. (2002). Optimal retrial and timeout strategies for accessing network resources. IEEE/ACM Transactions on Networking, 10(4), 551-564.

[18] R. Nobel, A Discrete-Time Retrial Queueing Model with One Server, (China), INFORMS/APS Conference, June (2004).

[19] Shuangchi He and J. G. Dai. Many-server queues with customer abandonment : numerical analysis of their diffusion models. arXiv preprint arXiv :1104.0347, 2011.

[20] Tkachenko Andrey. Multichannel queueing systems with balking and renenerative input ow, Basic research program, serie : Sceince, Technologie and innovation, WP BRP 14/STI/2013.

[21] Walrand, J. (1991). Communication Networks : A First Course. The Aksen Associates Series in Electrical and Computer Engineering. Richard D. Irwin, Inc., and Aksen Associates, Inc., Homewood, IL and Boston, MA.

[22] Wang, K-H. and 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 (2002), No.2, 169-180.

[23] Whitt. Fluid models for multiserver queues with abandonments. Operations Research, 54(1) :37-54, January 2006.

[24] Sergey Zeltyn and Avishai Mandelbaum. Call centers with impatient customers : Many-server asymptotics of the $M/M/n + G$ queue. Queueing Systems, 51(3-4) :361-402, December 2005.

[25] T. Yang and H. Li, $Geo/G/1$ Discrete Time Retrial Queue with Bernoulli Schedule, Eur. J. Oper. Res., Vol. 111 (3), pp. 629 649, (1998).

[26] T. Yang and H. Li, Steady-State Queue Size Distribution of Discrete-Time PH/Geo/1 Retrial Queues, Mathematical and Computer Modelling, Vol. 30, pp. 51 63 (1999).

Conclusion et perspectives

Dans cette thèse nous nous sommes intéréssés aux systèmes de files d'attente avec dérobade, abandon, rappel et feedback.

◇ ***Le premier chapitre*** est consacré à l'étude des chaînes de Markov qui ont un large champ d'application dans l'analyse dans différents systèmes dynamiques (réseaux de files d'attente, systèmes de communication et informatiques, biologiques, économiques...).

◇ ***Les chapitres deux et trois*** sont une introduction pour les systèmes de files d'attente avec dérobade, abandon, rappel et feedbck.

◇ Dans ***le quatrième chapitre*** nous avons réalisé l'analyse d'un modèle de files d'attente de deux serveurs hétérogènes $M/M/2/N$ (capacité du système N finie) avec dérobade, abandon et feedback. En utilisant le processus de Markov, les équations des probabilités d'état stable sont developpées, diverses mesures de performance du système ont été présentées. De plus, différents exemples numériques ont été réalisés afin de montrer comment les différents paramètres du modèle influencent sur le comportement du système.

◇ Dans ***le cinqième chapitre***, nous avons traité un système de file d'attente $Geo/G/1/1$ avec rappel et dérobade. Les client dans l'orbite tentent d'avoir un service ou abandonnent et quittent le système Après une certaine période aléatoire. Pour ce système la formule analytique de la fonction génératrice de la probabilité conjointe du nombre de clients dans l'orbite et du temps de service restant pour un client entrain d'être servi est obtenue. De plus la fonction génératrice de la probabilité du nombre total de clients dans le système est présentée. Notons que ces fonctions génératrices pourront être utilisés pour obtenir des mesures de performance importantes telles que la taille moyenne du système et le temps d'attente moyen.

Et comme perspectives,

▷ Nos résultats pourront être étendues à un système de file d'attente $Geo/Geo/s/s$ avec rappel, dérobade est deux orbites.

⋆ L'analyse cost-proffit du modèle peut être réalisée afin d'établir son étude économique.

⋆ Ce modèle peut être étudié sous les hypothèses de dépendance de taux d'arrivées et de taux de service qui emmène le système à un environnement plus réaliste.

Bibliographie

[1] Abou El-Ata, M. O. and Hariri, A. M. A. The $M/M/c/N$ queue with balking and reneging. Computers and Operations Research. 19, No. 13 713-716. (1992).

[2] Abdel-Karim Aboul-Hassan, Sherif I. Rabia and Ahmed Kadry, Analytical study of a discrete time retrial queue with balking customers and early arrival scheme, Alexandria Engineering Journal, 44, No. 6, 911-917. (2005).

[3] Allen, A. O. Probability, Statistics, and Queueing Theory with Computer Science Applications. Second edition, Academic Press, New York (First edition : 1978). (1990).

[4] Andersson, W.J.,Continuous Time Markov Chains :An Applications Orieted Approach,Springer New York, (1991).

[5] Andrejev, V. D., and Kudryavtsev, B. M. Markov Queueing Systems. Moscow Institute of Control (in Russian). (1982).

[6] Anisimov, V. V., Zakusilo, O. K., and Donchenko, V. S. Elements of Queueing Theory and Asymptotic System Analysis. Vishcha Shkola, Kiev (in Russian). (1987).

[7] Ancker, Jr., C. J. and Gafarian, A. V. Some queuing problems with balking and reneging : I. Operations Research. 11, No. 1, 88-100. (1963).

[8] Ancker, Jr., C. J. and Gafarian, A. V. Some queuing problems with balking and reneging : II. Operations Research. 11, No. 6, 928-937. (1963).

[9] J.R. Artalejo and A. Gomez-Corral. Channel idle periods in computer and telecommunication systems with customer retrials. Telecommunication Systems, 24, 29-46, (2003).

[10] J. R. Artalejo. Retrial queues with a finite number of sources. Journal of the Korean Mathematical Society, 35(3) :503-525, (1998).

[11] J.R. Artalejo and A. Gomez-Corral. Retrial Queueing Systems. A Computational Approach. Springer, (2008).

[12] J. R. Artalejo. Accessible bibliography on retrial queues. Mathematical and computer modelling 30. 1-6, (1999).

[13] J. R. Artalejo. Accessible bibliography on retrial queues : Progress in 2000- 2009. Mathematical and computer modelling 51, 1071-1081, (2010).

[14] I. Atencia and P. Moreno, $Geo/G/1$ Retrial Queue with General Retrial Times, Queueing Systems, Vol. 48, pp. 5-21 (2004).

[15] D'Avignon, G.R. and Disney, R.L., Single Server Queue with State Dependent Feedback,INFOR. 71-85. (1976).

[16] Ayyapan, G., Muthu Ganapathi Subramanian, A. and Sekar, G. $M/M/1$ Retrial Queueing System with Loss and Feedback under Non-pre-emptive Priority Service by Matrix Geometric Method, Applied Mathematical Sciences.4, No. 48, 2379-2389. (2010).

[17] Baccelli, F., and Bremaud, P. Mathematical Theory of Queues. Springer-Verlag, Berlin. (1990).

[18] Bagchi, T. P., and Templeton, J. P. C. Numerical Methods in Markov Chains and Bulk Queues. Lecture Notes in conomics and Mathematical Systems 72, Springer-Verlag, New York. (1972).

[19] Bharucha-Reid, A.T. Elements of the Theory of Markov Processes and their Applications.New York : McGraw-Hill, (1960).

[20] Bhat, U. N. A Study of the Queueing Systems $M/G/l$ and $GI/M/I$. Lecture Notes in Operations Research and athematical Economics 2, Springer - Verlag, New York. (1968).

[21] Billingsley, P. Statistical Inference for Markov Processes. Chicago : University of Chicago Press, (1961).

[22] S. Bocquet. Queueing Theory with Reneging. Defence Science and Thechnology Organisation. (2005).

[23] Borovkov, A. A. 1976. Stochastic Processes in Queueing Theory. Springer-Verlag,Berlin (Russian original : Nauka, Moscow). (1972).

[24] A. A. Bouchentouf, M. Kadi, A. Rabhi 2013 Analysis of Two Heterogeneous Server Queeing Model With Balking, Reneging and Feedback. Mathematical Sciences And Applications E-Notes Volume 2 No. 2 pp. 10-21 (2013).

[25] Bunday, B. D. Basic Queueing Theory. Edward Arnold, London. (1986).

[26] Chaudhry, M. L., and Templeton, J. G. C. A First Course in Bulk Queues.John Wiley and Sons, New York. (1983).

[27] Cheprasov, V. P. Elements of Queueing Theory. Kazan Aviation Institute (in Russian). (1985).

[28] B.D. Choi and J.W. Kim, Discrete-Time Geo1 ; $Geo2/G/1$ Retrial Queueing Systems with Two Types of Calls, Computers and Mathematics with Applications, Vol. 33, pp. 79 88 (1997).

[29] Chung,K.L.,Markov Chains with Stationary.

[30] Cisco : Behind the hype. Business Week. (2002).

[31] Claudie Chabriac. Processus stochastiques et modélisation. (2012-2013).

[32] J. W. Cohen. Basic Problems of Telephone Traffic Theory and Influance of Repeated Calls. Philips Telecom. Review, 18(2) :49-100, (1957).

[33] J. W. Cohen. Basic Problems of Telephone Traffic Theory and Influance of Repeated Calls. Philips Telecom. Review, 18(2) :49-100, (1957).

[34] Cooper, R. B. Introduction to Queueing Theory. Second edition, North-Holland Publishing Company, New York (First edition : Macmillan, New York, 1972). Republished by the Continuing Engineering Education Program, The George Washington University, Washington D.C., 1990.

[35] Denyiseva, O. M. Queueing Systems with Limited Waiting Times. Radio Svyaz, Moscow (in Russian). (1986).

[36] Ebrahimi, N. System reliability based on system wear. Stochastic Models, 22(1), 21-36. (2006).

[37] A. Eldin. Approach of Theorical Description of Repeated Call Attempts.Erricson Thecnics, 23(3) : 346-407, (1967).

[38] Erlang, A. K. Probability and Telephone Calls. Nyt Tidsskrift Matematik Series B, 20,33 - 39. (1909).

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

[40] Falin, G. A survey of retrial queues. Queueing Systems : Theory and Applications, 7(2), 127-167. (1990).

[41] Falin, G. I. and Artalejo, J. R. An infinite source retrial queue. European Journal of Operational Research, 108(2), 409. (1998).

[42] Fedosejev, Yu. N. Methods of Queueing Systems Analysis. Moscow Institute of Engineering and Physics (in Russian). (1982).

[43] Frédéric Sur. Programmation dynamique, chaînes de Markov, files d'attente. école des Mines de Nancy (2013)-(2014).

[44] O. Garnett, A. Mandelbaum, and M. Reiman. Designing a call center with impatient customers. Manufacturing & Service Operations Management, 4(3) :208-227, (June 2002).

[45] Gharbi, N. and Ioualalen, M. GSPN analysis of retrial systems with server breakdowns and repairs. (2006).

[46] Gelenbe, E.,and Pujolle, G. Introduction to Queuing Networks. John Wiley and Sons, Chichester, England (French Original : Introduction aux réseaux de files d'attente, Edition Eyrolles ; Paris). (1987).

[47] A. Gomez-Corral and M.F. Ramalhoto. On the waiting time distribution and the busy period of a retrial queue with constant retrial rate. Stochastic Modelling and Applications, 3, 37-47, (2000).

[48] Gross,D.and C.M,Harris,Fundamentals of Queueing Theory ,Willcy,New york, (1975 1984).

[49] O. Hashida and K. Kawashima. Buffer Behavior with Repeated Calls. Electronics and Communication in Japan, 62-B, 27, (1979).

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

[51] Haight, F. A. Queueing with reneging. Metrika 186-197. 2 (1959).

[52] Dr. J*á*nos Sztrik. Basic Queuing theory. p 25.

[53] Jean Louis Poss, Probabilité et statistique version 2.1. p74. Mai (2003).

[54] J. Lubacz and J. Roberts. A new approach to the single-server repeated attempt system with balking. Proceedings of the Third Int. Seminar on Teletraffic Theory. Moscow, 290-293, (1984).

[55] Kangzhou Wang, Na Li, and Zhibin Jiang. Queueing system with impatient customers : A review. In 2010 IEEE International Conference on Service Operations and Logistics and Informatics (SOLI), pages 82-87. (2010).

[56] Kashyap, B. R. K., and Chaudhry, M. L. An Introduction to Queueing Theory. Aarkay, Calcutta, India. (1988).

[57] Keilson, J. Markov Chain Models - Rarity and Exponentiality. Springer-Verlag, New York. (1979).

[58] Kendall, D. G. Stochastic processes occurring in the theory of queues and their analysis by the method of the imbedded Markov chain. Annals of Mathematical Statistics, 24(3),338 -354. (1953).

[59] Khintchine, A. Y Mathematical Methods in the Theory of Queueing. Second edition, Hafner Publishing Company, New York (First edition : Griffin,London, 1960 ; Russian original : 1955). (1969).

[60] V. G. Kulkarni and H. M. Liang. Retrial Queues Revisited. Frontiers in Queueing (J.H. Dshalov, ed.) CRC Press Boca Raton, pp 19-34, (1997).

[61] Kumar, R. and Sharma, S.K. $M/M/1/N$ Queuing System with Retention of Reneged Customers, Pakistan Journal of Statistics and Operation Research. 859-866. 8(2012).

[62] Kumar, R. and Sharma, S.K. An $M/M/1/N$ Queuing Model with Retention of reneged cus- tomers and Balking, American Journal of Operational Research. No. 1, 1-5. 2(2012).

[63] Lak*átós*. L*ászló* Szeidl. Miklós Telek. Introduction to Queueing Systems with Telecommunication Applications. Springer, p 88. (2010).

[64] LeGall, P. Les systèmes avec ou sans attente et les processus stochastiques, Tome I. Dunod, Paris. (1962).

[65] Libman, L. and Orda, A. Optimal retrial and timeout strategies for accessing network resources. IEEE/ACM Transactions on Networking, 10(4), 551-564. (2002).

[66] Liqiang Liu. Service Systems with Balking Based on Queueing Time. Chapel Hill Under the direction of Dr. Vidyadhar G. Kulkarni. (2007)

[67] J. Lubacz and J. Roberts. A new approach to the single-server repeated attempt system with balking. Proceedings of the Third Int. Seminar on Teletraffic Theory. Moscow, 290-293, (1984).

[68] Mahdy S. El-Paoumy and Hossam A. Nabwey, The Poissonian Queue with Balking Function, Reneging and two Heterogeneous Servers. International Journal of Basic & Applied Sciences IJBAS-IJENS. No. 6, 149-152. 11 (2011).

[69] Marrakesh International Conference on Probability and Statistics 3^{rd} Edition April 25-28, Cadi Ayyad University Marrakesh, Morocco. (2016).

[70] Nawel Arrar. Problèmes de convergence, optimisation d'algorithmes et analyse stochastique de systèmes de files d'attente avec rappels. Université Badji Mokhtar-Annaba, p 26 (2012).

[71] Neuts, M. F. Structured Stochastic Matrices of $M/G/I$ Type and Their Applications. Marcel Dekker, Inc., New York. (1989).

[72] Newell, G. F. Applications of Queueing Theory. Second edition, Chapman and Hall, London (First edition : 1971). (1982).

[73] R. Nobel, A Discrete-Time Retrial Queueing Model with One Server, (China), INFORMS/APS Conference, June (2004).

[74] Norris, J.R. Markov Chains. Cambridge : Cambridge University Press, (1997).

[75] O. Hashida and K. Kawashima. Buffer Behavior with Repeated Calls. Electronics and Communication in Japan, 62-B, 27, (1979).

[76] Robert, E. Reneging phenomenon of single channel queues. Mathematics of Operations Research 4(2) 162-178. (1979).

[77] Robert, E. Reneging phenomenon of single channel queues. Mathematics of Operations Research. No. 2, 162-178. 4 (1979).

[78] A. Rodrigo. Estimators of the retrial rate in $M/G/1$ retrial queues. Asia-Pacific Journal of Operational Research, 23, 193-213, (2006).

[79] Saaty, T. L. Elements of queueing theory with applications. New York : McGraw Hill. (1961).

[80] Thomopoulos, N. T. Strategic inventory management and planning. Carol Stream :Hitchcock Publishing Co. (1990).

[81] Samuel Karlin. Howard, M. Taylor. A first course in stochastic processes, second edition. Academic press New york San fransisco. (1975).

[82] Santhakumaran, A. and Thangaraj, V. A Single Server Queue with Impatient and Feedback Customers,Information and Management Science. 11, No. 3, 71-79. (2000).

[83] Sergey Zeltyn and Avishai Mandelbaum. Call centers with impatient customers : Many-server asymptotics of the $M/M/n+G$ queue. Queueing Systems, 51(3-4) :361-402, December (2005).

[84] Shuangchi He and J. G. Dai. Many-server queues with customer abandonment : numerical analysis of their diffusion models. arXiv preprint arXiv :1104.0347, (2011).

[85] Takacs, L. A Single Server Queue with Feedback,The Bell System Tech. Journal. 42, 134-149. (1963).

[86] J. G. C. Templeton, Retrial Queues, Top, 7 : 351-353, (1999).

[87] J. G. C. Templeton, Retrial Queues, Queueing Systems, 7, No. 2, 125-227,(1990).

[88] Thangaraj, V. and Vanitha, S. On the Analysis of $M/M/1$ Feedback Queue with Catastrophes using Continued Fractions,International Journal of Pure and Applied Mathematics. 53, No. 1, 131-151. (2009).

[89] Tkachenko Andrey. Multichannel queueing systems with balking and renenerative input ow, Basic research program, serie : Sceince, Technologie and innovation, WP BRP 14/STI/(2013).

[90] P. Tran-Gia and M. Mandjes. Modelling of customer retrial phenomenon in cellular mobile networks. IEEE Journal on Selected Areas in Communication, 15 :1406-1414, (1997) .

[91] van Doom, E. Stochastic Monotonicity and Queueing Applications of Birth-Death Processes. Lecture Notes in Statistics 4, Springer-Verlag, New York, (1981).

[92] Walrand, J. Communication Networks : A First Course. The Aksen Associates Series in Electrical and Computer Engineering. Richard D. Irwin, Inc., and Aksen Associates, Inc., Homewood, IL and Boston, MA, (1991).

[93] Wang, K-H. and 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).

[94] Whitt. Fluid models for multiserver queues with abandonments. Operations Research, 54(1) :37-54, (January 2006).

[95] T. Yang and H. Li, $Geo/G/1$ Discrete Time Retrial Queue with Bernoulli Schedule, Eur. J. Oper. Res., Vol. 111 (3), pp. 629 649, (1998).

[96] T. Yang and H. Li, Steady-State Queue Size Distribution of Discrete-Time $PH/Geo/1$ Retrial Queues, Mathematical and Computer Modelling, Vol. 30, pp. 51 63 (1999).

[97] Y. Velenik, Probabilités et Statistique, université de Genève, pp 195-202 Version du 24 mai (2012).

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

[99] Zakhar Kabluchko, Stochastic Processes (Stochastik II), University of Ulm Institute of Stochastics, p 23 (2013-2014).

ملخص

في هذه الأطروحة، نعتبر نظام M/M/N/2 حيث الزبون المستعجل الذي يجد امامه طابور انتظار يقرر عدم الدخول في النظام او بعد بقائه فترة معينة في الطابور يخرج دون الحصول على الخدمة، واثنين من الخوادم غير المتجانسة. باستخدام عملية ماركوف، استخرجنا معادلات الاحتمالات في حالة نظام ثابت، بعد ذلك اعطينا قوانين حساب بعض مقاييس اداء هذا النظام. وفي الأخير تطرقنا الى بعض الأمثلة العددية لإظهار مدى تأثير هذه المقاييس على اداء النظام.

في الجزء الثاني من الأطروحة قمنا بمعالجة نموذج طابور انتظار جديد مع الاهتمام بالعملاء الذين يصلون إلى مركز مشغول سوف يوجهون إلى مدار انتظار باحتمال ϑ ، أو يتركوا النظام باحتمال 1-ϑ . العميل في المدار يكرر طلبه الخدمة بعد فترة زمنية عشوائية ، أو يتخلى عن الخدمة ويترك النظام بعد وقت عشوائي. نعتبر و فترة نفاد صبر العميل في المدار هو أسي وكل هذه الفترات مستقلة عن بعضها البعض. الوقت بين محاولتين مختلفتين للحصول على الخدمة لا يزال مستقل ويتبع التوزيع الهندسي. لهذا النموذج، تحصلنا على صيغة تحليلية للدالة المولدة لوقت الخدمة المتبقي للعميل المتواجد حاليا عند الخادم وعدد من العملاء في المدار. و بهذا نستطيع الحصول على العدد الإجمالي للعملاء في النظام.

Résumé

Dans cette thèse, nous considérons une file d'attente M/ M /2/N avec dérobade, abandon et deux serveurs hétérogènes. En utilisant le processus de Markov, nous développons d'abord les équations des probabilités d'état stable. Ensuite, nous donnons quelques mesures de performance du système. Enfin, nous présentons quelques exemples numériques pour démontrer comment les différents paramètres du modèle influence sur le comportement du système.

En suite, on traite un nouveau modèle de file d'attente avec dérobade. Un client arrivant à un centre de service et le trouve occupé sera soit Orienté vers une orbite avec une probabilité ϑ, ou quitte le système avec la probabilité 1 - ϑ. Le client en orbite après une durée de temps aléatoire revient au système pour répéter sa demande de service, ou abandonne le service et quitte le système après un temps aléatoire. Le temps d'impatience de chaque client dans l'orbite est exponentielle et ces temps sont supposés être indépendants les uns des autres. Les temps entre les rappels sont encore indépendants et suivent une distribution géométrique. Pour ce modèle, nous obtenons la formule analytique pour la fonction génératrice conjointe du temps de service restant du client actuellement dans le serveur et le nombre de clients dans l'orbite. En outre, on obtient la fonction génératrice du nombre total de clients dans système.

Abstract

In this thesis, we consider an M/ M /2/N (capacity N) queueing system with balking, reneging, feedback and two heterogeneous servers. By using the Markov process method, we first develop the equations of the steady state probabilities. Then, we give some performance measures of the system. Finally, we present some numerical examples to demonstrate how the various parameters of the model in influence the behavior of the system.

After, we stady a retrial queueing model with balking customers in which an arriving customer that finds the service facility busy will either join the infinite buffer orbit with probability ϑ, or leave the system with probability 1 - ϑ. The customer in the orbit either attempts service again after a random time or gives up receiving service and leaves the system after a random time. The impatience time of each customer in the orbit is exponentially distributed and are assumed to be mutually independent. Inter-retrial times are independent and follow a geometric distribution. For this model we derive analytic formula for the joint probability generating function of the remaining service time of the customer currently in the server and the number of customers in the orbit. Furthermore, we obtain the probability generating function of the total number of customers in the system.