



N° Attribué par la bibliothèque

--	--	--	--	--	--	--	--	--	--



Année univ.: 2017/2018

# Étude d'un système de files d'attente avec serveurs hétérogènes, dérobade et abandon des clients.

Mémoire présenté en vue de l'obtention du diplôme de

Master Académique

Université de Saida - Dr Moulay Tahar

Discipline : MATHÉMATIQUES

Spécialité : Analyse Stochastique, Statistique des Processus et  
Applications (ASSPA)

par

**Melle. Khadidja HAFSI<sup>1</sup>**

Sous la direction de

**Dr. Mokhtar KADI**

Soutenu le 24/06/2018 devant le jury composé de

<b>Dr. L. BOUSMAHA</b>	Université Dr Tahar Moulay - Saïda	Présidente
<b>Dr. M. KADI</b>	Université Dr Tahar Moulay - Saïda	Encadreur
<b>Pr. A. KANDOUCI</b>	Université Dr Tahar Moulay - Saïda	Examineur
<b>Mme. W. BENZATOUT</b>	Université Dr Tahar Moulay - Saïda	Examinatrice

---

1. e-mail : khadidja.haf@yahoo.com

# Table des matières

<b>1</b>	<b>Les processus stochastiques</b>	<b>9</b>
1.1	Introduction . . . . .	9
1.2	Le processus de comptage . . . . .	10
1.3	Processus de renouvellement . . . . .	12
1.4	Processus de Poisson . . . . .	13
1.4.1	La distribution de Poisson . . . . .	13
1.4.2	La distribution exponentielle . . . . .	14
1.4.3	Loi de Poisson et loi exponentielle . . . . .	15
1.4.4	Perte de mémoire . . . . .	15
1.4.5	Processus de Poisson . . . . .	17
1.5	Processus de naissance et de mort . . . . .	18
<b>2</b>	<b>Files d'attente classiques</b>	<b>20</b>
2.1	Description d'une file d'attente simple . . . . .	21
2.2	Les différents types de files d'attente . . . . .	23
2.3	Notation de Kendall . . . . .	24
2.4	Loi de Little . . . . .	25
2.5	Quelque modèles de files d'attente . . . . .	26
2.5.1	Modèle d'attente M/M/1 . . . . .	26
2.5.2	Modèle d'attente M/M/1/N . . . . .	29
2.5.3	Modèle d'attente M/M/c . . . . .	31
<b>3</b>	<b>Systèmes de files d'attente avec dérobade, abandon et serveurs hétérogènes</b>	<b>35</b>
3.1	Systèmes de files d'attente avec dérobade . . . . .	36
3.1.1	Description du modèle . . . . .	36
3.1.2	Taux de service . . . . .	37
3.1.3	Modèle de files d'attente M/M/1 avec dérobade . . . . .	38
3.2	Systèmes de files d'attente avec abandon . . . . .	39

3.3	Modèle de file d'attente $M/M/2/N$ avec dérobade, abandon et serveurs hétérogènes . . . . .	43
3.3.1	Résultats numériques . . . . .	51
	<b>Bibliographie</b>	<b>57</b>

# Dédicaces

Au nom du dieu le clément et le miséricordieux louange à **ALLAH** le tout puissant.

Je dédie ce modeste travail en signe de respect, reconnaissance et de remerciement :

À mes très chers parents :

Ma mère, qui a œuvré pour ma réussite, de par son amour, son soutien, tous les sacrifices consentis et ses précieux conseils, pour toute son assistance et sa présence dans ma vie, reçois à travers ce travail aussi modeste soit-il, l'expression de mes sentiments et de mon éternelle gratitude.

Mon père "**MOHAMED**", qui peut être fier et trouver ici le résultat de longues années de sacrifices et de privations pour m'aider à avancer dans la vie. Puisse Dieu faire en sorte que ce travail porte son fruit ; je suis très fière d'être votre fille. Merci pour les valeurs nobles, l'éducation et le soutien permanent venu de votre part.

À mes chers frères "**SOFIANE**", "**ZAKARIA**" et "**HOUCINE**" qui n'ont cessé d'être pour moi des exemples de persévérance, de courage et de générosité, et pour leur complicité, à qui je souhaite un avenir radieux plein de réussite.

À mes deux grand-mères pour leur douceur et leur gentillesse.

À mes tantes et mes oncles pour toute l'affection qu'ils m'ont donné et pour leurs précieux encouragements.

À tous les membres de ma promotion.

À mes amis et tous ceux qui me sont chers je leur dédie ce mémoire pour tous les moments de joie qu'on a passé ensemble.

Je prie Dieu, le tout puissant de leur accorder santé, bonheur et succès et j'espère que la vie leur réserve le meilleur.

# Remerciements

En tout premier lieu, je remercie le bon Dieu, tout puissant, de m'avoir donné la force pour survivre, ainsi que l'audace pour dépasser toutes les difficultés.

Je voudrais adresser toute ma gratitude à mon encadreur de mémoire et mon professeur **Dr. Mokhtar KADI** de m'avoir encadré durant ces mois et pour sa patience, sa disponibilité et surtout ses judicieux conseils, qui ont contribué à alimenter ma réflexion orienté aidé et conseillé, et ainsi pour la confiance et la liberté qu'il m'a accordé, sans oublier ses efforts fournis pendant ces deux années.

Je remercie **Dr.L.BOUSMAHA** d'avoir accepté à présider le jury de mon mémoire.

Je souhaite également remercier **Pr.A.KANDOUCI** et **Mme W.BENZATOUT** de m'avoir fait l'honneur d'accepter d'être examinateurs de mon mémoire. Leurs remarques et précieux conseils ont abouti à une amélioration du ce document.

Je remercie le Professeur **A. KANDOUCI** et tout les membres du laboratoire LMSSA pour m'avoir accueilli au sein du labo. Je tiens également à leur remercier pour leur disponibilité, la confiance qu'ils m'ont accordé et pour leurs conseils et leurs commentaires fort utiles qui ont fortement enrichi ma formation.

Mes vifs remerciements aux **Mrs. Abdelhak GUENDOUCI** et **Lahcen YAHIAOUI** pour m'aider à réaliser la partie numérique ainsi que la simulation.

Ma gratitude va également aux enseignants du département de mathématiques que j'ai eu durant mon cursus.

J'adresse mes sincères remerciements à tous les professeurs qui m'ont enseigné.

je remercie mes très chers parents et toute ma famille que Dieu les garde pour leur soutien qu'ils m'ont apporté pendant ces années d'études.

Enfin Je voudrais exprimer ma reconnaissance envers mes amis proches et collègues qui m'ont apporté leur support moral et intellectuel tout au long de ma démarche et je les remercie pour les bons moments passés ensemble et aussi leur encouragements.

Afin de n'oublier personne, mes vifs remerciements s'adressent à tous ceux qui m'ont aidée à la réalisation de ce modeste mémoire.

# Introduction générale

L'attente est devenue un phénomène de la vie contemporaine. Elle varie de l'attente dans une banque, dans un magasin, en guichet de poste, cinéma, les stations d'essence d'un autobus pour rejoindre le travail, d'un ascenseur dans un immeuble, chez le médecin, ...etc. L'analyse de tels phénomènes est devenue des thèmes de recherche traités par la théorie des files d'attente.

Cette théorie date du début du XX<sup>ème</sup> siècle par les travaux de l'ingénieur danois Agner Krarup Erlang (1878, 1929) 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 intégrées dans la modélisation des systèmes informatiques et aux réseaux de communication. Cette intégration dans ces domaines et d'autres a permis une évolution de cette théorie surtout dans l'évaluation des paramètres de performances des ces systèmes. 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.

Dans ce mémoire on s'intéresse à un système de files d'attente avec dérobade, abandon et serveurs hétérogènes. Le phénomène de l'attente provoque l'impatience des clients, qui provoque à son tour une mauvaise qualité de service. Pour éviter ces pertes, des entreprises ont mis en place des processus d'amélioration continue dont le but ultime est l'élimination de toute forme d'attente ou d'impatience. Tous ces exemples révèlent l'importance de l'analyse des files d'attente avec clients impatientes.

Plusieurs chercheurs ont étudié des systèmes de files d'attente avec fonction de dérobade constante ( $0 \leq \beta \leq 1$ ). Certains d'entre eux ont traité le cas de la dérobade en fonction du nombre de clients dans le système  $b_n$  comme Habib [13] et Abou El-ata et Hariri [1] tandis que Morse [27] a discuté le modèle  $M/M/1$  avec la même fonction de dérobade mais sans serveurs hétérogènes. Krishna Kumar et al [20] ont trouvé la solution

transitoire d'une file d'attente  $M/M/2$  avec serveurs hétérogènes.

Ce manuscrit est composé de trois chapitres, dans le premier chapitre nous abordons les processus à la base de l'étude de tels systèmes d'attente comme :

- Le processus de comptage,
- processus de renouvellement,
- processus de Poisson,
- et processus de naissance et de mort.

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 KENDALL, la formule de LITTLE ...) sont notamment données. Ensuite nous étudions quelque modèles de files d'attente markoviennes ( $M/M/1$ ,  $M/M/1/N$ ,  $M/M/c$ ) et l'évaluation de leurs paramètres de performance.

Et enfin le troisième chapitre représente une étude du modèle  $M/M/2/N$  de file d'attente avec dérobade, abandon et serveurs hétérogènes. Quelques exemples numériques pour démontrer comment les différents paramètres du modèle influence sur le comportement du système sont donnés.



# Chapitre 1

## Les processus stochastiques

### 1.1 Introduction

Les processus stochastiques sont un outil très puissant pour la modélisation des phénomènes aléatoires évoluant dans le temps. Leurs utilisations pour la description et l'analyse des propriétés des systèmes dynamiques (files d'attente, réseaux informatiques et téléphoniques, physique, biologie ou économie ... etc.)

#### Définitions et propriétés de base

##### Définition 1.1.1. (*Processus stochastique*)

Un processus stochastique  $X = (X_t)_{t \in T}$  est une famille de variables aléatoires  $(X(t), t \in T)$  ou 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 le processus stochastique  $(X(t), t \in T)$  est dit un processus à temps discret.
- Si  $T$  est un intervalle de  $[0; \infty)$ , alors le processus stochastique  $(X(t), t \in T)$  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.

### Exemples

- **Processus à temps discret** : Le nombre de clients arrivés à une bank, Le nombre d'avion attirer dans un aéroport.
- **Processus à temps continue** : Le temps d'arrivé des clients dans un super marché, les vagues en mer.

## 1.2 Le processus de comptage

Parmi les plus élémentaires des processus, nous trouvons les processus de comptage qui seront bien étudiés par la suite.

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

### Définition 1.2.1. (*Processus de comptage*)

Un processus stochastique  $[N(t); t \in \mathbb{R}_+]$  est un processus de comptage si  $N(t)$  représente le nombre total d'événements qui se sont produits entre 0 et  $t$ , il doit donc satisfaire

- $N(t) \geq 0$  ;
- $N(t)$  a des valeurs entières uniquement ;
- Si  $s < t$ , alors  $N(s) < N(t)$  ;
- Pour  $s < t$ ,  $N(t) - N(s)$  est le nombre d'événements qui ont eu lieu entre  $s$  et  $t$ .

Un processus de comptage est un processus discret à temps continu. Un second processus peut être associé au processus des temps d'occurrence ; le processus des temps d'inter-arrivées  $\{W_n, n \in N_0\}$  ou  $\forall n \in N_0$  la variable aléatoire  $W_n$  est le temps d'attente entre les  $(n-1)$ <sup>ième</sup>,  $n$ <sup>ième</sup> occurrences, c-à-d :

$$W_n = T_n - T_{n-1}$$

avec  $T_n$  est le temps d'arrivé du  $n$ <sup>ième</sup> client.

### Proposition 1.2.1. [36]

Les relations suivantes sont triviales tel que  $T_0 = 0$  à vérifier :

1.  $T_n = W_1 + W_2 + \dots + W_n \quad \forall n \geq 1$  ;
2.  $N(t) = \sup\{n \geq 0 : T_n \leq t\}$  ;
3.  $\mathbb{P}[N(t) = n] = \mathbb{P}[T_n \leq t < T_{n+1}]$  ;
4.  $\mathbb{P}[N(t) \geq n] = \mathbb{P}[T_n \leq t]$  ;
5.  $\mathbb{P}[s < T_n < t] = \mathbb{P}[n(s) < n \leq N(t)]$ .

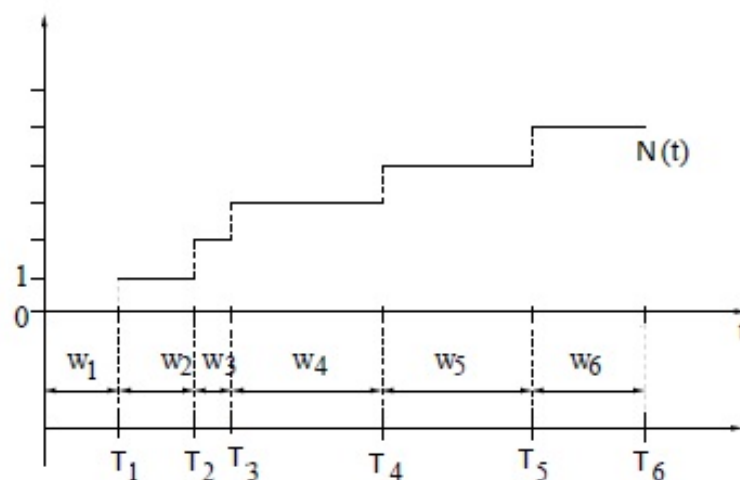


FIGURE 1.1 – Processus de comptage

**Démonstration de (1) :** on a  $W_n = T_n - T_{n-1}$

$$\begin{aligned} W_1 + W_2 + \dots + W_n &= T_1 - T_0 + T_2 - T_1 + T_3 - T_2 + \dots + T_{n-1} - T_{n-2} + T_n - T_{n-1} \\ &= -T_0 + T_n \\ &= T_n \quad \text{car } T_0 = 0. \end{aligned}$$

Dans la suite nous nous intéresserons souvent aux deux propriétés suivantes des processus stochastiques :

**Définition 1.2.2. (Processus à accroissements indépendants)**

Le processus stochastique  $\{X_t, t \in T\}$  est appelé processus à **accroissements indépendants** si

$\forall k \in \mathbb{N}_0$  et  $\forall t_0 < t_1 < t_2 < \dots < t_k$ , les variables aléatoires  $X_{t_1} - X_{t_0}, X_{t_2} - X_{t_1}, \dots, X_{t_k} - X_{t_{k-1}}$  sont indépendantes.

**Définition 1.2.3. (Processus à accroissements stationnaires)**

Le processus stochastique  $\{X_t, t \in T\}$  est dit à **accroissements stationnaires** si  $\forall h > 0$  les variables aléatoires  $X_{t+h} - X_t$  ont la même distribution  $\forall t \in \mathbb{R}_+$ .

**Définition 1.2.4.** Un processus de comptage  $\{N(t), t \geq 0\}$  est un processus de Poisson d'intensité  $\lambda > 0$  si

- $N(0) = 0$ ,
- le processus est à des accroissements indépendants,
- le processus est à des accroissements stationnaires,
- $\forall 0 \leq s < t$  la variable aléatoire  $N(t) - N(s)$  suit une loi de Poisson de paramètre  $\lambda(t - s)$ .

### 1.3 Processus de renouvellement

#### Inter-arrivées indépendantes :

Un processus de renouvellement a pour fonction de dénombrer les occurrences d'un phénomène donné, lorsque les délais entre deux occurrences consécutives sont des variables aléatoires indépendantes et identiquement distribuées. Il peut s'agir de compter le nombre de pannes d'un matériel électronique en théorie de la fiabilité (le matériel est alors renouvelé après chaque panne, d'où la dénomination), de dénombrer les arrivées de clients dans une file d'attente, de recenser les occurrence d'un sinistre pour une compagnie d'assurance . . .

#### Définition 1.3.1. (*Processus de renouvellement*)

Un processus de comptage dont la suite des inter-arrivées forme une suite de variables aléatoires indépendantes et identiquement distribuées s'appelle processus de renouvellement.

**Définition 1.3.2.** La suite  $(S_n)$  est appelée processus de renouvellement associé aux  $(X_n)_{n \geq 0}$  et le processus  $R_t$  est le processus de comptage. Par abus de langage, on appelle également  $R_t$  Processus de renouvellement. Soit  $(X_n)_{n \geq 0}$  une suite de variables aléatoire positives on note  $S_n$  la suite des sommes partielles :  $S_0 = 0$  et  $S_n = X_n + S_{n-1}$  pour tout  $n \geq 1$  on considère alors le processus  $R_t$  défini comme suit :

$$R_t = \text{card}\{n \geq 1, S_n \leq t\} = \sum_{n \geq 1} \mathbf{1}_{\{S_n \leq t\}},$$

Par exemple, si les  $X_n$  modélisent les durées de vie d'une ampoule  $R_t$  représente le nombre d'ampoules changées avant l'instant  $t$ ; les  $X_n$  peuvent également représenter le temps séparant deux ventes successives, ou deux sinistres successifs pour une compagnie d'assurance.  $R_t$  désignera alors, suivant le cas, le nombre d'articles vendus ou le nombre sinistres survenus au cours de l'intervalle de temps  $[0; t]$ .

## 1.4 Processus de Poisson

### Introduction

Le processus de Poisson est un outil extrêmement utilisé pour modéliser les processus stochastiques. Il peut servir à modéliser plusieurs phénomènes de la vie réelle par exemple : les appels téléphoniques arrivant dans une centrale, les temps d'arrivée de clients à une caisse, occurrence d'accidents dans une ville, passage de véhicules a un péage d'autoroute, pannes de machines dans une usine ... etc.

Cette collection des dates d'arrivée s'appelle le processus des arrivées. Lorsque les dates d'arrivées sont imprévisibles, elles sont modélisées par des variables aléatoires.

**Remarque 1.4.1.** *De nombreux phénomènes aléatoires se manifestent par des "arrivées" survenant une par une à des instants aléatoires successifs.*

### 1.4.1 La distribution de Poisson

La loi de Poisson est attribuée à Simeon D. Poisson, mathématicien français (1781 – 1840). Cette loi fut proposée par Poisson dans un ouvrage qu'il publia en 1837 sous le titre " Recherche sur la probabilité de jugements en matière criminelle et en matière civile".

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

$$P_k = \frac{e^{-\lambda} \lambda^k}{k!} \quad \text{pour } k = 0, 1, 2, \dots \quad (1.1)$$

Soit  $X$  une variable aléatoire ayant la distribution de Poisson dans (1.1) 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

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

alors

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

tandis que

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

## 1.4.2 La distribution exponentielle

Soit  $t$  une variable aléatoire avec  $t > 0$  qui suit une distribution exponentielle. La densité de probabilité de  $t$  est  $f(t) = \mu e^{-\mu t}$  et la distribution cumulée correspondante est  $F(t) = 1 - e^{-\mu t}$ .

L'espérance et la variance de  $t$  sont :

$$E(t) = \frac{1}{\mu} \quad \text{et} \quad V(t) = \frac{1}{\mu^2}$$

, respectivement.

### Relation entre la distribution exponentielle et la distribution de Poisson

La densité de probabilité d'une distribution exponentielle  $f(t) = \mu e^{-\mu t}$ . Supposons  $\tau$  est exponentielle avec une espérance  $\frac{1}{\mu}$ , et  $n$  est de Poisson de moyenne  $\mu$  on a :

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

Notons  $\mathbb{P}(n, t)$  la probabilité d'avoir  $n$  unités dans le temps  $t$ .

$$\begin{aligned}
\mathbb{P}(0, t) &= e^{-\mu t} \\
\mathbb{P}(1, t) &= \int_{\tau=0}^t \mathbb{P}(0, \tau) f(1 - \tau) d\tau = \mu t e^{-\mu t} \\
\mathbb{P}(2, t) &= \int_{\tau=0}^t \mathbb{P}(1, \tau) f(1 - \tau) d\tau = (\mu t)^2 e^{-\mu t} / 2! \\
&= \dots \\
\mathbb{P}(n, t) &= \int_{\tau=0}^t \mathbb{P}(n - 1, \tau) f(1 - \tau) d\tau = (\mu t)^n e^{-\mu t} / n!
\end{aligned}$$

### 1.4.3 Loi de Poisson et loi exponentielle

**Définition 1.4.1.** Une variable aléatoire  $X$  à valeurs entières suit une loi de Poisson de paramètre  $\lambda > 0$  si :

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

**Définition 1.4.2.** Une variable aléatoire  $Y$  à valeurs réelles strictement positives suit une loi exponentielle de paramètre  $\mu > 0$  si :

$$\forall t > 0, \quad \mathbb{P}(Y = t) = \mu e^{-\mu t}.$$

### 1.4.4 Perte de mémoire

En probabilités et statistique, la perte de mémoire est une propriété de certaines lois de probabilité : la loi exponentielle et la loi géométrique. On dit que ce sont des lois sans mémoire.

Cette propriété est le plus souvent exprimée en termes de « temps d'attente ». Supposons qu'une variable aléatoire  $X$  soit définie comme le temps passé dans un magasin de l'heure d'ouverture (disons neuf heures du matin) à l'arrivée du premier client. On peut donc voir  $X$  comme le temps qu'un serveur attend avant l'arrivée du premier client.

La propriété de perte de mémoire fait une comparaison entre les lois de probabilité du temps d'attente du serveur de neuf heures à l'arrivée du premier client, et celle du temps d'attente du serveur pour qu'un client arrive à compter d'un délai arbitraire après l'ouverture (disons, par exemple, une heure après l'ouverture soit à partir de dix heures du matin) sachant qu'aucun client n'est arrivé de l'ouverture à l'écoulement de ce délai arbitraire.

La propriété de perte de mémoire affirme que ces lois sont les mêmes.

Ainsi, dans notre exemple, ce n'est pas parce que le serveur a déjà attendu, en vain, pendant une heure l'arrivée d'un premier client qu'il peut espérer que le délai avant qu'arrive effectivement son premier client soit plus faible qu'au moment de l'ouverture.

Les termes de perte de mémoire et sans mémoire ont parfois été utilisés différemment pour faire référence à des processus de Markov, dans ce cas la propriété de Markov assure que les propriétés des variables aléatoires dans le futur dépendent uniquement des informations du temps présent, pas des informations issues du passé. Cependant ces différentes versions de perte de mémoire sont proches d'un point de vue théorique.

L'unique loi de probabilité continue à perte de mémoire est la loi exponentielle, ainsi la propriété de perte de mémoire caractérise la loi exponentielle parmi toutes les lois continues.

### Propriété sans mémoire de la distribution exponentielle

Quand une variable aléatoire  $t$  est exponentielle, la densité de probabilité est

$$f(t) = \mu e^{-\mu t},$$

et la distribution cumulée correspondante est

$$F(t) = 1 - e^{-\mu t}.$$

Pour un accroissement de temps  $h$ , la probabilité que  $t$  est supérieur à  $h$  devient

$$P(t > h) = e^{-\mu h},$$

de plus pour  $t = (t' + h)$ , la probabilité de  $t$  est plus grande que  $(t' + h)$  est

$$P(t > (t' + h)) = e^{-\mu(t'+h)},$$

la probabilité conditionnelle de  $t > (t' + h)$  sachant  $t > t'$  est

$$P(t > t' + h | t > t') = e^{-\mu(t'+h)} / e^{-\mu t'} = e^{-\mu h}.$$

Du fait que les deux probabilités sont les mêmes, la distribution exponentielle est appelée une distribution de probabilité sans mémoire.



On s'intéresse ici au comptage du nombre d'occurrences d'un événement, par exemple la naissance d'un individu. On note  $N_t$  le nombre d'événements survenus dans l'intervalle  $[0, t]$ . Un tel processus a une trajectoire en escalier

$$\begin{aligned} N_t &= 0 \quad \text{si } t < t_1 \\ &= 1 \quad \text{si } t_1 \leq t \leq t_2 \\ &\vdots \\ &= k \quad \text{si } t_k \leq t \leq t_{k+1} \\ &\text{etc.} \end{aligned}$$

### 1.4.5 Processus de Poisson

De tels phénomènes peuvent se définir par la famille  $(T_n)_{n \in \mathbb{N}^*}$  des temps d'arrivées qui sont des variables aléatoires. Mais on peut aussi le faire à partir du processus de comptage  $(N_t)_{n \in \mathbb{R}}$ .  $N_t$  est le nombre d'événements apparus jusqu'à l'instant  $t$ .

#### Définition 1.4.3. (*Processus de Poisson*)

Soit  $(X_n)$  une suite de variables aléatoires indépendantes et identiquement distribuées de loi exponentielle  $\varepsilon(\lambda)$  avec  $\lambda > 0$  si  $S_n = X_1 + X_2 + \dots + X_n$ ,  $N_0 = 0$  et pour tout  $t > 0$

$$N_t = \sum_{n=1}^{\infty} \mathbb{1}_{S_n \leq t},$$

$(N_t)$  est un processus de Poisson d'intensité  $\lambda$ .

#### Définition 1.4.4. (*Processus de Poisson*)

Un processus de Poisson  $N = (N_t)$  d'intensité  $\lambda$  est un processus de comptage à trajectoires continues à droite tel que :

- $N(0) = 0$ ,
- $N$  est un processus à accroissements indépendants et stationnaires,
- pour tout  $t \geq 0$ ,  $N_t$  suit la loi de poisson  $\mathcal{P}(\lambda t)$ .

**Proposition 1.4.1.**  $(N_t)_{t \geq 0}$  est un processus de Poisson si et seulement si les quatre propriétés suivantes sont vérifiées :

- $N(0) = 0$ ,
- pour tout  $t > 0$ ,  $N_t$  une loi de Poisson d'espérance  $\lambda t$ ,

- pour tous  $t > s \geq 0$ ,  $N_t - N_s$  suit une loi de Poisson d'espérance  $\lambda(t - s)$ ,
- pour tous  $0 \leq t_1 < t_2 < \dots < t_n$  les variables aléatoires  $N_{t_1}, N_{t_2} - N_{t_1}, \dots, N_{t_n} - N_{t_{n-1}}$  sont indépendantes.

### Caractérisation d'un processus de Poisson par ses temps d'arrivé :

soit  $T_n$  l'instant de la  $n^{\text{ième}}$  arrivé :  $T_n = \inf\{T \geq 0, N_t = n\}$  et  $W_n$  est le temps séparant le  $(n - 1)^{\text{ième}}$  évènement du  $n^{\text{ième}}$  évènement pour  $n \in \mathbb{N}^*$   $W_n = T_n - T_{n-1}$  (en convenant  $T_0 = 0$ ) on a :

$$T_n = \sum_{i=1}^n W_i,$$

$$N_t = \max\{n \geq 0, T_n \leq t\}.$$

## 1.5 Processus de naissance et de mort

### Introduction

Ces processus permettent de façon générale de décrire l'évolution temporelle de la taille d'une population d'un type donné.

Dans le cas d'un système d'attente, on considère par exemple des populations comprenant tous les clients qui sont dans le système à l'instant  $t$ .

Les processus de naissance et de mort sont des processus stochastiques à temps continu et à espace d'états discret  $n = 0, 1, 2, \dots$ . Ils sont sans mémoire, et à partir d'un état donné  $n$ , des transitions ne sont possibles que vers l'un ou l'autre des états voisins

$$(n + 1) \text{ et } (n - 1) \quad \text{pour } n \geq 1$$

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.

### Définition 1.5.1. (*Processus de naissance et de mort*)

On peut réaliser un processus de naissance et de mort de la façon suivante :

- Les arrivées et les départs d'entités obéissent à des lois exponentielles de taux respectifs  $\lambda(n)$  et  $\mu(n)$ .
- A l'aide d'hypothèse de régularité : deux événements ne peuvent pas se produire en même temps, donc la probabilité que deux événements se produisent dans un intervalle de temps  $dt$  est négligeable.
- Il y a une transition vers un état voisin, soit par l'arrivée d'un client (naissance), soit par le départ d'un client (mort).

C'est le point de départ de la théorie des files d'attente. On introduit les données suivantes :

- $\lambda_n$  : taux de naissances quand le nombre de population égale à  $n$ .
- $\mu_n$  : taux de morts quand le nombre de population égale à  $n$ .

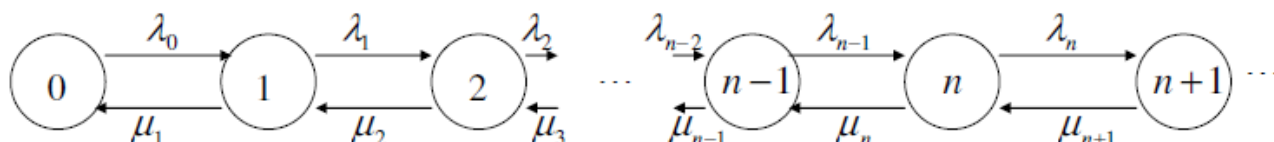


FIGURE 1.2 – Diagramme de transition d'un processus de naissance et de mort

### Exemple

Les modèles de files d'attente les plus simples et les plus fréquents sont ceux qui ont un processus de Poisson pour les arrivées et un temps de service exponentiellement distribué. Dans ce cas la taille de la file d'attente constitue un processus de naissance et de mort car :

- Naissance  $\longleftrightarrow$  arrivée du client.
- Mort  $\longleftrightarrow$  départ du système après son service.

Pour le processus d'arrivée Poissonien nous avons :

1.

$$\mathbb{P}(N(t+h) - N(t) = 1) = \lambda h + o(h)$$

avec  $N(t)$  : est le nombre de clients arrivants à l'instant  $t$ .

2.  $\mathbb{P}(\text{aucune arrivée dans } [t, t+h]) = \mathbb{P}(N(t+h) - N(t) = 0) = 1 - \lambda h + o(h)$ .

3. De même la propriété exponentielle de la distribution exprimée par son taux constant implique que :

$$\mathbb{P}(\text{un service est achevé en } [t, t+h] \text{ / le service a commencé à l'instant } t) = \mu h + o(h).$$

4.  $\mathbb{P}(\text{un service n'est pas achevé en } [t, t+h] \text{ / le service a commencé à l'instant } t) =$

$$\mathbb{P}(N(t+h) - N(t) = 0) = 1 - \mu h + o(h).$$

Dans le processus est de naissance et de mort :

- Naissance  $\longleftrightarrow$  Le temps s'écoulant entre deux naissances consécutives est distribué exponentiellement.
- Mort  $\longleftrightarrow$  Le temps s'écoulant entre deux morts consécutives est aussi distribué exponentiellement.

# Chapitre 2

## Files d'attente classiques

Les files d'attente peuvent être considérées comme un phénomène caractéristique de la vie contemporaine. On les rencontre dans les domaines d'activité les plus divers. L'étude mathématique des phénomènes d'attente constitue un champ d'application important des processus stochastiques.

Les travaux de l'ingénieur danois Agner Krarup Erlang (1878, 1929) sur le trafic téléphonique de Copenhague sont les premières études sur les files d'attente. Ensuite les applications dans les réseaux informatiques et télécommunications au début du XXème siècle ont permis une évolution de cette théorie.

La Théorie des files d'attente est une technique de la recherche opérationnelle qui permet de modéliser un système admettant un phénomène d'attente, de calculer ses performances et de déterminer ses caractéristiques pour aider les gestionnaires dans leurs prises de décisions.

On s'intéresse essentiellement à deux grandeurs : le nombre de clients dans le système, et le temps passé par un client dans le système. Ce dernier se décompose en un temps d'attente et un temps de service.

◊ Une file d'attente est définie par :

- La suite des instants d'arrivées des clients.
- La suite des temps de service des clients.
- La discipline de service qui donne l'ordre dans lequel seront servi les clients.
- La capacité de la file.
- Le nombre de serveurs.
- Population totale de clients (rare).

## 2.1 Description d'une file d'attente simple

### File d'attente simple

#### Définition 2.1.1. (*File d'attente simple*)

Une file simple (ou station) est un système constitué d'un ou plusieurs serveurs et d'un espace d'attente qui contient une ou plusieurs places. Les clients arrivent de l'extérieur, patientent éventuellement dans la file d'attente, reçoivent un service, puis quittent la station.

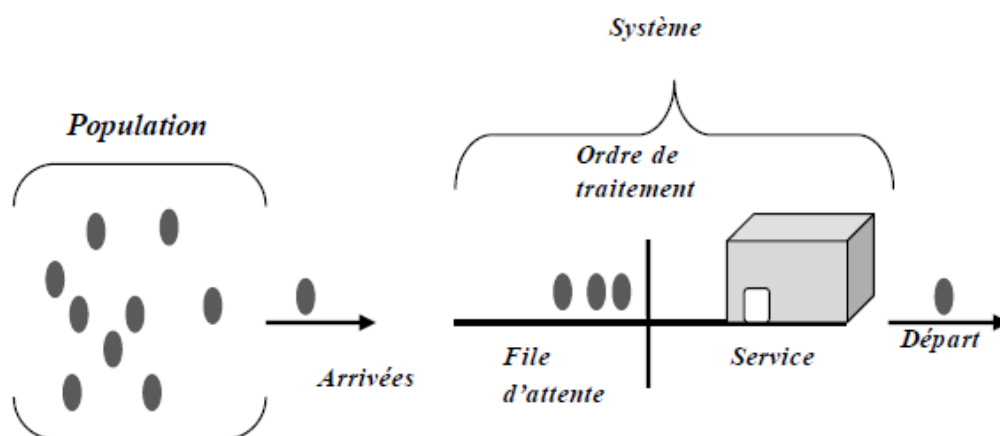


FIGURE 2.1 – Système de File d'attente simple

**Clients** : sont les objets ou les personnes qui arrivent dans le système. On suppose que les clients arrivent un par un, et sont servis par leur ordre d'arrivée.

**Interarrivée** : est l'intervalle de temps séparant l'arrivée de deux clients successifs. On suppose que ces Inter-arrivées sont indépendantes les unes des autres et suivent la même loi de probabilité.

**Service** : le temps que le client passe au guichet par exemple. Le temps de séjour total dans le système est la somme du temps d'attente et de la durée du service effectif.

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

- Le processus d'arrivée : Pour le processus d'arrivée, on s'intéresse aux instants d'arrivées des clients dans la file. Ils sont en général aléatoires. Certaines hypothèses sont faites 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. Évidemment d'autres cas peuvent se présenter : temps d'interarrivés constants, de loi uniforme, loi normale ou encore gamma.

- Le processus de service : qui compris (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.

### Terminologie et notations

- $\lambda_n$  : Taux moyen d'arrivée (espérance mathématique du nombre d'arrivées par unité de temps).
- $\frac{1}{\lambda_n}$  : Temps moyen entre les arrivées lorsque les clients sont dans le système.
- $\mu_n$  : Taux moyen de service d'un client.
- $\frac{1}{\mu_n}$  : Temps moyen de service du  $n^{\text{ième}}$  client.

Quand nous commençons à analyser un système de file d'attente, l'état de ce dernier dépend beaucoup de l'état initial et du temps écoulé. Nous disons alors que le système est en situation transitoire, et son étude est alors très complexe.

C'est pourquoi dans la théorie des files d'attente, nous préférons faire l'étude une fois que le système a atteint sa stabilité d'où les états du système sont essentiellement indépendantes de l'état initial et du temps déjà écoulé.

### Notation et terminologie en situation d'équilibre

- $P_n$  : Probabilité qu'il y ait  $n$  clients dans le système.
- $L_s$  : Nombre moyen (espérance mathématique) de client dans le système.
- $L_q$  : Nombre moyen de client dans la file d'attente (excluant ceux dans le service).
- $W_s$  : Temps moyen dans le système.
- $W_q$  : Temps moyen dans la file (excluant le temps de service).
- $c$  : Nombre de serveurs.

## 2.2 Les différents types de files d'attente

Les figures suivantes représentent les différents systèmes de files d'attente selon l'espace d'attente et l'espace de service :



FIGURE 2.2 – File d'attente avec un seul espace d'attente et un seul serveur

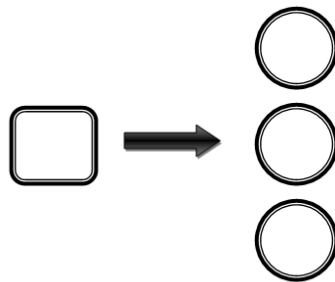


FIGURE 2.3 – File d'attente avec un seul espace d'attente et plusieurs serveurs

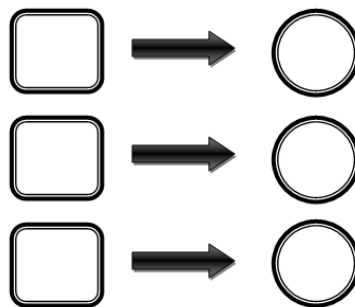


FIGURE 2.4 – File d'attente avec plusieurs espaces d'attente et plusieurs serveurs

## 2.3 Notation de Kendall

Il existe évidemment, de très nombreux systèmes de files d'attente. La notation de Kendall permet de décrire le système par une suite de 6 symboles :

$$A/B/m/K/N/Z$$

Voici la signification de ces symboles :

1. A : Processus d'arrivée des clients (distribution d'interarrivée). Les symboles utilisés sont :
  - ▷ M : (loi exponentielle) Inter-arrivées des clients sont identiquement distribuées selon une loi exponentielle. Il correspond à un processus de Poisson ponctuel (propriété sans mémoire).
  - ▷ D : (loi déterministe) Les temps inter-arrivées des clients ou les temps de service sont constants et toujours les mêmes.
  - ▷ GI : (général indépendant) Inter-arrivées des clients ont une distribution générale (il n'y a aucune hypothèse sur la distribution mais les Inter-arrivées sont indépendantes et identiquement distribuées).
  - ▷ G : (loi générale) Inter-arrivées des clients ont une distribution générale (il n'y a aucune hypothèse sur la distribution mais les inter-arrivées sont indépendantes et identiquement distribuées).
  - ▷  $E_k$  : (loi d'Erlang d'ordre 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. B : Processus de service (distribution de durée de service de clients).
3. m : nombre de serveurs.
4. K : capacité maximale de la file d'attente.
5. N : nombre de clients utilisant le système.
6. Z : discipline du système qui décrit la façon dont les clients sont ordonnancés.

Le nombre de serveurs peut varier de 1 à l'infini noté ( $\infty$ ), de même pour  $K$  et  $N$ . En ce qui concerne  $Z$ , les ordonnancements les plus utilisés sont :

- 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 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 mono serveur.
- RANDOM or SIRO (Served In Random Order) : les prochains clients sont servis aléatoirement.
- Round-Robin (cyclique avec un quantum  $Q$ ) : Tous les clients de la file d'attente entrent en service à tour de rôle, effectuant un quantum  $Q$  de leur temps de service et sont replacés dans la file, jusqu'à ce que leur service soit totalement accompli.
- PS (Processor Sharing) : C'est le cas limite de la distribution Round-Robin lorsque le quantum de temps  $Q$  tend vers 0. Tous les clients sont servis en même temps, mais avec une vitesse inversement proportionnelle au nombre de clients simultanément présents. Si le taux du serveur est égal à  $\mu$  et qu'à un instant donné il y a  $n$  clients à la station, tous les clients sont donc servis simultanément avec un taux  $\frac{\mu}{n}$ .
- PRIOR (Avec priorité, avec préemption ou sans préemption) or PNP (Priorité service), les clients sont servis selon leur priorité. Tous les clients de la plus haute priorité sont servis premiers, puis les clients de priorité inférieure sont servis, et ainsi de suite.

## 2.4 Loi de Little

### Introduction

La loi de Little (1961) est une relation très général qui s'applique à une grande classe de systèmes. Elle ne concerne que le régime permanent du système. Aucune hypothèse sur les variables aléatoires qui caractérisent le système (temps d'inter arrivées, temps de service, ...) n'est nécessaire. La seule condition d'application de la loi de Little est que le système soit stable.

$$d_s = d_e = d$$

La loi de Little s'exprime telle que dans le théorème 2.4.1 suivant :

**Théorème 2.4.1. (Formule de Little)**

Le nombre moyen de clients  $L$ , le temps moyen passé dans le système  $W$  et le débit moyen d'un système stable  $d$  en régime permanent se relient de la façon suivante :

$$L = W \times d.$$

**Remarque 2.4.1.** La loi de Little s'applique à tous les modèles de file d'attente rencontrés en pratique (pas seulement à la file  $M/M/1$ ).

## 2.5 Quelques modèles de files d'attente

### 2.5.1 Modèle d'attente $M/M/1$

#### Description du modèle

La file d'attente  $M/M/1$  est la file d'attente la plus simple, elle comporte un seul serveur offrant un service à des clients qui arrivent selon le processus de Poisson de paramètre  $\lambda > 0$ , le temps de service pour chaque client suit une loi exponentielle de paramètre  $\mu > 0$ .

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 inter arrivées sont indépendants. Ces clients reçoivent le service selon une distribution exponentielle de paramètre  $\mu$ . Ces temps de service sont également supposés indépendants [23]. En outre, toutes les variables aléatoires concernées sont censés être indépendants les uns des autres. Les clients sont servis dans leur ordre d'arrivée. Cette file est un cas particulier du processus de naissance et de mort chaque arrivée d'un client pouvant être assimilée à une naissance (taux de naissance est égal à  $\lambda$ ), et chaque départ correspond à une mort (taux de mort est égal à  $\mu$ ).

$$\begin{cases} \lambda_n = \lambda & (n \geq 0) \\ \mu_n = \mu & (n \geq 1) \end{cases}$$

$$\forall n \in \mathbb{N}$$

i.e. indépendants du nombre de client dans le système. sous l'hypothèse que  $\lambda < \mu$  (le taux d'arrivée est plus petit que le taux de service)

on a :

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

c'est l'intensité du trafic, avec  $\rho < 1$  :

$$\left\{ \begin{array}{l} \pi_n = \pi_0 \rho^n \\ \pi_0 = \frac{1}{\sum_{n=0}^{+\infty} \rho^n} = 1 - \rho \end{array} \right.$$

**Remarque 2.5.1.** on note  $\pi_0$  la probabilité d'être servi immédiatement à l'arrivée.

La condition  $\lambda < \mu$  est bien entendu équivalente à la condition  $\rho < 1$ . Ainsi, un régime stationnaire peut exister si et seulement si l'intensité de trafic est inférieure à 1. On voit aussi le rôle de ce paramètre lorsque l'on calcule le nombre moyen de clients en régime stationnaire

avec

$$\sum_{n=0}^{\infty} \left( \frac{\lambda}{\mu} \right)^n = \frac{1}{1 - \frac{\lambda}{\mu}},$$

$\pi_n$  : c'est la probabilité d'avoir n client dans le système.

$$\pi_n = \rho^n \pi_0,$$

avec

$$\pi_0 = 1 - \rho.$$

**Débit  $d$  :**

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

Ici  $d = \lambda$  car  $\lambda_n = \lambda$  pour tout  $n \geq 0$ .

$$d = \sum_{n=0}^{\infty} \pi_n \mu = (1 - \pi_0) \mu = \rho \mu = \lambda,$$

donc

$$d = d_e = d_s.$$

**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

$$U = 1 - \pi_0 = \rho.$$

Nombre moyen de clients dans le système  $L$  :

$$\sum_{n=0}^{\infty} n\pi_n = \frac{\rho}{1-\rho} = \frac{\frac{\lambda}{\mu}}{1 - \frac{\lambda}{\mu}} = \frac{\lambda}{\mu - \lambda}.$$

Temps moyen de séjour  $W$  :

Ce paramètre est obtenu en utilisant la loi de Little :

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

qui peut se décomposer en :

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

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

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

La durée moyenne d'attente en régime stationnaire :

$$E(W) = \frac{\rho}{\lambda(1-\rho)}.$$

**Proposition 2.5.1.** *Soit  $0 < \lambda < \mu$ . Dans le système M/M/1 la variable aléatoire  $W$  égale à la durée de séjour des clients dans le système en régime stationnaire suit une loi exponentielle de paramètre :  $\mu(1-\rho) = \mu - \lambda$ .*

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  $L_q$  :

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

## 2.5.2 Modèle d'attente M/M/1/N

### Description du modèle

On considère un système à serveur simple identique à la file  $M/M/1/N$  excepté que la capacité de la file d'attente est finie. On a donc toujours les hypothèses suivantes :

- Le processus d'arrivée des clients dans la file est un processus de Poisson de taux  $\lambda$ .
- Le temps de service d'un client est une variable aléatoire exponentielle de taux  $\mu$ .
- Soit  $N$  la capacité de la file d'attente : c'est le nombre maximal de clients qui peuvent être présents dans le système, soit en attente, soit en service.

Quand un client arrive alors qu'il y a déjà  $N$  clients présents dans le système, il est perdu. Ce système est connu sous le nom de file  $M/M/1/N$ . L'espace d'états  $E$  est maintenant fini :  $E = \{0, 1, 2, \dots, N\}$ .

La capacité de la file étant limitée, même si les clients arrivent en moyenne beaucoup plus vite que ce que le serveur de la file est capable de traiter, dès que celle-ci est pleine, les clients qui se présentent sont rejetés. Le nombre de clients dans la file ne peut donc jamais "partir" à l'infini. De plus, dès qu'un client est autorisé à entrer, il sortira un jour et son temps de séjour dans la file est fini, puisqu'il correspond au temps de service de tous les clients devant lui et que ce nombre est limité par  $N$ . Le débit de sortie sera donc bien égal au débit d'entrée, ce qui correspond bien à la stabilité inconditionnelle du système. Donc ce processus est considéré comme un processus de naissance et de mort avec :

- Un taux de naissance  $\lambda_n = \lambda$ , pour tout  $n < N$
- Et le taux de mortalité  $\mu_n = \mu$ , pour tout  $n \neq 0$ .

Soit  $\pi_n, n = 0, 1, \dots, N$ , la probabilité pour qu'il ait  $n$  clients dans le système à l'instant  $t$ .

$$\begin{aligned} \lambda\pi_0 &= \mu\pi_1 \\ (\lambda + \mu)\pi_n &= \lambda\pi_{(n-1)} + \mu\pi_{(n+1)} \quad \text{pour } n = 1, 2, \dots, N-1, \\ \lambda\pi(N-1) &= \mu\pi(N). \end{aligned}$$

Le calcul de  $\pi_n$  se fait de la manière suivante :

$$\pi_n = \begin{cases} \frac{(1-\rho)\rho^n}{1-\rho^{N+1}} & \text{pour } n \leq N, \\ 0 & \text{pour } n > N. \end{cases}$$

Cas particulier  $\rho = 1$

Dans ce cas

$$\pi_n = \begin{cases} \frac{1}{N+1} & \text{pour } n \leq N, \\ 0 & \text{pour } n > N. \end{cases}$$

**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  $\mu$  (le débit du sortie du service) donc :

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

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

$$\begin{aligned} d_e &= \mathbb{P}(\text{[file non pleine aux instants d'arrivée]})\lambda \\ &= \sum_{n=0}^{N-1} \pi_n \lambda = (1 - \pi_N)\lambda = \frac{1 - \rho^N}{1 - \rho^{N+1}}\lambda. \end{aligned}$$

Comme

$$\rho = \frac{\lambda}{\mu},$$

donc

$$d_e = d_s = d.$$

**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 occupé donc au moins il y a un client dans la file

$$\begin{aligned} U &= \sum_{n=1}^N \pi_n = 1 - \pi_0 \\ &= \rho \frac{1 - \rho^N}{1 - \rho^{N+1}}. \end{aligned}$$

Remarque :

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

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

Nombre moyen de clients L :

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

Pareille que le taux d'utilisation du serveur, lorsque  $N \rightarrow +\infty$  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 :

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

### 2.5.3 Modèle d'attente M/M/c

**Description du modèle**

Pour ce modèle de file d'attente, il existe  $c$  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  $\lambda$  et le temps de service d'un client est une variable aléatoire exponentielle de paramètre  $\mu$ .

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

$$\lambda_n = \lambda,$$

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

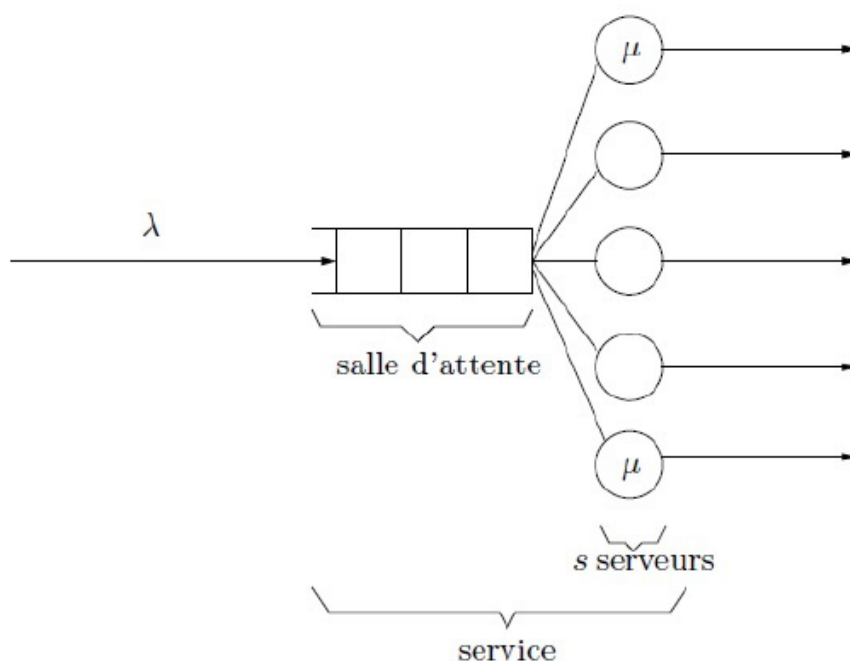


FIGURE 2.5 – Représentation schématique d'une file M/M/c

Pour calculer  $\pi_n$ , le système doit être stable ( $\lambda < c\mu$  donc  $\rho < c$ ) 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 que les serveurs de la file sont capables de traiter par unité de temps.

Alors

$$\pi_n = \begin{cases} \pi_0 \frac{\rho^n}{n!} & \text{pour } n = 1, 2, \dots, c-1 \\ \pi_0 \rho^n \frac{c^{c-n}}{c!} & \text{pour } n \geq c \end{cases}$$



avec

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

**Remarque :**

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

$$\pi_n = (1 - \rho)\rho^n$$

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

**Débit  $d$  :**

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

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

En remplaçant les expressions obtenues pour les probabilités  $\pi_n$  et  $\pi_0$ , on trouve bien que la file est stable :

$$d = \lambda,$$

pour des raison de calcul pour la file  $M/M/c$ , 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,  $L_q$  :

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

On en déduit l'expression du temps moyen de séjour

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

Nombre moyen de clients  $L$  :

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

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

## Chapitre 3

# Systemes de files d'attente avec dérobade, abandon et serveurs hétérogènes

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 impatients 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 décide 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.

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 intéressés par l'étude de ces modèles, tels que Haight, 1957, [14] qui a étudié le modèle d'attente  $M/M/1$  avec dérobade, puis, en 1959, [15] il a traité ce même modèle  $M/M/1$  avec abandon. Ce sont ensuite Ancker et Gafarian [3, 4] qui ont étudié l'effet de la combinaison entre le dérobade et l'abandon 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 [34] ont fait l'extension 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 Systèmes de files d'attente avec dérobage

Les systèmes de files d'attente avec dérobage sont des modèles largement utilisés dans des problèmes de la vie réelle tels que les centrales d'appels téléphoniques, les urgences des hôpitaux .... À son arrivée au système, le client découragé par la longueur de la file d'attente par exemple, décide de ne pas rejoindre la file d'attente, c'est le cas du dérobage. Cette impatience a poussé beaucoup de mathématiciens [12] 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

$$P_n = \begin{cases} 1 & \text{pour } n < c, \\ 0 & \text{pour } n \geq c. \end{cases}$$

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  $\lambda$ . 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

$$\lambda_n = \lambda P_n \quad \text{pour } n = 0, 1, \dots .$$

Et dans le cas d'un seul serveur,  $\mu_n = \mu$  pour  $n = 1, 2, \dots$  et nous pouvons évaluer la distribution stationnaire  $\pi_n$  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

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

La vitesse à laquelle les clients sont perdus est

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

et le taux 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  $M/M/c$ . A son arrivée un client entre dans le système, si et seulement si un serveur est libre. Pour cela nous avons

$$\lambda_N = \begin{cases} \lambda & \text{pour } N = 0, 1, \dots, c-1, \\ 0 & \text{pour } N = c, \end{cases}$$

et

$$\mu_N = N\mu \quad \text{pour } N = 0, 1, \dots, c,$$

afin de déterminer la distribution, nous avons

$$\pi_N = \frac{1}{N!} \left( \frac{\lambda}{\mu} \right)^N \quad \text{pour } N = 0, 1, \dots, c,$$

et par conséquent

$$\pi_N = \frac{\frac{1}{N!} \left( \frac{\lambda}{\mu} \right)^N}{\sum_{j=0}^c \frac{1}{j!} \left( \frac{\lambda}{\mu} \right)^j} \quad \text{pour } N = 0, 1, \dots, c.$$

La proportion des clients perdus est  $\pi_c q_c = \pi_c$ ; d'où  $q_c = 1$ .

### 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 Poissonniennes et les taux de service sont sans mémoire, les paramètres de naissance et de mort appropriés sont

$$\lambda_N = \lambda \quad \text{pour } N = 0, 1, \dots,$$

et

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

Plus généralement, nous considérons des arrivées Poissonniennes  $\lambda_N = \lambda$  pour  $N = 0, 1, \dots$  et les taux de service arbitraires  $\mu_N$  pour  $N = 0, 1, \dots$ . La distribution stationnaire dans ce cas, est donnée par

$$\pi_N = \frac{\pi_0 \lambda^N}{\mu_1 \mu_2 \cdots \mu_N} \quad \text{pour } N \geq 1, \quad (3.1)$$

où la condition de normalisation nous permet d'écrire :

$$\pi_0 = \left[ 1 + \sum_{N=1}^{\infty} \frac{\lambda^N}{\mu_1 \mu_2 \cdots \mu_N} \right]^{-1}. \quad (3.2)$$

### 3.1.3 Modèle de files d'attente M/M/1 avec dérobage

Dans ce modèle, un client arrivant à un système qui trouve  $N$  clients devant lui, il devient impatient et décide de ne pas rejoindre la file. Notons  $b_N$  la probabilité qu'un client rejoigne le système lorsque il y a  $N$  clients dans le système au moment de son arrivée. Donc le nombre de clients dans le système est un processus de naissance et de mort avec : des taux de naissances

$$\lambda_N = \lambda b_N \quad N = 0, 1, \dots,$$

tel que

$$b_N = \frac{1}{N+1} \quad N = 0, 1, \dots,$$

ainsi

$$P_N = \frac{\rho^N}{N!} P_0 \quad N = 0, 1, \dots,$$

puis en utilisant la condition de normalisation, nous obtenons

$$P_N = \frac{\rho^N}{N!} e^{-\rho} \quad N = 0, 1, \dots$$

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  $U$  :**

$$\begin{aligned} U &= 1 - P_0 = 1 - e^{-\rho}, \\ &= \frac{\mathbb{E}(\delta)}{\frac{1}{\lambda} + \mathbb{E}(\delta)}, \end{aligned}$$

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

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

Nombre moyen de clients dans le système  $L_s$  :

$$L_s = \rho$$

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

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

$$L_q = L_s - U_s = \rho - (1 - e^{-\rho}) = \rho + e^{-\rho} - 1$$

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

$$d = \sum_{N=0}^{\infty} \lambda_N P_N = \sum_{N=1}^{\infty} \mu_N P_N = \mu(1 - e^{-\rho}),$$

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.

## 3.2 Systèmes de files d'attente 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 [31]. Après un temps passé dans la file, le client impatient décide de quitter le système sans avoir le service.

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  $\mu$ .
- 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  $\alpha$ .

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,

$$\frac{d}{dt}P_0(t) = -\lambda P_0(t) + \mu P_1(t) \quad (3.3)$$

$$\begin{aligned} \frac{d}{dt}P_n(t) = & - \left[ \left( \frac{\lambda}{n+1} \right) + \mu + (n-1)\alpha P \right] P_n(t) + (\mu + n\alpha P)P_{n+1}(t) \\ & + \left( \frac{\lambda}{n} \right) P_{n-1}(t); \quad n = 1, \dots, N-1 \end{aligned} \quad (3.4)$$

$$\frac{d}{dt}P_N(t) = -[\mu + (N-1)\alpha P] P_N(t) + \left( \frac{\lambda}{N} \right) P_{N-1}(t) \quad (3.5)$$

En état d'équilibre,  $\lim_{t \rightarrow +\infty} P_n(t) = P_n$  et,  $\frac{d}{dt}P_n(t) = 0$  et donc, la solution des équations (3.3) à (3.5) donne :

$$0 = -\lambda P_0 + \mu P_1 \quad (3.6)$$



$$0 = - \left[ \left( \frac{\lambda}{n+1} \right) + \mu + (n-1)\alpha P \right] P_n + (\mu + n\alpha P) P_{n+1} + \left( \frac{\lambda}{n} \right) P_{n-1}; \quad n = 1, \dots, N-1 \quad (3.7)$$

$$0 = - [\mu + (N-1)\alpha P] P_N + \left( \frac{\lambda}{N} \right) P_{N-1}. \quad (3.8)$$

En résolvant les équations de (3.6) à (3.8), nous obtenons

$$P_n(t) = - \left[ \frac{1}{n!} \prod_{k=1}^n \frac{\lambda}{\mu + (k-1)\alpha P} \right] P_0; \quad n = 1, \dots, N \quad (3.9)$$

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

$$P_0 = \frac{1}{1 + \sum_{n=1}^N \left( \frac{1}{n!} \prod_{k=1}^n \frac{\lambda}{\mu + (k-1)\alpha P} \right)}. \quad (3.10)$$

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  $L_s$  :

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

Nombre moyen de clients dans la file  $L_q$  :

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

Temps moyen de séjour dans le système  $W_s$  :

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

Temps moyen de séjour dans la file  $W_q$  :

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

Le nombre moyen de clients servis  $E(\text{client servis})$  :

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

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

Taux d'abandon  $R_{\text{abond}}$  :

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

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

Le nombre de clients en attente  $E(\text{client en attente})$  :

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

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

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

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

Où  $P_0$  est la valeur calculée en (3.10).

### 3.3 Modèle de file d'attente $M/M/2/N$ avec dérobade, abandon et serveurs hétérogènes

#### Introduction

Dans cette partie, on va étudier un modèle de files d'attente  $M/M/2/N$  (la capacité du système  $N$  étant limité). Avec une fonction de dérobade abandon et deux serveurs hétérogènes. La discipline de la file est FCFS (first come, first served).

Après en calcul les paramètres de performance de ce système.

#### Description du modèle

Les clients arrivent de l'extérieur au système selon la distribution de Poisson de paramètre  $\lambda$  et rejoignent la file d'attente avec la probabilité  $e^{-\frac{\gamma(n-1)}{\mu_3}}$  où  $\gamma$  est la mesure moyenne de l'impatience des clients et  $\frac{n-1}{\mu_3}$  est l'estimation du client de son temps d'attente.

Une fois dans la file d'attente, si le temps d'attente expire le client peut abandonner selon une distribution exponentielle  $f(t) = \alpha e^{-\alpha t}$ ,  $t > 0$  avec le paramètre  $\alpha$  et la probabilité d'abandon dans une courte période de temps  $\Delta_t$  est :  $r_n = (n-2)\alpha\Delta_t$  ( $n = 2, 3, \dots, N$ ).

Les temps de service des clients sont exponentiellement distribué avec des taux

- $\mu_1$  : Pour le premier serveur.
- $\mu_2$  : Pour le deuxième serveur.
- ( $\mu_3 = \mu_1 + \mu_2$ ).

Les clients sont servis selon les conditions suivantes :

1. Si les deux serveurs sont occupés, les clients rejoignent la file d'attente.
2. Si l'un des deux serveurs est libre, le client à la tête de la file l'occupe.
3. Si les deux serveurs sont libres, le client à la tête de la file d'attente choisit le serveur 1 avec probabilité  $\pi_1$  ou serveur 2 avec probabilité  $\pi_2$ ,  $\pi_1 + \pi_2 = 1$ .

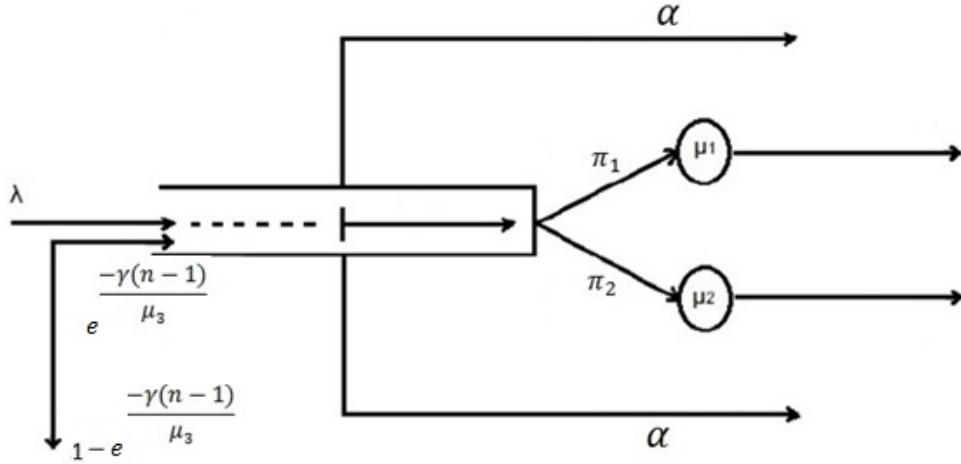


FIGURE 3.1 – Modèle  $M/M/2/N$  avec dérobade, abandon, et deux serveurs hétérogènes

### Les équations d'équilibre et leur solutions

Nous définissons les probabilités d'équilibre comme suit :

- $P_{0,0} = \mathbb{P}(\text{Il n'y a aucun client dans le système}),$
- $P_{1,0} = \mathbb{P}(\text{Il y a un client entrain d'être servi par le serveur 1}),$
- $P_{0,1} = \mathbb{P}(\text{Il y a un client entrain d'être servi par le serveur 2}),$
- $P_n = \mathbb{P}(\text{Il y a } n \text{ clients dans le système}), \quad n = 2, 3, \dots, N.$

Et on pose encore,

$$P_0 = P_{0,0}, \quad P_1 = P_{1,0} + P_{0,1} \text{ et } P_2 = P_{1,1}.$$

Par conséquent, les équations de probabilité à l'état stable sont :

$$\lambda P_0 = \mu_1 P_{1,0} + \mu_2 P_{0,1}, \quad n = 0 \tag{3.11}$$

$$n = 1, \begin{cases} (\lambda + \mu_1) P_{1,0} & = \mu_2 P_{0,1} + \lambda \pi_1 P_0 \\ (\lambda + \mu_2) P_{0,1} & = \mu_1 P_{1,1} + \lambda \pi_2 P_0 \end{cases} \tag{3.12}$$

$$[\lambda \nu + \mu_3] P_2 = (\mu_3 + \alpha) P_3 + \lambda P_1, \quad n = 2 \tag{3.13}$$

$$[\lambda\nu^{n-1} + \mu_3 + (n-2)\alpha] P_n = [\mu_3 + (n-1)\alpha] P_{n+1} + \lambda\nu^{n-2}P_{n-1}, \quad n = 3, 4, \dots, N-1 \quad (3.14)$$

$$[\mu_3 + (N-2)\alpha]P_N = \lambda\nu^{N-2}P_{N-1}, \quad n = N \quad (3.15)$$

En résolvant l'équation (3.12) et en utilisant (3.11), on déduit :

$$P_{1,0} = \frac{\lambda[\lambda + \pi_1\mu_3]}{\mu_1[2\lambda + \mu_3]}P_0 \quad (3.16)$$

et

$$P_{0,1} = \frac{\lambda[\lambda + \pi_2\mu_3]}{\mu_2[2\lambda + \mu_3]}P_0 \quad (3.17)$$

Donc :  $P_1 = P_{0,1} + P_{1,0} = \Delta P_0$

$$\text{où } \Delta = \frac{\lambda[\lambda\mu_3 + \mu_3(\pi_1\mu_2 + \mu_1\pi_2)]}{\mu_1\mu_2(2\lambda + \mu_3)} \quad (3.18)$$

En utilisant le processus de naissance et de mort, on obtient :

$$P_{n+1} = \frac{\xi\nu^{n-1}}{\delta + n - 1}, \quad n = 1, 2, \dots, N. \quad (3.19)$$

En mettant  $n = 1, 2, \dots, N-1$  dans l'équation (3.19) et en utilisant la méthode itératif, on obtient :

$$P_n = \frac{(\xi\nu^{\frac{n-2}{2}})^{n-1}}{(\delta)_{n-1}}\Delta P_0, \quad n = 1, 2, \dots, N.$$

Où

$$(\delta)_n = \delta(\delta + 1)\dots(\delta + n - 1), \quad (\delta)_0 = 1, \quad \xi = \frac{\lambda}{\alpha}, \quad \delta = \frac{\mu_3}{\alpha} \text{ et } \nu = e^{-\frac{\gamma}{\mu_3}}.$$

En utilisant la condition de normalisation  $\sum_{n=0}^N P_n = 1$  on déduit :

$$P_0 = \left[ 1 + \Delta \sum_{n=1}^N \frac{(\xi\nu^{\frac{n-2}{2}})^{n-1}}{(\delta)_{n-1}} \right]^{-1}.$$

Nombre moyen de clients dans le système  $L_s$  :

le nombre moyen de clients dans le système est :

$$\begin{aligned} L_s &= \sum_{n=1}^N nP_n \\ &= \Delta \left[ 1 + \sum_{n=2}^N \frac{n(\xi\nu^{\frac{n-2}{2}})^{n-1}}{(\delta)_{n-1}} \right] P_0 \end{aligned}$$

Nombre moyen de clients dans la file  $L_q$  :

le nombre moyen de clients dans la file est :

$$\begin{aligned} L_q &= \sum_{n=3}^N (n-2)P_n \\ &= \Delta \left[ \sum_{n=3}^N \frac{(n-2)(\xi\nu^{\frac{n-2}{2}})^{n-1}}{(\delta)_{n-1}} \right] P_0 \end{aligned}$$

Temps moyen de séjour dans le système  $W_s$  :

et le temps d'attente prévu dans le système :

$$W_s = \frac{L}{\lambda'}$$

Temps moyen de séjour dans la file  $W_q$  :

$$W_q = \frac{L_q}{\lambda'}$$

où

$$\lambda' = \frac{\mu_3}{2}(L - L_q).$$

**Cas spéciaux**

Certains systèmes de files d'attente peuvent être obtenus comme des cas spéciaux de ce modèle.

1. Si  $\mu_1 = \mu_2 = \mu, \pi_1 = \pi_2 = \frac{1}{2}$ , on obtient un modèle de file d'attente avec deux serveurs homogènes :  $M/M/2/N$  avec un dérobade ( $\nu = \beta$  fonction constante) et abandon, les résultats sont :

$$P_1 = \frac{\lambda}{\mu} P_0, \quad \Delta = \frac{\lambda}{\mu} = r,$$

$$P_2 = \frac{\lambda}{2\mu} P_0,$$

$$P_n = \frac{r^2}{2} \left( \rho \beta^{\frac{n-1}{2}} \right)^{n-2} P_0, \quad n = 2, 3, \dots, N.$$

Où  $\rho = \frac{\lambda}{2\mu}$  avec le travail de Habib [13] à  $c = 2$ .

De plus, si  $\alpha = 0$  et  $\beta = 1$  ( $\gamma = 0$ ), on obtient un modèle de file d'attente avec deux serveurs homogènes  $M/M/2/N$  sans dérobage et abandon qui a été étudié par White et al [35], Medhi [26] et Bunday [7].

2. Si on met  $\mu_1 = \mu_2$ ,  $\pi_1 = 1$ ,  $\pi_2 = 0$ ,  $\alpha = 0$ ,  $\nu = \beta$  et  $N \rightarrow \infty$ , on obtient la file d'attente  $M/M/2$  avec dérobage (balking) seulement, qui a été étudié par Singh [32].

#### Exemple numérique

Dans le système ci-dessus, en laissant  $N = 6$ , c'est-à-dire la file d'attente :  $M/M/2/6$ , on obtient :

$$P_0 = \left[ 1 + \Delta \sum_{n=1}^6 \frac{(\xi \nu^{\frac{n-2}{2}})^{n-1}}{(\delta)_{n-1}} \right]^{-1},$$

$$P_n = \frac{(\xi \nu^{\frac{n-2}{2}})^{n-1}}{(\delta)_{n-1}} \Delta P_0, \quad n = 1, 2, \dots, 6,$$

$$L = \left[ 1 + \sum_{n=2}^6 \frac{n(\xi \nu^{\frac{n-2}{2}})^{n-1}}{(\delta)_{n-1}} \right] \Delta P_0,$$

et

$$L_q = \Delta \left[ \sum_{n=3}^6 \frac{(n-2)(\xi \nu^{\frac{n-2}{2}})^{n-1}}{(\delta)_{n-1}} \right] P_0.$$

Maintenant, on calcule les probabilités et quelques mesures de performance pour :  $\lambda = 4$ ,  $\mu_1 = 3$ ,  $\mu_2 = 5$ ,  $\alpha = 0$ ,  $\pi_1 = 0.2$ ,  $\pi_2 = 0.8$ , et  $\gamma = 2$ .

Substituer par ces valeurs dans les résultats ci-dessus, on obtient :

$$P_0 = 0.90596$$

$$P_1 = 0.055868$$

$$P_2 = 0.026291$$

$$P_3 = 0.0091004$$

$$P_4 = 0.0023239$$

$$P_5 = 0.00043961$$

$$P_6 = 0.000061536$$

$$L = 18.771$$

$$L_q = 9.8324$$

$$W = 0.52500$$

et

$$W_q = 0.27500.$$



**Annexes**

Annexe (a) :

Multiplier la première équation et la deuxième équation de les relations (3.12) par  $\mu_1$  et  $-\mu_2$  respectivement, on obtient :

$$\mu_1(\lambda + \mu_1)P_{1,0} = \mu_1\mu_2P_{1,1} + \lambda\mu_1\pi_1P_0 \quad (3.20)$$

$$-\mu_2(\lambda + \mu_2)P_{0,1} = -\mu_1\mu_2P_{1,1} + \lambda\pi_2\mu_2P_0 \quad (3.21)$$

Puis, en ajoutant (3.20) et (3.21), on obtient :

$$\mu_1(\lambda + \mu_1)P_{1,0} - \mu_2(\lambda + \mu_2)P_{0,1} = \lambda[\mu_3\pi_1 - \mu_2]P_0 \quad (3.22)$$

Ensuite, on mettent :

$P_{0,1} = \frac{1}{\mu_2}(\lambda P_0 - \mu_1 P_{1,0})$  dans (3.22) , on trouvent :

$$(2\lambda\mu_1 + \mu_1\mu_3)P_{0,1} = \lambda[\lambda + \mu_3\pi_1]P_0$$

où

$$P_{0,1} = \frac{\lambda[\lambda + \mu_3\pi_1]}{\mu_1[2\lambda + \mu_3]}P_0.$$

Annexe (b) :

De la même façon et en mettent :  $P_{1,0} = \frac{1}{\mu_1}(\lambda P_0 - \mu_2 P_{1,0})$  dans (3.22), on obtient :

$$P_{1,0} = \frac{\lambda[\lambda + \mu_3\pi_2]}{\mu_2[2\lambda + \mu_3]}P_0.$$

Annexe (c) : Des relations (3.11) et (3.12), on observe que :

$$\mu_3P_2 = \lambda P_1.$$

Ainsi, à partir de (3.13) et en utilise le résultat ci-dessus, on obtient :

$$\lambda\nu P_2 = (\mu_3 + \alpha)P_3$$

De même (3.14), on obtient :

$$\lambda\nu^{n-1}P_n = [\mu_3 + (n-1)\alpha]P_{n+1}$$

où

$$[\mu_3 + (n - 2)\alpha]P_n = \lambda\nu^{n-2}P_{n-1}, \quad n = 3, 4, \dots, N - 1.$$

Aussi, à partir des résultats ci-dessus, nous pouvons obtenir :

$$P_{n+1} = \frac{\xi\nu^{n-1}}{\delta + n - 1}P_n, \quad n = 1, 2, \dots, N - 1$$

où :

$$\delta = \frac{\mu_3}{\alpha} \quad \text{et} \quad \xi = \frac{\lambda}{\alpha}.$$

### 3.3.1 Résultats numériques

Pour examiner l'impact des paramètres du modèle sur le comportement des mesures de performances du système, certains calculs numériques sont présentés sous formes des tableaux et des graphiques. La capacité du système est fixé à  $N = 6$ , et la fonction de dérobade et donnée par  $e^{-\frac{\gamma(n-1)}{\mu_3}}$ .

Ici nous avons pris  $\lambda = 15$ ,  $\alpha = 1$ ,  $\mu_1 = 3$ ,  $\mu_2 = 5$ ,  $\pi_1 = 0.2$ ,  $\pi_2 = 0.8$ ,  $\gamma = 2$ , à moins qu'ils ne soient considérés comme des variables ou que leurs valeurs soient mentionnées dans les tableaux et figures respectifs, avec

- $P_i = P_0$  : La probabilité que le serveur est libre,
- et  $P_b = 1 - P_0$  : La probabilité que les serveurs soient occupés.

TABLE 3.1 – Variation des paramètres de performance vs.  $\lambda$

$\lambda$	$P_i$	$P_b$	$L_s$	$L_q$	$W_s$	$W_q$	$R_a$
4	0.47808051	0.52191949	0.74327463	0.04911739	0.26768958	0.01768958	0.04911739
6	0.28685624	0.71314376	1.24710436	0.17779351	0.29156731	0.04156731	0.17779351
8	0.17281204	0.82718796	1.72789969	0.37851082	0.32012634	0.07012634	0.37851082
10	0.15827990	0.84172010	1.87934186	0.48080548	0.26875838	0.06875838	0.48080548
13	0.05141282	0.94858718	2.75816320	1.01934032	0.39655609	0.14655609	1.01934032

Pour ce premier tableau [3.1], on remarque que la croissance de  $\lambda$  (le taux des inter-arrivées) provoque une décroissance de  $P_i$  (la probabilité que le système soit vide), par contre pour les autres paramètres  $P_b$  (c'est la probabilité que les serveurs soient occupés),  $L_s$  (le nombre moyen de clients dans le système),  $L_q$  (le nombre moyen de clients dans la file d'attente),  $W_s$  (Temps moyen dans le système),  $W_q$  (Temps moyen dans la file d'attente),  $R_a$  (le taux d'abandon), on voit une croissance ce qui est logique.

TABLE 3.2 – Variation des paramètres de performance vs.  $\mu_1$ 

$\mu_1$	$P_i$	$P_b$	$L_s$	$L_q$	$W_s$	$W_q$	$R_a$
3	0.02617315	0.97382685	3.24179359	1.39552640	0.43896593	0.18896593	1.39552640
5	0.04611964	0.95388036	2.99497858	1.22557679	0.33853007	0.13853007	1.22557679
7	0.06717486	0.93282514	2.73543404	1.04745443	0.27008956	0.10342289	1.04745443
9	0.08807877	0.91192123	2.49002917	0.88291387	0.22133972	0.07848258	0.88291387
11	0.10787529	0.89212471	2.27063229	0.74016697	0.18545277	0.06045277	0.74016697

TABLE 3.3 – Variation des paramètres de performance vs.  $\mu_2$ 

$\mu_2$	$P_i$	$P_b$	$L_s$	$L_q$	$W_s$	$W_q$	$R_a$
6	0.03346343	0.96653657	3.13046863	1.31709161	0.38362662	0.16140440	1.31709161
8	0.04991268	0.95008732	2.88495010	1.14379605	0.30125788	0.11943970	1.14379605
10	0.06800816	0.93199184	2.63742917	0.97288335	0.24376519	0.08991904	0.97288335
12	0.08674952	0.91325048	2.40679927	0.81885947	0.20208988	0.06875655	0.81885947
14	0.10534117	0.89465883	2.20103004	0.68659198	0.17098402	0.05333696	0.68659198

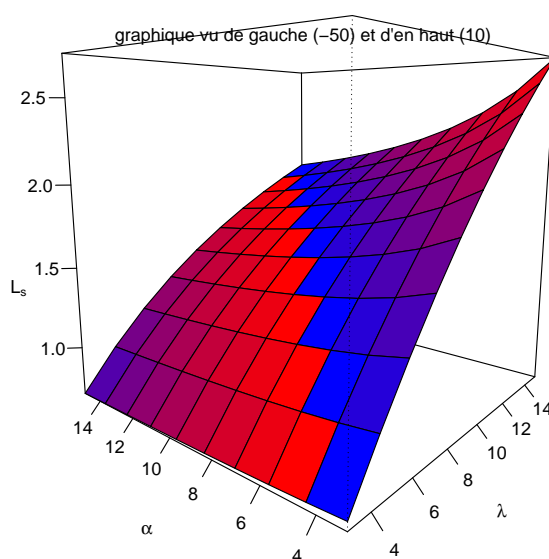
Tableaux [3.2, 3.3] :

Ici la croissance de  $\mu_i$ , ( $i = 1, 2$ ) (le taux de service du  $i^{\text{ème}}$  serveur) a un impact négatif sur les différents paramètres de performance c'est à dire (ils décroissent) à savoir  $P_b$ ,  $L_s$ ,  $L_q$ ,  $W_s$ ,  $W_q$ ,  $R_a$  mais la probabilité pour que le système soit vide ( $P_i$ ) croit cela est interprété par le fait que plus le taux de service est grand plus le système se vide rapidement.

TABLE 3.4 – Variation des paramètres de performance vs.  $\alpha$ 

$\alpha$	$P_i$	$P_b$	$L_s$	$L_q$	$W_s$	$W_q$	$R_a$
1	0.02617315	0.97382685	3.24179359	1.39552640	0.43896593	0.18896593	1.39552640
2.5	0.03418728	0.96581272	2.83565309	1.03645838	0.39401698	0.14401698	2.59114594
4	0.04003207	0.95996793	2.59332584	0.82846161	0.36735486	0.11735486	3.31384643
5.5	0.04451982	0.95548018	2.43103841	0.69253379	0.34958757	0.09958757	3.80893582

Pour le tableau [3.4] lorsque le taux d'abandon ( $\alpha$ ) augmente la probabilité  $P_i$  pour que le système soit vide et le taux d'abandon  $R_a$  augmentent et les autres paramètres décroissent.

FIGURE 3.2 –  $L_s$  en fonction de  $\lambda$  pour différents valeurs de  $\alpha$  (En trois dimensions)

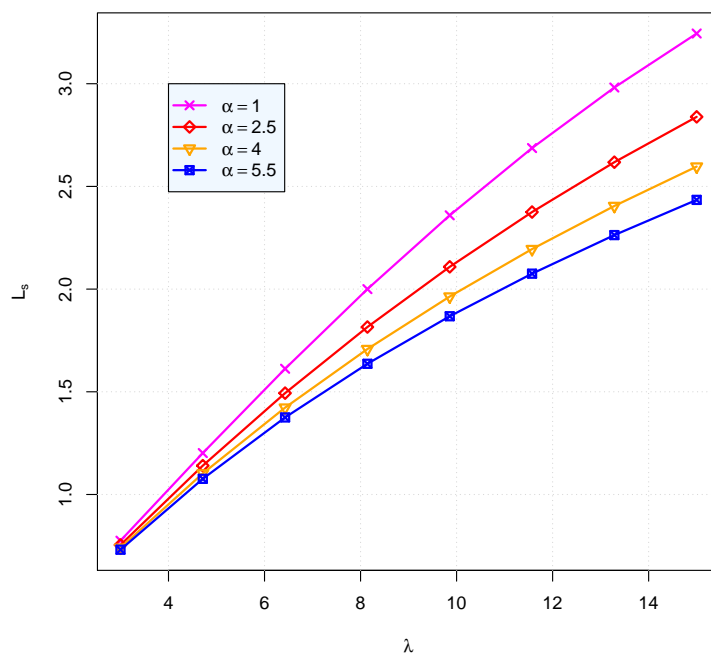


FIGURE 3.3 –  $L_s$  en fonction de  $\lambda$  pour différents valeurs de  $\alpha$  (En deux dimensions)

Figures [3.2, 3.3] :

Par la croissance de  $\lambda$ ,  $L_s$  (le nombre moyen de clients dans le système) croît aussi. Cette croissance de  $L_s$  est plus importante dans le cas de  $\alpha = 1$  que dans le cas  $\alpha = 5.5$ . Cela est interprété par le fait que  $\alpha$  représente le taux d'abandon.

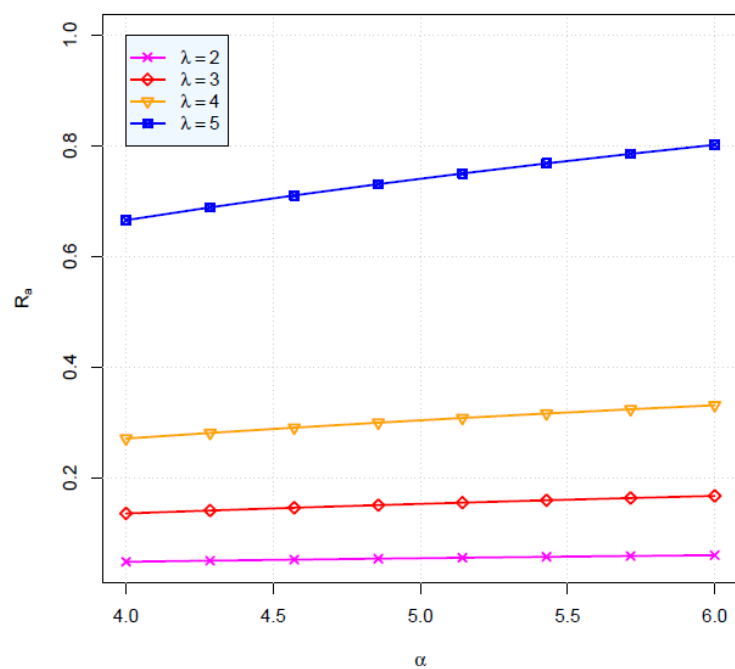
FIGURE 3.4 –  $R_a$  en fonction de  $\alpha$  pour différents valeurs de  $\lambda$ 

Figure [3.4] :

Le taux d'abandon  $R_a$  croît par la croissance de  $\alpha$  (le paramètre de l'abandon).

De plus cette croissance est plus importante dans le cas  $\lambda = 2$  ce qui est logique.

# Conclusion générale

Dans ce mémoire, nous analysons le modèle :  $M/M/2/N$  avec fonction de dérobade, abandon et serveurs hétérogènes. Les expressions explicites pour les probabilités d'état stable sont dérivées. Quelques mesures de performance du système et certains cas spéciaux sont obtenus. Ainsi, les valeurs numériques de nos résultats analytiques sont également présentées.

La généralisation du modèle à  $C$  serveurs hétérogènes peut être traitée. L'analyse économique ainsi que l'optimisation du différents paramètres du modèle pour le mieux gérer peuvent être le but d'un travail.



# 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] Ammar SI, Transient analysis of a two-heterogeneous servers queue with impatient behavior, *Journal of the Egyptian Mathematical Society*, 22 (1), 90-95, 2014.
- [3] Ancker, Jr., C. J. and Gafarian, A. V. Some queuing problems with balking and reneging : I. *Operations Research*. 11, No. 1, 88-100. (1963).
- [4] Ancker, Jr., C. J. and Gafarian, A. V. Some queuing problems with balking and reneging : II. *Operations Research*. 11, No. 6, 928-937. (1963).
- [5] Bouchentouf A. A., Kadi M., Rabhi A. 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)*.
- [6] Bunday, B. D. *Basic Queueing Theory*. Edward Arnold, London. (1986).
- [7] Bunday, B.D, "An Introduction to Queueing theory", John Wiley, New York, 1996.
- [8] Chaudhry, M. L., and Templeton, J. G. C. *A First Course in Bulk Queues*. John Wiley and Sons, New York. (1983).
- [9] A. Choudhury and P. Medhi, Balking and reneging in multiserver Markovian queuing system. *Int. J. Math. Oper. Res.* 3 (2011) 377-394.
- [10] 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.
- [11] Fedosejev, Yu. N. *Methods of Queueing Systems Analysis*. Moscow Institute of Engineering and Physics (in Russian). (1982).
- [12] O. Garnett, A. Mandelbaum, and M. Reiman. Designing a call center with impatient customers. *Manufacturing & Service Operations Management*, 4(3 : 208-227, (June 2002).

- [13] Habib, M.E., "The M/M/C/N Queueing system with balking", Academy of Business Journal, AL-Azhar University, vol.1, pp.1-10, 2007.
- [14] Haight, F. A., Queuing with balking, I, Biometrika, 44, 360-369. (1957).
- [15] Haight, F. A., Queueing with reneging. Metrika 186-197. 2 (1959).
- [16] Jean Louis Poss, Probabilité et statistique version 2.1. p74. Mai (2003).
- [17] 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).
- [18] 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).
- [19] 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).
- [20] Krishna Kumar B., S. Pavai Madheswari, and K. S Venkatakrisnan " Transient Solution of an M/M/2 Queue with Heterogeneous Servers Subject to Catastrophes", Information and Management Sciences, vol. 18, no. 1, pp. 63-80,2007.
- [21] Krishnamoorthy A. and C. Sreenivasan, An M/M/2 Queueing System with Heterogeneous Servers Including One with Working Vacation, International Journal of Stochastic Analysis Volume 2012, Article ID 145867, 16 pages.
- [22] Kumar, R. and Sharma, S.K. *M/M/1/N* Queueing System with Retention of Reneged Customers, Pakistan Journal of Statistics and Operation Research. 859-866. 8(2012).
- [23] Kumar, R. and Sharma, S.K. An *M/M/1/N* Queueing Model with Retention of reneged customers and Balking, American Journal of Operational Research. No. 1, 1-5. 2(2012).
- [24] Liqiang Liu. Service Systems with Balking Based on Queueing Time. Chapel Hill Under the direction of Dr. Vidyadhar G. Kulkarni. (2007).
- [25] 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).
- [26] Medhi, J., "Stochastic Models in Queueing Theory", Academic Press, New York, 1991.
- [27] Morse, M., "Queues, Inventory and Maintenance", John Wiley, New York, 1957.

- [28] Saaty, T. L. Elements of queueing theory with applications. New York : McGraw Hill. (1961).
- [29] Santhakumaran, A. and Thangaraj, V. A Single Server Queue with Impatient and Feedback Customers, Information and Management Science. 11, No. 3, 71-79. (2000).
- [30] 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).
- [31] Shuangchi He and J. G. Dai. Many-server queues with customer abandonment : numerical analysis of their diffusion models. arXiv preprint arXiv :1104.0347, (2011).
- [32] Singh, V.P., " Two-server Markovian Queues with Balking : Heterogeneous VS. Homogeneous Servers ", Oper. Res., No. 18, pp. 145-159. 1970.
- [33] Van Doorn, E. Stochastic Monotonicity and Queueing Applications of Birth-Death Processes. Lecture Notes in Statistics 4, Springer-Verlag, New York, (1981).
- [34] 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).
- [35] White, J.A., J.w. Schmidt and G.K. Bennett, " Analysis of Queueing System", Academic press, New York. 1975.
- [36] Yves Caumel. Probabilité et Processus Stochastiques. Springer-Verlag France, (2011).