



N° Attribué par la bibliothèque

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



Année univ.: 2017/2018

# Systemes De File D'Attente Avec Découragement

Mémoire présenté en vue de l'obtention du diplôme de

Master Académique

Université de Saïda - Dr Moulay Tahar

Discipline : MATHEMATIQUES

Spécialité : Analyse stochastique, statistique des processus et  
applications (ASSPA)

par

**Hammache Saouaguia<sup>1</sup>**

Sous la direction de

**Dr Mokhtar Kadi**

Soutenu le 25/06/2018 devant le jury composé de

<b>Dr S. Rahmani</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>Dr F. Benziadi</b>	Université Dr Tahar Moulay - Saïda	Examinatrice

---

1. e-mail : kanzahemache@gmail.com

# *Remerciements*

*Par ces quelques lignes, je tiens à remercier toutes les personnes qui ont participé de près ou de loin au bon déroulement de ce mémoire, en espérant n'avoir oublié personne ...*

*Je tiens d'exprimer toute ma reconnaissance à mon encadreur de mémoire et mon professeur Monsieur : **Kadi Mokhtar** de m'avoir encadré orienté aidé et conseillé, et ainsi pour la confiance et la liberté qu'il m'a accordé, sans oublier ses efforts fournis pendant cette période.*

*Mes vifs remerciements vont aux **Mr. A.Guendouzi** et **Mr. L.Yahyaoui** pour l'aide considérable qui m'ont donné à réaliser la partie simulation.*

*Mes sincères remerciements et ma gratitude vont aussi à **Dr. S.Rahmani** pour avoir accepté de juger ce travail et d'en présider le jury de soutenance. Que vous soyez assuré de mon entière reconnaissance.*

*Je souhaite également remercier **Dr. F.Benziadi** et **Pr. A.Kandouci** de m'avoir fait l'honneur d'accepter d'être examinateurs de mon mémoire. Leurs remarques et précieux conseils ont abouti à une amélioration de ce document.*

*J'adresse mes sincères remerciements à tous Les membres du laboratoire des Modèles Stochastiques, Statistique et Applications et à la tête le directeur **Pr. A.Kandouci**.*

*Je remercie également tous les enseignants de départements de Mathématiques.*

*Je remercie mes très chers parents, qui ont toujours été là pour moi.*

*Enfin je remercie mes collègues étudiantes et étudiants de Mathématiques pour les bons moments passés ensemble et aussi pour leur soutiens et leur encouragements. Et pour finir, merci à toutes les personnes que j'ai oublié de citer et qui m'ont permis de mener à bien ce mémoire.*

*«Merci à vous tous»*

## *dédicaces*

*Je dédie ce travail à :*

*Mes chère parents que dieu me les garde et qui m'ont toujours encouragé et m'aidé à terminer mes études.*

*A mes chères sœurs et mes chers frères.*

*A toute les familles : Hammache, Lechelag, Taibi, Lazragi et sadouki.*

*Une spéciale dédicace à mon adorable amie M.Zahira qui m'a bien aidé le long de mon parcours universitaire et bien sur sans oublier mon adorable amie Houda.*

*A tous mes collègues de la promotion de Master A.S.S.P.A 2018.*

*A toutes les personnes qui m'aiment.*

# Table des matières

<b>1</b>	<b>Les processus stochastiques</b>	<b>6</b>
1.1	Processus de comptage . . . . .	7
1.2	Rappels : loi de Poisson et loi exponentielle . . . . .	8
1.3	Processus de Poisson . . . . .	10
1.4	Processus de naissance et de mort . . . . .	13
1.4.1	Processus de naissance . . . . .	13
1.4.2	Processus de mort . . . . .	14
<b>2</b>	<b>Les systèmes de files d'attente classiques</b>	<b>16</b>
2.1	File d'attente simple . . . . .	17
2.2	Notation de Kendall . . . . .	20
2.3	Loi de Little . . . . .	22
2.4	Quelque modèles de files d'attente . . . . .	23
2.4.1	Modèle d'attente $M/M/1$ . . . . .	23
2.4.2	Modèle d'attente $M/M/1/N$ . . . . .	27
2.4.3	Modèle d'attente $M/M/c$ . . . . .	30
2.4.4	Modèle d'attente $M/M/\infty$ . . . . .	33
<b>3</b>	<b>Les systèmes de file d'attente avec découragement</b>	<b>35</b>
3.1	Le modèle d'attente $M/M/1/N$ avec découragement . . . . .	35
3.2	Résultats numériques . . . . .	41
	<b>Bibliographie</b>	<b>48</b>

# Introduction générale

La théorie des files d'attente joue un rôle important dans la modélisation des problèmes de la vie réelle. Elles ont été intégrées dans la modélisation de divers domaines d'activité à l'évaluation des performances des systèmes informatiques, aux réseaux de communication et aux secteurs de la santé et sciences médicales etc. Grâce à tous ces développements, la théorie des files d'attente est aujourd'hui largement utilisée et ses applications sont multiples.

Dans ce mémoire on s'intéresse aux systèmes de files d'attente avec découragement. 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 phénomène de l'impatience des clients a un impact très négatif sur les systèmes de files d'attente, il est parmi les causes principales des pertes économiques considérables des entreprises du secteur de gouvernement et privé.

Les systèmes de files d'attente avec abandon ont été traités pour la première fois par Haight [17]. Ancker et Gafarian [3] ont étudié un modèle de files d'attente  $M/M/1/N$  avec dérobade et abandon et effectué son analyse à l'état stationnaire. Ancker et Gafarian [4] ont aussi obtenu des résultats pour un système avec dérobade pur (aucun abandon). Ces systèmes de files d'attente avec multiserveurs et clients impatientes ont une large application dans de nombreuses situations de la vie réelle comme dans les hôpitaux, informatique-communication, magasins de détail etc. Xiong et Altioek [44] ont traité des files d'attente de plusieurs serveurs avec abandon déterministe. Wang et al. [43] ont présenté une bibliographie générale des systèmes de files d'attente avec clients impatientes. Une file d'attente markovienne de serveur unique avec des clients impatientes et situations où les clients abandonnent le système en même temps a fait l'objet d'un travail de Kapodistria [19]. Il considère deux scénarios d'abandon. Dans le premier, tous les clients présents sont impatientes et effectuent des abandons synchronisés; alors que dans le second scénario, le client de service est exclu de la procédure d'abandon. Il étend également cette analyse à la file d'attente  $M/M/c$  sous le deuxième scénario de l'abandon. Kumar Sharma [25] ont appliqué le modèle de files d'attente  $M/M/1/N$  pour la modélisation des

situations avec des clients impatientes.

Les modèles de files d'attente où les clients potentiels sont découragés par la longueur de la file d'attente sont étudiés par de nombreux chercheurs dans leurs travaux de recherche. Natvig [33] a étudié le processus de naissance-mort dans une files d'attente avec un seul serveur et l'état du système dépendant des paramètres  $\lambda_n = \frac{1}{n+1}\lambda$ ,  $n \geq 0$  et  $\mu_n = \mu$ ,  $n \geq 1$ . Raynolds [37] a présenté un modèle de files d'attente avec découragement et multiserveurs. Il obtient la distribution d'équilibre de la longueur de la file d'attente et les mesures de performance. Ammar et al. [2] ont traité un modèle de files d'attente avec un seul serveur, capacité finie et découragement en utilisant la méthode matricielle.

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 Poisson et processus de naissance et de mort...etc).

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

Et enfin le troisième chapitre représente une étude du modèle  $M/M/1/N$  de file d'attente avec découragement. Ensuite nous présentons quelques exemples numériques pour voir comment les différents paramètres du modèle influencent sur le comportement du système.

# Chapitre 1

## Les processus stochastiques

L'étude des processus stochastiques s'insère dans la théorie des probabilités dont elle constitue l'un des objectifs les plus profonds. Elle soulève des problèmes mathématiques intéressants et souvent très difficiles.

Par exemple, le prix d'un baril du Pétrole, il a connu au cours de ces cinq dernières années des fluctuations, qui ont tiré l'attention de beaucoup des spécialistes économiques. En effet, Ce prix, dans la bourse, varie tout le temps, cette variation nous donne l'idée d'établir un processus aléatoire, ou encore un processus stochastique, d'où la modélisation par une famille de variables aléatoires  $X(t)_{t \in T}$  où  $T$  est l'ensemble des temps pendant lesquels le phénomène est observé. La famille  $X(t)_{t \in T}$  est appelée processus aléatoire, ou encore processus stochastique.

**Définition 1.0.1.** *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  $t$  [46].*

- *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.*

## 1.1 Processus de comptage

### Définition 1.1.1. (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;
- 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'interarrivées  $\{T_n, n \in \mathbb{N}_0\}$  ou  $\forall n \in \mathbb{N}_0$  la variable aléatoire  $T_n$  est le temps d'attente entre les  $(n-1)$ <sup>ième</sup> et  $n$ <sup>ième</sup> occurrences [45], c-à-d :

$$T_n = A_n - A_{n-1}$$

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

**Proposition 1.1.1.** Les relations suivantes sont triviales à vérifier tel que  $A_0 = 0$  :

1.  $A_n = T_1 + T_2 + \dots + T_n \quad \forall n \geq 1$ ;
2.  $N(t) = \sup\{n \geq 0 : A_n \leq t\}$ ;
3.  $\mathbb{P}[N(t) = n] = \mathbb{P}[A_n \leq t < A_{n+1}]$ ;
4.  $\mathbb{P}[N(t) \geq n] = \mathbb{P}[A_n \leq t]$ ;
5.  $\mathbb{P}[s < A_n < t] = \mathbb{P}[N_s < n \leq N(t)]$ .

**démonstration de (1) :** on a  $T_n = A_n - A_{n-1}$

$$\begin{aligned} T_1 + T_2 + \dots + T_n &= A_1 - A_0 + A_2 - A_1 + A_3 - A_2 + \dots + A_{n-1} - A_{n-2} + A_n - A_{n-1} \\ &= -A_0 + A_n \\ &= A_n. \quad \text{car } A_0 = 0 \end{aligned}$$

### Définition 1.1.2. (processus de comptage)

Un processus de comptage  $N(t)_{t \in \mathbb{R}^+}$  est un processus croissant, si pour tout  $s \leq t$ ,  $N_s \leq N_t$ .

La variable aléatoire  $N_t - N_s$  est alors appelée accroissement du processus sur  $]s, t]$ .

Par exemple :

- $N(t)$  = nombre de poissons capturés dans l'intervalle de temps  $[0, t]$ ;
- $N(t)$  = taille d'une population à la date  $t$ .

**Définition 1.1.3. (processus à accroissements indépendants)**

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

**Définition 1.1.4. (processus à accroissements stationnaire)**

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

## 1.2 Rappels : loi de Poisson et loi exponentielle

**Définition 1.2.1.** Une variable aléatoire  $X$  à valeurs entières suit une loi de Poisson de paramètre  $\lambda > 0$  si :

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

**Définition 1.2.2.** Une variable aléatoire  $Y$  à valeurs réelles strictement positives suit une loi exponentielle de paramètre  $\mu > 0$  si :

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

**Distribution de Poisson :**

Soit  $n$  une variable aléatoire discrète avec  $n = 0, 1, \dots$  qui suit une distribution de Poisson.

La distribution de probabilité de  $n$  est  $P_n = \frac{\lambda^n}{n!} e^{-\lambda}$ .

L'espérance et la variance de  $n$  sont :

$$\mathbb{E}(n) = \lambda \quad \text{et} \quad \text{Var}(n) = \lambda, \quad \text{respectivement.}$$

La distribution de Poisson peut également être définie en unités de temps  $t$ . Dans ce cas, la variable discrète  $n$  représente le nombre d'occurrences dans le temps  $t$  devient,

$$P(n, t) = \frac{(\lambda t)^n}{n!} e^{-(\lambda t)}$$

**Distribution exponentielle :**

Soit  $t$  une variable aléatoire continue avec  $t \geq 0$  qui suit une distribution exponentielle.

La densité de probabilité de  $t$  est  $f(t) = \mu e^{-(\mu t)}$  et la distribution cumulée correspondante est  $F(t) = 1 - e^{-(\mu t)}$ .

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

$$\mathbb{E}(t) = \frac{1}{\mu} \quad \text{et} \quad \text{Var}(t) = \frac{1}{\mu^2}, \quad \text{respectivement.}$$

**Relation entre la distribution exponentielle et la distribution de Poisson :**

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

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

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

$$P(0, t) = e^{-(\mu t)}$$

$$P(1, t) = \int_{\tau=0}^t P(0, \tau) f(1 - \tau) d\tau = \mu t e^{-(\mu t)}$$

$$P(2, t) = \int_{\tau=0}^t P(1, \tau) f(1 - \tau) d\tau = (\mu t)^2 e^{-(\mu t)} / 2!$$

...

$$P(n, t) = \int_{\tau=0}^t P(n-1, \tau) f(1 - \tau) d\tau = (\mu t)^n e^{-(\mu t)} / n!$$

**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  $h$  deviennent

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

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.

### 1.3 Processus de Poisson

Nous allons à présent introduire un processus de nature différente, dont le domaine d'applicabilité est très important : le processus de Poisson. Dans le cadre qui va nous intéresser ici, celui-ci décrit la répartition aléatoire et uniforme de points sur la droite réelle positive.

Il peut servir à modéliser par exemple :

- les appels téléphoniques arrivant dans une centrale.
- les temps d'arrivée de clients à une caisse.
- les temps d'occurrence de sinistres à dédommager par une compagnie d'assurance...etc.

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.

#### Définition 1.3.1. (processus de Poisson)

Soit  $N_t$  le nombre d'occurrences d'un événement aléatoire dans l'intervalle de temps  $(0, t]$ ,  $t > 0$  et  $N_0 = 0$ . Si  $N_t$  satisfait aux deux conditions suivantes, nous appellerons  $N$  un processus de Poisson d'intensité (ou taux)  $\lambda > 0$ ,

- Quelque soient les deux suites  $(s_i)$  et  $(t_i)$ , avec  $0 \leq s_1 \leq t_1 \leq s_2 \leq t_2 \leq \dots \leq s_n \leq t_n < +\infty$ , les variables aléatoires  $N_{t_1} - N_{s_1}, N_{t_2} - N_{s_2}, \dots, N_{t_n} - N_{s_n}$  sont indépendantes.
- $\forall t > 0$  on a :

$$\mathbb{P}(N_{t+h} - N_t = k) = \begin{cases} \lambda(h) + o(h) & \text{si } k = 1 \\ o(h) & \text{si } k \geq 2 \\ 1 - \lambda(h) + o(h) & \text{si } k = 0 \end{cases}$$

$$\text{avec } \lim_{h \rightarrow 0} \frac{o(h)}{h} = 0$$

**Définition 1.3.2. (processus de Poisson)**

Soit  $\lambda > 0$  et  $(S_n)_{n \geq 1}$  une suite de variables aléatoires indépendantes de même loi exponentielle  $\varepsilon(\lambda)$ . On pose  $A_n = S_1 + \dots + S_n$ . On définit alors le processus de comptage  $N = (N_t)_{t \geq 0}$  à valeurs dans  $\mathbb{N} \cup \{+\infty\}$  par

$$N_t = \sum_{n \geq 1} \mathbb{1}_{\{A_n \leq t\}}$$

Ce processus s'appelle le processus de Poisson d'intensité  $\lambda$ .

**Remarque 1.3.1.** Le processus se réécrit aussi sous la forme  $N_t = \sup(n \geq 0 : A_n \leq t)$ .

Réciproquement, on remarque que  $A_n = \inf(t \geq 0 : N_t = n)$ .

Comme pour  $t > s$ , on a  $N_t - N_s = \sum_{n \geq 1} \mathbb{1}_{\{s < A_n \leq t\}}$ ,  $N$  est un processus à accroissements indépendants et stationnaires.

**Définition 1.3.3. (équivalence du processus de Poisson)**

Un processus de Poisson  $N = (N_t)_{t \geq 0}$  d'intensité  $\lambda$  est un processus de comptage à trajectoires continues à droite [39] 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  $\mathcal{P}(\lambda t)$ .

**Caractérisation d'un processus de Poisson par ses temps d'arrivée :**

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

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

**Théorème 1.3.1.**  $(N_t)_{t \in \mathbb{R}_+}$  est un processus de Poisson de paramètre  $\lambda$  si et seulement si les variables aléatoires  $T_n$  sont indépendantes de même loi exponentielle  $\varepsilon(\lambda)$  de densité

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

**Preuve :**

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

Où  $F_1$  est la fonction de répartition de  $T_1$ . On a donc bien  $T_1$  qui suit la loi exponentielle  $\varepsilon(\lambda)$ .

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

D'après l'indépendance des accroissements.

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

Donc  $T_2$  est bien indépendante de  $T_1$  et de même loi exponentielle  $\varepsilon(\lambda)$ .

De façon plus générale,

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

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

La réciproque sera admise.

**Propriété 1.3.1.** *Si  $(N_t)_{t \in \mathbb{R}_+}$  est un processus de Poisson de paramètre  $\lambda$ , le temps aléatoire  $U$  qui sépare un instant  $\theta$  du prochain évènement et le temps aléatoire  $V$  qui sépare  $\theta$  du dernier évènement suivent la loi exponentielle  $\varepsilon(\lambda)$ .*

**Preuve :**

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

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

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

Car  $[V > x]$  signifie que pendant la durée  $x$  qui précède  $\theta$ , il n'y a eu aucune arrivée.

Remarque : On a alors  $\mathbb{E}(U + V) = \mathbb{E}(U) + \mathbb{E}(V) = \frac{2}{\lambda}$  alors que  $\mathbb{E}(T_n) = \frac{1}{\lambda}$  pour tout  $n \in \mathbb{N}^*$ .

C'est donc que  $U + V = T_{N_\theta}$  n'a pas même loi que les  $T_n$  alors que sur  $[N_\theta = n]$ , on a  $T_{N_\theta} = T_n$ .

On peut terminer ce paragraphe en remarquant que :

$$\mathbb{E}(N_1) = \lambda \quad \text{et} \quad \mathbb{E}(T_n) = \frac{1}{\lambda}$$

Ainsi, plus  $\lambda$  est grand, plus le nombre moyen d'arrivées par unité de temps est important, et plus l'intervalle entre 2 arrivées est court, ce qui semblait a priori évident. Pour cette raison, on appelle également le paramètre  $\lambda$  l'intensité du processus.

## 1.4 Processus de naissance et de mort

Un processus est une collection de variables aléatoires  $\{Z_t, t \geq 0\}$ , indexée par le temps. Ici, il nous servira à décrire l'évolution aléatoire au cours du temps d'un nombre d'individus, dans une population ou un système d'attente.

Les variables aléatoires  $Z_t$  prennent donc leurs valeurs dans l'ensemble des entiers  $\mathbb{N}$ . Le processus évolue comme un processus markovien de saut : le nombre d'individus reste constant pendant une certaine durée exponentielle, puis il saute vers une autre valeur. S'agissant d'une population ou d'une file d'attente, nous ne considérerons que des sauts vers les deux valeurs voisines : la taille de la population peut soit augmenter de 1 (naissance ou arrivée) soit diminuer de 1 (mort ou départ). L'intensité de ces sauts est gouvernée par deux suites de réels positifs  $(\lambda_n)_{n \in \mathbb{N}}$  (taux de naissance) et  $(\mu_n)_{n \in \mathbb{N}^*}$  (taux de mort). Pour éviter les cas particuliers, nous supposerons dans un premier temps que ces taux sont tous strictement positifs.

### 1.4.1 Processus de naissance

**Définition 1.4.1.** Soit  $(\lambda_n)_{n \in \mathbb{N}}$  une suite de réels strictement positifs. Un processus de naissance de taux de naissance  $(\lambda_n)$  est un processus markovien de saut  $\{Z_t, t \geq 0\}$  à valeurs dans  $\mathbb{N}$  tel que pour tout  $n \geq 0$  le taux de transition de  $n$  vers  $n + 1$  est  $\lambda_n$ .

- Le processus de naissance est la généralisation direct d'un processus de Poisson lorsque le paramètre d'intensité  $\lambda$  dépend de l'état courant du processus, il va nous permettre d'introduire le concept "d'explosion".

### 1.4.2 Processus de mort

**Définition 1.4.2.** Soit  $(\mu_n)_{n \in \mathbb{N}^*}$  une suite de réels strictement positifs. Un processus de mort de taux de mort  $(\mu_n)$  est un processus markovien de saut  $\{Z_t, t \geq 0\}$  à valeurs dans  $\mathbb{N}$  tel que pour tout  $n \geq 1$  le taux de transition de  $n$  vers  $n - 1$  est  $\mu_n$ .

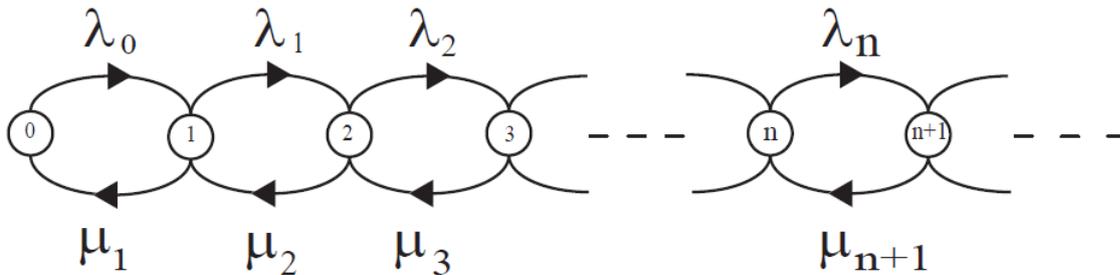


FIGURE 1.1 – Graphe de transition d'un processus de naissance et de mort.

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  $\lambda_n$  sont nuls, on parle de processus de mort.
- Si tous les  $\mu_n$  sont nuls, on parle de processus de naissance.

Pour comprendre la dynamique d'un processus de naissance et de mort, on peut revenir à la construction d'un processus de saut par sa chaîne incluse. Ici c'est une chaîne de Markov à valeurs dans  $\mathbb{N}$ , qui saute de  $n$  vers  $n + 1$  avec probabilité  $\frac{\lambda_n}{\lambda_n + \mu_n}$  et de  $n$  vers  $n - 1$  avec probabilité  $\frac{\mu_n}{\lambda_n + \mu_n}$ . Le passage de la chaîne au processus se fait en rajoutant des temps de séjour aléatoires, indépendants entre eux et indépendants de la chaîne, dont la loi dépend de l'état courant : le temps de séjour dans l'état  $n$  suit la loi exponentielle  $\varepsilon(\lambda_n + \mu_n)$ .

#### Hypothèses :

- Le temps s'écoulant entre deux arrivées consécutives est distribué exponentiellement.
- Le temps de service est aussi distribué exponentiellement.

Sous les hypothèses précédentes, une file d'attente peut être vu comme un processus de naissance et de mort :

- naissance  $\longleftrightarrow$  arrivée du client.
- mort  $\longleftrightarrow$  départ du client du système après son service.

Dans le processus de naissance et de mort :

Hyp.1 : Naissance  $\longleftrightarrow$  Le temps s'écoulant entre deux naissances consécutives est distribué exponentiellement.

Hyp.2 : Mort  $\longleftrightarrow$  Le temps s'écoulant entre deux morts consécutives est aussi distribué exponentiellement.

Hyp.3 : Chaque transition à partir de l'état  $n$  est de type  $n \rightarrow (n+1)$  (une seule naissance) ou  $n \rightarrow (n-1)$  (une seule mort).

## Chapitre 2

# Les systèmes de files d'attente classiques

La théorie des files d'attente fournit un outil très puissant et efficace pour la modélisation des systèmes admettant un phénomène d'attente. Elle peut s'appliquer à différentes situations : gestion des avions au décollage ou à l'atterrissage, attente des clients et des administrés aux guichets, ou bien encore stockage des programmes informatiques avant leur traitement, ateliers de réparation...etc. Ce domaine de recherches, né en 1917, des travaux de l'ingénieur danois Erlang sur la gestion des réseaux téléphoniques de Copenhague entre 1909 et 1920, étudie notamment les systèmes d'arrivée dans une queue, les différentes priorités de chaque nouvel arrivant, ainsi que la modélisation statistique des temps d'exécution. C'est grâce aux apports des mathématiciens Khintchine, Palm, Kendall, Pollaczek et Kolmogorov que la théorie s'est vraiment développée.

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) de savoir par exemple si le nombre de serveurs dans le système est adéquat pour gérer le flux de demandes ou encore d'appréhender les effets d'une modification des conditions de fonctionnement, et ainsi prendre des décisions sur le nombre minimum de ressources nécessaires.

Dans l'étude de systèmes de files d'attente, 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.

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 quelques modèles de files d'attente markoviennes ( $M/M/1$ ,  $M/M/1/N$ ,  $M/M/c$ ,  $M/M/\infty$ ) et l'évaluation des paramètres de performance.

**Classification des files d'attente :** [34]

Pour décrire une file d'attente, on doit donc se donner les éléments suivants :

1. La nature du processus des arrivées qui est définie par la distribution des intervalles séparant deux arrivées consécutives.
2. La distribution du temps aléatoire de service.
3. Le nombre  $c$  des stations de service.
4. La capacité  $N$  du système. Si  $N < \infty$ , la file ne peut dépasser une longueur de  $N - s$  unités. Dans ce cas, certains clients qui arrivent vers le système n'ont pas la possibilité d'y entrer.

**Terminologie et notations :** [5]

En lien avec la loi exponentielle :

- $\lambda$  : Le taux d'arrivée; le nombre moyen d'arrivées par unité de temps.
- $\frac{1}{\lambda}$  : L'intervalle de temps moyen séparant deux arrivées consécutives.
- $\mu$  : Le taux de service; le nombre moyen de clients servis par unité de temps.
- $\frac{1}{\mu}$  : Temps moyen de service d'un client dans le système.

L'analyse d'un système de file d'attente dépend de l'état initial et du temps écoulé. C'est la situation transitoire où l'étude est très complexe. Dans la théorie des files d'attente l'étude se fait une fois que le système atteint sa situation d'équilibre; où les états du système sont essentiellement indépendants de l'état initial et du temps déjà écoulé. On suppose que le système est en opération depuis un très long moment.

**En situation d'équilibre on note :** [32]

- $\pi_n$  : Probabilité qu'il y ait  $n$  clients dans le système.
- $L_s$  : Le nombre moyen de clients dans le système.
- $L_q$  : Le nombre moyen de clients dans la file d'attente.
- $W_s$  : La durée de séjour moyenne dans le système (attente + service).
- $W_q$  : La durée d'attente moyenne d'un client dans la file d'attente.
- $c$  : Le nombre de serveurs.

## 2.1 File d'attente 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.

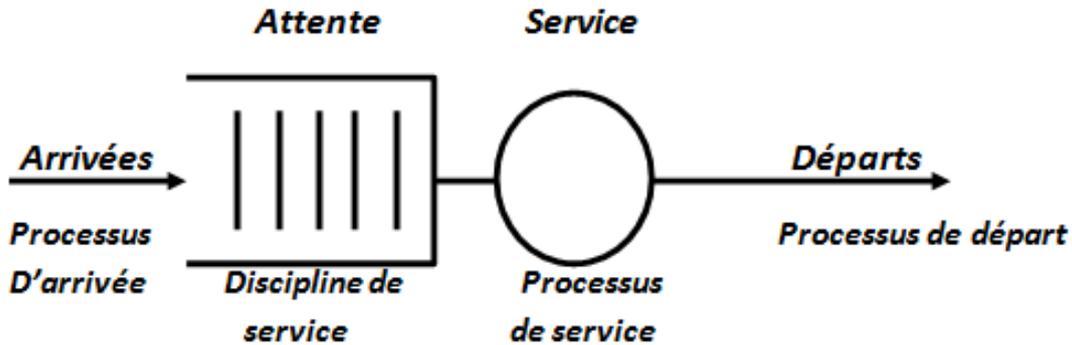


FIGURE 2.1 – Représentation schématique d'une file d'attente simple.

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

**Le processus d'arrivée :**

L'arrivée des clients à le système sera décrite à l'aide d'un processus stochastique de comptage  $(N_t)_{t \geq 0}$ .

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

Si  $T_n$  désigne la variable aléatoire mesurant le temps séparant l'arrivée du  $(n - 1)^{\text{ième}}$  client et du  $n^{\text{ième}}$  client [22], on a alors :

$$T_n = A_n - A_{n-1}$$

**Le processus de service :**

Considérons tout d'abord une file à serveur unique.

On note  $D_n$  la variable aléatoire mesurant l'instant de départ du  $n^{\text{ième}}$  client du système et  $Y_n$  la variable aléatoire mesurant le temps de service du  $n^{\text{ième}}$  client (le temps séparant le début et la fin du service). Un instant de départ correspond toujours à une fin de service, mais ne correspond pas forcément à un début de service [36]. Il se peut en effet qu'un client qui quitte la station laisse celle-ci vide. Le serveur est alors inoccupé jusqu'à l'arrivée du prochain client.

On note  $\mu$  le taux de service :

$$\frac{1}{\mu} \text{ est la durée moyenne de service.}$$

**Structure de la file :**Nombre de serveurs :

Une station peut disposer de plusieurs serveurs en parallèle. Soit  $c$  le nombre de serveurs. Dès qu'un client arrive à la station, soit il y a un serveur libre, le client entre instantanément en service, soit tous les serveurs sont occupés et le client se place dans la file en attente de libération d'un des serveurs. Mais on suppose à la plupart du temps que les serveurs sont identiques et indépendants les uns des autres. Une station particulière est la station IS (infini serveurs) dans laquelle le nombre de serveurs est infini. Cette station ne comporte donc pas de file d'attente.

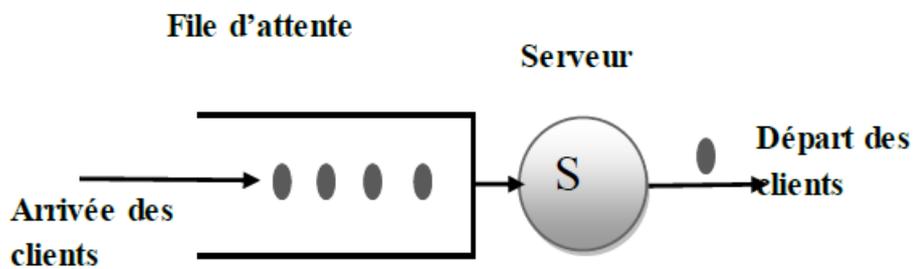
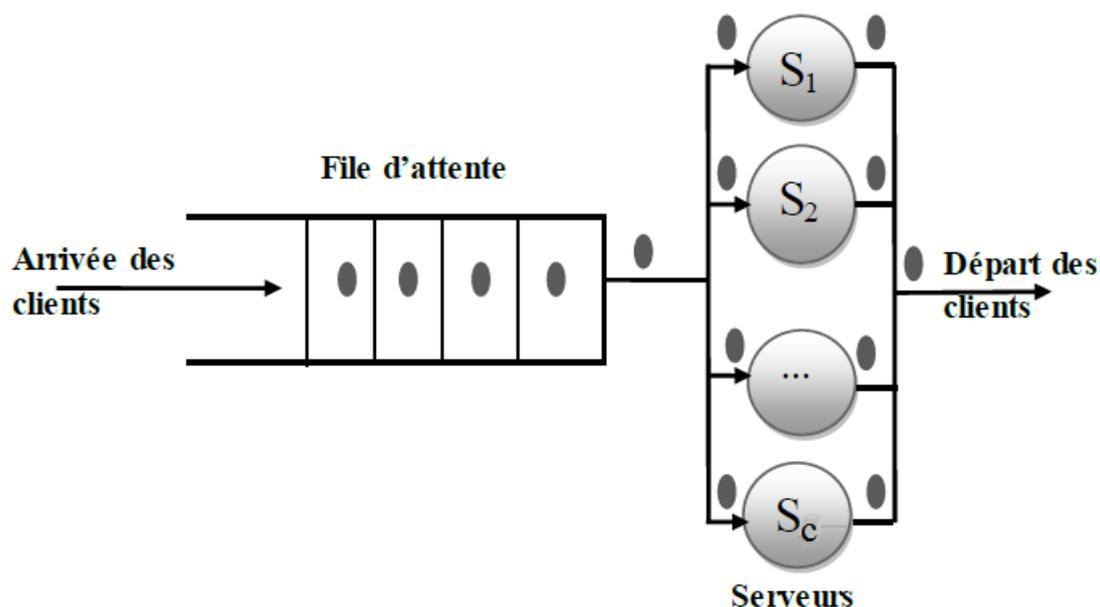
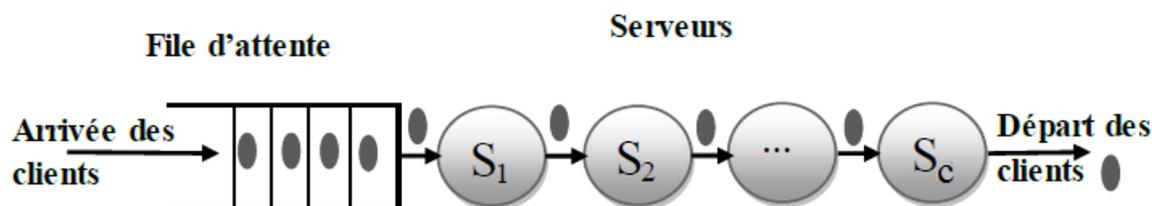


FIGURE 2.2 – Système de file d'attente avec un serveur unique.

FIGURE 2.3 – Système de file d'attente à  $C$  serveurs parallèles.FIGURE 2.4 – Système de file d'attente à  $C$  serveurs en série.

Capacité de la file :

La capacité de la file à accueillir des clients en attente de service peut être finie ou infinie. Soit  $N$  la capacité de la file, une file à capacité illimitée vérifie  $N = +\infty$ .

## 2.2 Notation de Kendall

La notation suivante, appelée la notation de Kendall [10], est largement utilisée pour classer les différents systèmes de files d'attente :

$$T/Y/C/N/m/Z$$

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, qui est un entier positif.
4.  $N$  : 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.
- PNP (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 inférieure 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 2.2.1.** *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.*

## 2.3 Loi de Little

La loi de Little est une relation très générale qui s'applique à une grande classe de systèmes. Elle ne concerne que le régime permanent du système. Aucune hypothèse sur les variables aléatoires qui caractérisent le système (temps d'interarrivées, temps de service, etc). La seule condition d'application de la loi de Little est que le système soit stable. Le débit du système est alors indifféremment soit le débit d'entrée, soit le débit de sortie. La loi de Little s'exprime telle que dans le théorème 2.3.1 suivante :

**Théorème 2.3.1.** *(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 :*

$$L = W \times d$$

où  $d$  est le taux d'entrée dans le système ( $d_e = \lambda$  pour une file M/M/1)

*On a vu que la loi de Little nous dit qu'il existe une relation entre le nombre moyen de clients dans la file (en attente ou en service) et le temps moyen total de séjour d'un client dans la file (temps d'attente + temps de service).*

*La loi de Little peut aussi s'appliquer en considérant uniquement l'attente dans la queue (sans le service). Elle permet alors de relier le nombre moyen de clients en attente  $L_q$ , au temps moyen d'attente d'un client avant service  $W_q$ , par la relation :*

$$L_q = W_q \times d$$

*Enfin, on peut appliquer la loi de Little en ne considérant que le serveur. Dans ce cas, elle relie le nombre moyen de clients en service  $L_s$ , au temps moyen de séjour d'un client*

dans le serveur qui n'est rien d'autre que le temps moyen de service  $\frac{1}{\mu}$ , par la relation :

$$L_s = \frac{1}{\mu} \times d$$

On a obtenu trois relations en appliquant la loi de Little successivement au système entier, à la file d'attente seule et enfin, au serveur seul. Ces trois relations ne sont bien sûr pas indépendantes. On peut en effet déduire l'une d'entre elles à partir des deux autres en remarquant que :

$$W = W_q + \frac{1}{\mu} \quad \text{et} \quad L = L_q + L_s$$

**Remarque 2.3.1.** La loi de Little s'applique à tous les modèles de file d'attente rencontrés en pratique (pas seulement à la file  $M/M/1$ ).

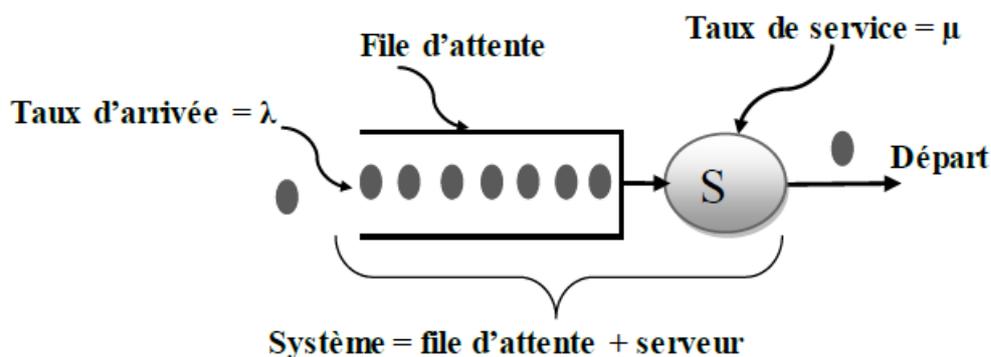
## 2.4 Quelque modèles de files d'attente

### 2.4.1 Modèle d'attente $M/M/1$

#### Description du modèle :

Le système de files d'attente  $M/M/1$  est le système le plus élémentaire de la théorie des files d'attente.

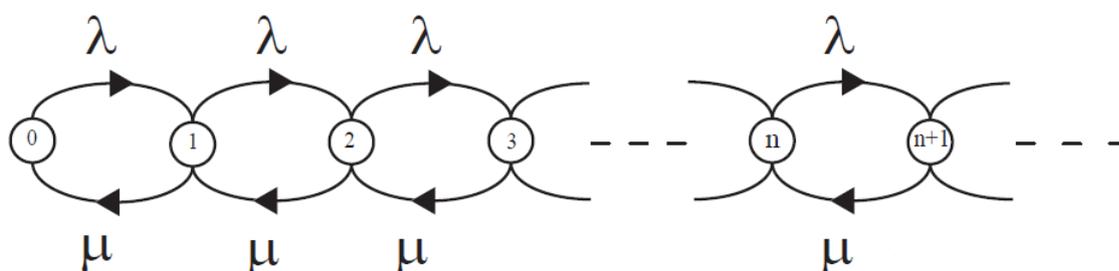
La file d'attente  $M/M/1$  est un modèle caractérisé par des arrivées suivant un processus de Poisson de taux  $\lambda$ , temps de service exponentiel de paramètre  $\mu$  et un seul serveur. Les clients arrivent à la station selon un processus de Poisson de taux  $\lambda$ , si le serveur est vide 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  $\mu$ . Ces temps de service sont également supposés indépendants [23]. En outre, toutes les variables aléatoires concernés sont censés être indépendantes les unes des autres. Les clients sont servis dans leur ordre d'arrivée.

FIGURE 2.5 – La file  $M/M/1$ .

La file peut être considérée comme un processus de naissance et de mort, pour lequel :

$$\lambda_n = \lambda \quad \forall n \geq 0$$

$$\mu_n = \begin{cases} \mu & \forall n \geq 1 \\ 0 & \text{si } n = 0 \end{cases}$$

FIGURE 2.6 – Graphe de transition d'une file  $M/M/1$ .

Sous l'hypothèse que  $\lambda < \mu$  (le taux d'arrivée est plus petit que le taux de service).  
On a :

$$\rho = \frac{\lambda}{\mu} < 1$$

La quantité  $\rho$  mentionne l'intensité du trafic.

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

$$\begin{cases} \pi_n = \pi_0 \rho^n, \\ \pi_0 = \frac{1}{\sum_{n=0}^{+\infty} \rho^n} = 1 - \rho. \end{cases}$$

Donc

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

$\pi = \{\pi_n\}_{n \geq 0}$  est appelé distribution stationnaire (la probabilité stationnaire d'avoir  $n$  clients dans la file).

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

#### **Débit $d$ :**

Le débit du système (noté  $d$ ) représente la probabilité pour que la file ne soit pas vide. Ici  $d = \lambda$  car  $\lambda_n = \lambda$  pour tout  $n \geq 0$ .

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

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

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

#### **Nombre moyen de clients dans le système $L$ :**

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

$$\begin{aligned}
L &= \sum_{n=1}^{+\infty} n\pi_n, \\
&= \sum_{n=1}^{+\infty} n(1-\rho)\rho^n, \\
&= \rho(1-\rho) \sum_{n=0}^{+\infty} (n+1)\rho^n, \\
&= \rho(1-\rho)(1 + 2\rho + 3\rho^2 + \dots), \\
&= \rho(1-\rho) \frac{d}{d\rho} (\rho + \rho^2 + \rho^3 + \dots), \\
&= \rho(1-\rho) \frac{d}{d\rho} \left( \frac{1}{1-\rho} - 1 \right).
\end{aligned}$$

Soit

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

**Temps moyen de séjour  $W$  :**

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 :

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

**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  $L_q$  :**

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

## 2.4.2 Modèle d'attente $M/M/1/N$

### Description du modèle :

On considère une file d'attente simple avec 1 serveur et une capacité  $N$ . Les hypothèses sont les mêmes que pour la file  $M/M/1$  :

- 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.

Dans la pratique, les files d'attente sont toujours finie. Dans ce cas, quand un client arrive alors qu'il y a déjà  $N$  clients présents devant lui dans le système, il est perdu, (par exemple, les appels téléphoniques). Ce système est connu sous le nom de file  $M/M/1/N$ .

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.

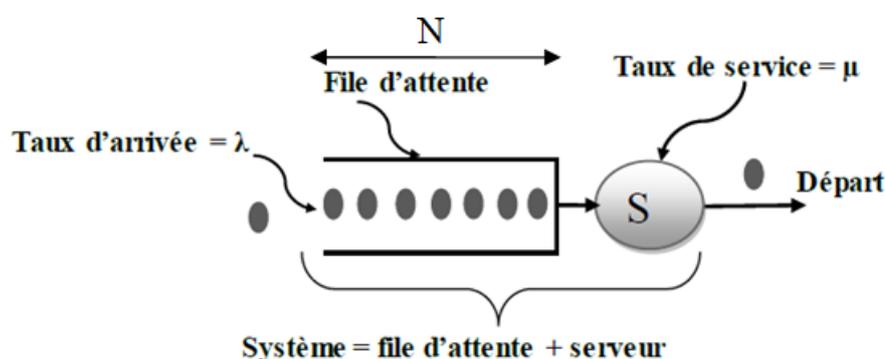
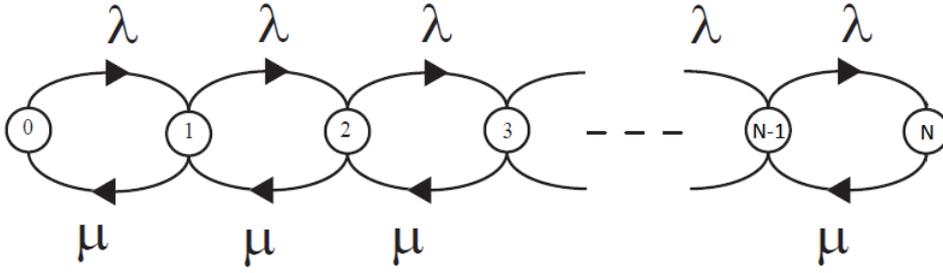


FIGURE 2.7 – La file  $M/M/1/N$ .

Le processus de naissance et de mort modélisant ce type de file d'attente est alors défini de la façon suivante :

$$\lambda_n = \begin{cases} \lambda & \text{si } n < N \\ 0 & \text{si } n \geq N \end{cases}$$

$$\mu_n = \begin{cases} \mu & \text{si } n \neq 0 \\ 0 & \text{si } n = 0 \end{cases}$$

FIGURE 2.8 – Graphe de transition d'une file  $M/M/1/N$ .

Soit  $\pi_n, n = 0, 1, \dots, N$ , la probabilité stationnaire d'avoir  $n$  clients dans la file est :

$$\pi_n = \begin{cases} \pi_0 \rho^n & \text{si } n \leq N \\ 0 & \text{si } n > N \end{cases}$$

$$\pi_0 = \frac{1}{\sum_{n=0}^N \rho^n} = \frac{1 - \rho}{1 - \rho^{N+1}} \quad \text{si } \lambda \neq \mu \quad (\text{et } \frac{1}{N+1} \quad \text{si } \lambda = \mu).$$

### 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  $d_e$ . On s'attend bien sûr à obtenir l'égalité de ces deux débits.

1. Le débit effectif d'entrée dans la file est égal à  $\lambda$  dès l'instant qu'un client arrive lorsque la file n'est pas pleine :

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

2. Le débit en sortie du système est égal à  $\mu$  dès l'instant où la file n'est pas vide :

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

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

**Taux d'utilisation du serveur  $U(N)$  :**

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

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

**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 dans le système  $L$  :**

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

Pareil que le taux d'utilisation du serveur, lorsque  $N \rightarrow +\infty$  et  $\rho < 1$  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 :

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

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

$$L_q = \sum_{n=1}^{\infty} (n-1)\pi_n = L - (1 - \pi_0).$$

Temps moyen passé dans la file d'attente  $W_q$  :

$$W_q = \frac{L_q}{d}.$$

### 2.4.3 Modèle d'attente $M/M/c$

**Description du modèle :**

On considère un système identique à la file  $M/M/1$  excepté qu'il comporte  $c$  serveurs identiques et indépendants les uns des autres. On conserve les hypothèses :

- 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  $\mu$ .

Ce système est connu sous le nom de file  $M/M/c$  [15]. L'espace d'états  $E$  est comme pour la  $M/M/1$  infini :  $E = \{0, 1, 2, \dots\}$ . La file d'attente est de capacité infini.

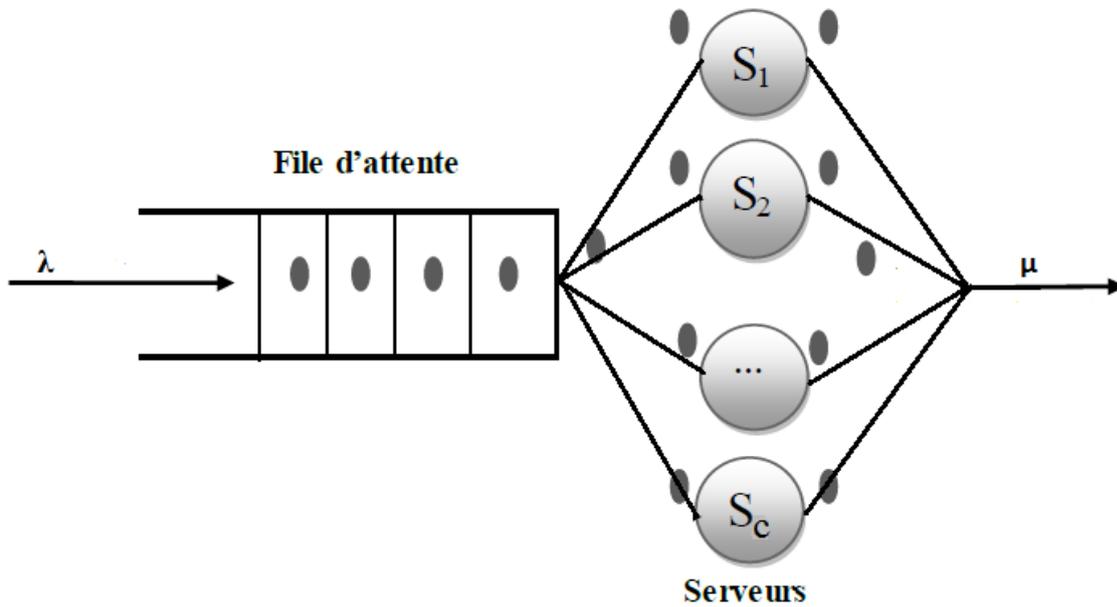


FIGURE 2.9 – La file  $M/M/c$ .

Dans ce cas aussi, Le processus de naissance et de mort modélisant ce type de file d'attente est alors défini de la façon suivante :

$$\lambda_n = \lambda \quad \forall n \geq 0$$

$$\mu_n = \begin{cases} 0 & \text{si } n = 0 \\ n\mu & \text{si } n = 1, \dots, c \\ c\mu & \text{si } n \geq c \end{cases}$$

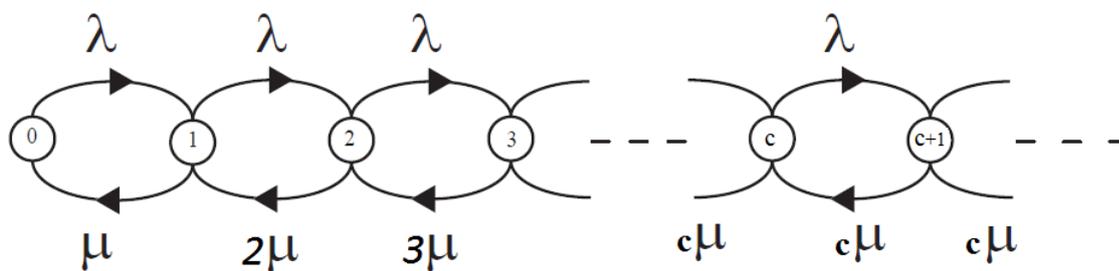


FIGURE 2.10 – Graphe de transition d'une file  $M/M/c$ .

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

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

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, le débit moyen de sortie est égal au débit moyen d'entrée :

$$d = \lambda.$$

Pour des raisons 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,  $L_q$  :

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

On en déduit l'expression du temps moyen de séjour

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

#### Nombre moyen de clients $L$ :

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

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

#### Temps moyen passé dans la file d'attente $W_q$ :

$$W_q = \frac{L_q}{d}.$$

### 2.4.4 Modèle d'attente $M/M/\infty$

#### **Description du modèle :**

On considère un système composé d'un nombre illimité de serveurs identiques et indépendants les uns des autres. Dès qu'un client arrive, il rentre donc instantanément en service. Dans cette file particulière, il n'y a pas donc d'attente. On suppose toujours que le processus d'arrivée des clients est Poissonien de taux  $\lambda$  et que les temps de service sont exponentiels de taux  $\mu$  (pour tous les serveurs). Ce système est connu sous le nom de file  $M/M/\infty$ .

Comme cela a été fait pour la file  $M/M/c$ , on peut facilement démontrer que le taux de

transition d'un état  $n$  quelconque vers l'état  $n - 1$  est égal à  $n\mu$  et correspond au taux de sortie d'un des  $n$  clients en service [14]. De même, le taux de transition d'un état  $n$  vers l'état  $n + 1$  est égal à  $\lambda$  et correspond au taux d'arrivée d'un client.

De façon intuitive, la capacité de traitement de la file est infinie puisque tout nouveau client se présentant à l'entrée de la file est instantanément traité. La condition de stabilité exprimant que "le nombre moyen de client arrivant à la file par unité de temps doit être inférieure à la capacité de traitement de la file" est donc toujours satisfaite.

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

$$\pi_{n-1}\lambda = \pi_n n\mu \quad \text{pour } n = 1, 2, \dots$$

avec

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

Et on obtient finalement :

$$\pi_n = \frac{\rho^n}{n!} \pi_0 \quad \text{pour } n = 1, 2, \dots$$

Par le calcul de  $\pi_n$  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 :

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

On retrouve la stabilité inconditionnelle de la file.

**Taux d'utilisation du serveur  $U$  :**

$$U = 1 - \pi_0 = 1 - e^{-\rho}.$$

**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 :

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

# Chapitre 3

## Les systèmes de file d'attente avec découragement

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.

Les files d'attente avec découragement peuvent être considérées comme un phénomène caractéristique de la vie contemporaine. Les travaux de plusieurs mathématiciens tels que Haight, 1959 [17], qui a étudié le modèle d'attente  $M/M/1$  avec abandon.

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

- Avec abandon Après un temps passé dans la file, le client impatient décide de quitter le système sans avoir le service.

### 3.1 Le modèle d'attente $M/M/1/N$ avec découragement

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.

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  $\xi$ .

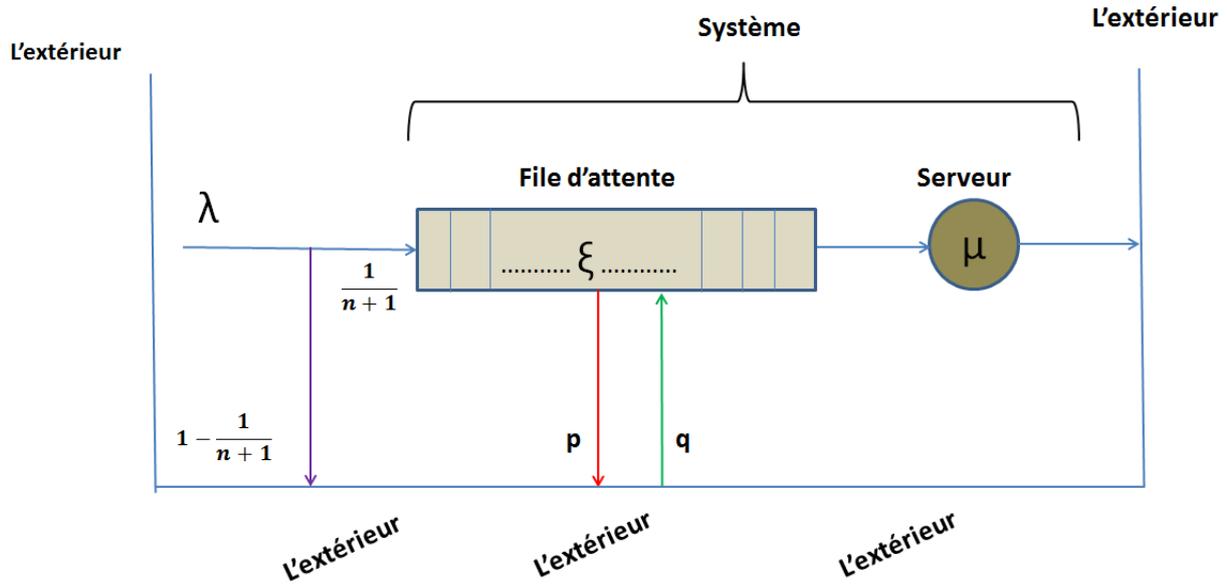


FIGURE 3.1 – Modèle  $M/M/1/N$  avec découragement.

Soit  $P_n(t)$  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 :

$$\frac{d}{dt}P_0(t) = -\lambda P_0(t) + \mu P_1(t) \quad (3.1)$$

$$\begin{aligned} \frac{d}{dt}P_n(t) = & - \left[ \left( \frac{\lambda}{n+1} \right) + \mu + (n-1)\xi p \right] P_n(t) + (\mu + n\xi p)P_{n+1}(t) \\ & + \left( \frac{\lambda}{n} \right) P_{n-1}(t); \quad n = 1, \dots, N-1 \end{aligned} \quad (3.2)$$

$$\frac{d}{dt}P_N(t) = - [\mu + (N-1)\xi p] P_N(t) + \left( \frac{\lambda}{N} \right) P_{N-1}(t) \quad (3.3)$$

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 (3.1) à (3.3) donne :

$$0 = -\lambda P_0 + \mu P_1 \quad (3.4)$$

$$0 = - \left[ \left( \frac{\lambda}{n+1} \right) + \mu + (n-1)\xi p \right] P_n + (\mu + n\xi p) P_{n+1} + \left( \frac{\lambda}{n} \right) P_{n-1}; \quad n = 1, \dots, N-1 \quad (3.5)$$

$$0 = - [\mu + (N-1)\xi p] P_N + \left( \frac{\lambda}{N} \right) P_{N-1} \quad (3.6)$$

En résolvant les équations de (3.4) à (3.6), on obtient

$$P_n = \left[ \frac{1}{n!} \prod_{k=1}^n \frac{\lambda}{\mu + (k-1)\xi p} \right] P_0; \quad n = 1, \dots, N \quad (3.7)$$

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

$$P_0 = \frac{1}{1 + \sum_{n=1}^N \left( \frac{1}{n!} \prod_{k=1}^n \frac{\lambda}{\mu + (k-1)\xi p} \right)} \quad (3.8)$$

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  $E(\text{client servis})$  :

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

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

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

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

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

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

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

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

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 (3.8).

**Cas particuliers :**

Lorsqu'il n'y a pas de rétention des clients abandonnés (i.e.  $q = 0$ ).

Le système de file d'attente est réduit à un système avec abandon tel que :

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

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

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

Quand il n'y a aucun abandon.

On traite deux sous-cas :

1. Le modèle  $M/M/1/N$  avec rétention de clients abandonnés (il a été étudié par Kumar Sharma [26])

$$P_n = \prod_{k=1}^n \frac{\lambda}{\mu + (k-1)\xi p} P_0; \quad n = 1, \dots, N-1$$

Aussi pour  $n = N$  on obtient

$$P_N = \prod_{k=1}^N \frac{\lambda}{\mu + (k-1)\xi p} P_0$$

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

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

2. Lorsqu'il n'y a pas d'abandon (c-à-d que les clients ne s'impatientent pas).

Dans ce cas, la probabilité d'abandonner ( $P$ ) est égal à zéro, implique que  $\xi = 0$ . Tous les clients qui entrent dans le système quittent après l'obtention de service. Par conséquent, des équations (3.7) et (3.8), on obtient :

$$P_n = \left( \frac{\lambda}{\mu} \right)^n P_0; \quad n = 1, \dots, N$$

Et en utilisant la condition de normalisation, on obtient :

$$P_0 = \frac{1}{1 + \sum_{n=0}^N \left( \frac{\lambda}{\mu} \right)^n}$$

Il est évident que le modèle se réduit à un simple modèle de files d'attente  $M/M/1/N$ .

## 3.2 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 forme des tableaux et des graphiques. La capacité du système est fixée à  $N = 8$ , et la fonction de dérobade est donnée par  $1 - \frac{1}{n+1}$ . Ici, nous avons pris  $\lambda = 4$ ,  $\mu = 3$ ,  $\xi = 0.2$ ,  $q = 0.6$  et  $p = 0.4$ , 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 système soit vide.
- $P_b = 1 - P_0$  : La probabilité que le serveur est occupé.
- $R_e = \sum_{n=1}^N (n-1)\xi q P_n$  La rétention des clients abandonnés.

TABLE 3.1 – Variations des paramètres de performance vs.  $N$

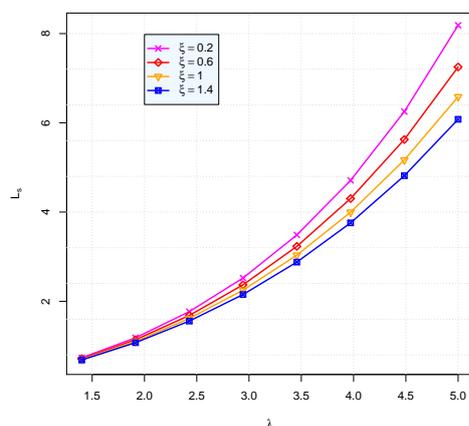
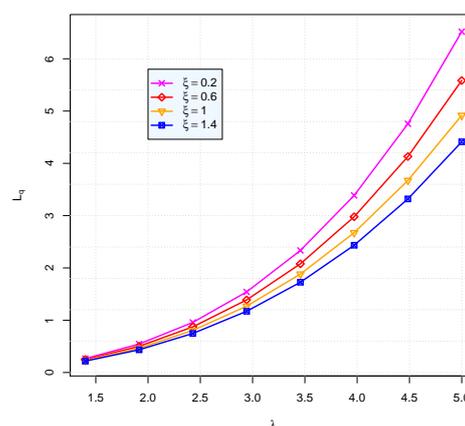
$N$	$P_i$	$P_b$	$L_s$	$L_q$	$W_s$	$W_q$	$R_a$	$R_e$
3	0.158245	0.841755	4.097921	2.764588	1.024480	0.691147	0.491751	0.327834
4	0.127298	0.872702	4.516646	3.183313	1.129161	0.795828	0.541998	0.361332
5	0.112129	0.887871	4.636969	3.303636	1.159242	0.825909	0.556436	0.370958
6	0.104326	0.895674	4.663708	3.330374	1.165927	0.832594	0.559645	0.373097
7	0.100324	0.899676	4.668500	3.335166	1.167125	0.833792	0.560220	0.373480
8	0.098357	0.901643	4.669213	3.335879	1.167303	0.833970	0.560306	0.373537
9	0.097461	0.902539	4.669303	3.335969	1.167326	0.833992	0.560316	0.373544
10	0.097093	0.902907	4.669312	3.335979	1.167328	0.833995	0.560317	0.373545
11	0.096960	0.903040	4.669313	3.335980	1.167328	0.833995	0.560318	0.373545
12	0.096919	0.903081	4.669313	3.335980	1.167328	0.833995	0.560318	0.373545

Pour ce premier tableau (3.1), on remarque que la croissance de  $N$  (la capacité du système) provoque une décroissance de  $P_i$  (la probabilité que le système soit vide), par contre pour les autres paramètres on voit une croissance ce qui est logique.

TABLE 3.2 – Variations des paramètres de performance vs.  $\lambda$ 

$\lambda$	$P_i$	$P_b$	$L_s$	$L_q$	$W_s$	$W_q$	$R_a$	$R_e$
1,4	0.610799	0.389201	0.728277	0.261611	0.520198	0.186865	0.087393	0.058262
1.7	0.541623	0.458377	0.971929	0.405262	0.571723	0.238390	0.116631	0.077754
2.0	0.476028	0.523972	1.256304	0.589637	0.628152	0.294819	0.150756	0.100504
2.3	0.413599	0.586401	1.586848	0.820181	0.689934	0.356600	0.190422	0.126948
2.6	0.353956	0.646044	1.969650	1.102983	0.757558	0.424224	0.236358	0.157572
2.9	0.296714	0.703286	2.411508	1.444842	0.831555	0.498221	0.289381	0.192921
3.2	0.241428	0.758572	2.920010	1.853343	0.912503	0.579170	0.350401	0.233601
3.5	0.187491	0.812509	3.503610	2.336943	1.001031	0.667698	0.420433	0.280289
3.8	0.134021	0.865979	4.171724	2.905058	1.097822	0.764489	0.500607	0.333738
4.1	0.080788	0.919212	4.934829	3.568162	1.203617	0.870283	0.592179	0.394786

Pour ce tableau (3.2), la croissance de  $\lambda$  (taux d'arrivée) implique une croissance de tout les paramètres de performances sauf la probabilité que le système soit vide cela est du au nombre important de clients qui arrivent au système.

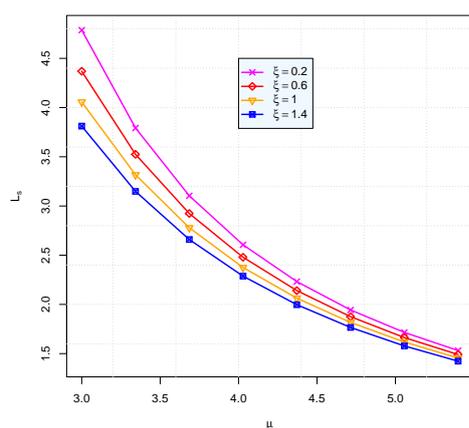
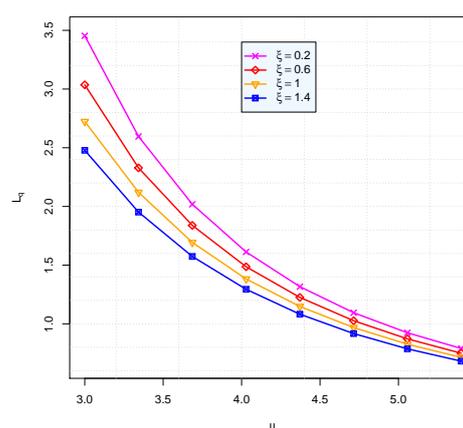
FIGURE 3.2 –  $L_s$  en fonction de  $\lambda$  pour différentes valeurs de  $\xi$ FIGURE 3.3 –  $L_q$  en fonction de  $\lambda$  pour différentes valeurs de  $\xi$ 

Par la croissance de  $\lambda$  (taux d'arrivée)  $L_s$ ,  $L_q$  croissent aussi, mais dans le cas  $\xi = 0.2$  (taux d'impatience) la croissance de  $L_s$  est plus important que dans le cas  $\xi = 1.4$  car ( $0.2 < 1.4$ ) ce qui confirme les interprétations précédentes (figure 3.2, 3.3).

TABLE 3.3 – Variations des paramètres de performance vs.  $\mu$ 

$\mu$	$P_i$	$P_b$	$L_s$	$L_q$	$W_s$	$W_q$	$R_a$	$R_e$
3.0	0.098357	0.901643	4.669213	3.335879	1.167303	0.833970	0.560306	0.373537
3.2	0.141642	0.858358	4.071074	2.821074	1.017769	0.705269	0.488529	0.325686
3.4	0.180423	0.819577	3.591438	2.414967	0.897859	0.603742	0.430973	0.287315
3.6	0.215310	0.784690	3.200469	2.089358	0.800117	0.522339	0.384056	0.256038
3.8	0.247003	0.752997	2.877174	1.824542	0.719293	0.456136	0.345261	0.230174
4.0	0.276017	0.723983	2.606438	1.606438	0.651609	0.401609	0.312773	0.208515
4.2	0.302733	0.697267	2.377163	1.424782	0.594291	0.356195	0.285260	0.190173
4.4	0.327445	0.672555	2.181049	1.271958	0.545262	0.317990	0.261726	0.174484
4.6	0.350389	0.649611	2.011791	1.142226	0.502948	0.285557	0.241415	0.160943
4.8	0.371760	0.628240	1.864529	1.031195	0.466132	0.257799	0.223743	0.149162

Le nombre de clients dans le système et dans la file diminue ainsi que les temps d'attente dans la file et dans le système. Proportionnellement la moyenne des clients abandonnés et la moyenne des clients retenus après que leur temps d'impatience expire décroissent. Cela est interprété par la croissance du taux de service c-à-d les clients sont servis rapidement (tableau 3.3).

FIGURE 3.4 –  $L_s$  en fonction de  $\mu$  pour différentes valeurs de  $\xi$ FIGURE 3.5 –  $L_q$  en fonction de  $\mu$  pour différentes valeurs de  $\xi$ 

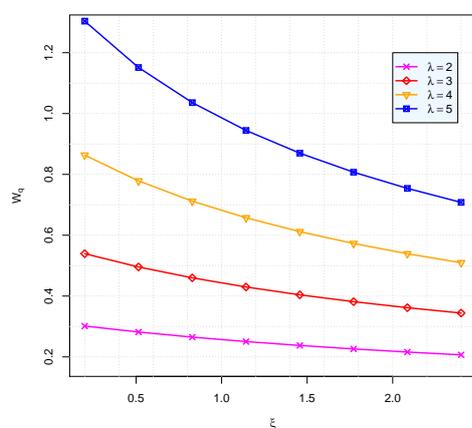
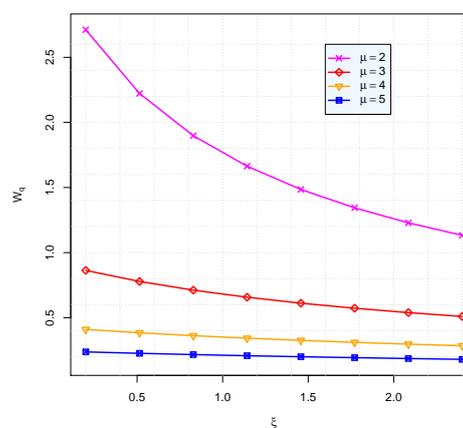
Par la croissance de  $\mu$  (taux de service)  $L_s$ ,  $L_q$  décroissent. Cette décroissance est plus importante dans le cas  $\xi = 0.2$  que dans le cas  $\xi = 1.4$  (figure 3.4, 3.5).

TABLE 3.4 – Variations des paramètres de performance vs.  $\xi$ 

$\xi$	$P_i$	$P_b$	$L_s$	$L_q$	$W_s$	$W_q$	$R_a$	$R_e$
0.1	0.081274	0.918726	4.849711	3.516377	1.212428	0.879094	0.290983	0.193988
0.3	0.110831	0.889169	4.510447	3.177114	1.127612	0.794279	0.811881	0.541254
0.5	0.128465	0.871535	4.242412	2.909079	1.060603	0.727270	1.272724	0.848482
0.7	0.141148	0.858852	4.023186	2.689853	1.005796	0.672463	1.689738	1.126492
0.9	0.151187	0.848813	3.839363	2.506030	0.959841	0.626507	2.073256	1.382171
1.1	0.159546	0.840454	3.682289	2.348956	0.920572	0.587239	2.430311	1.620207
1.3	0.166722	0.833278	3.546060	2.212726	0.886515	0.553182	2.765927	1.843951
1.5	0.173012	0.826988	3.426473	2.093139	0.856618	0.523285	3.083825	2.055884
1.7	0.178608	0.821392	3.320437	1.987104	0.830109	0.496776	3.386846	2.257897
1.9	0.183644	0.816356	3.225619	1.892286	0.806405	0.473072	3.677206	2.451471

Le tableau (3.4), montre que le long de la croissance de  $\xi$  (le taux d'impatience) le nombre moyen de clients dans le système et dans la file ainsi que les temps d'attente dans la file et dans le système décroissent cela est du à l'abandon des clients qui croit par la croissance de  $\xi$ . Et comme l'abandon des clients croit la rétention des clients dans le système augmente ce qui est logique.

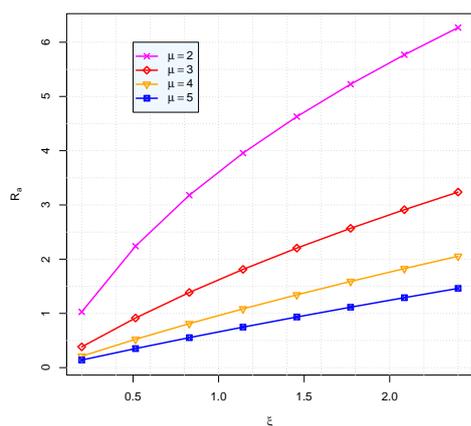
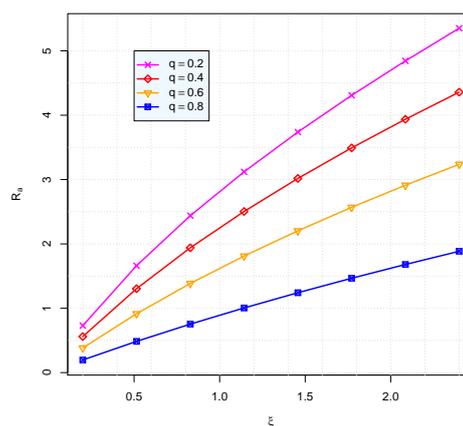
L'effet de la rétention est bien présent dans les résultats du tableau (3.5). Car la croissance de la probabilité de la rétention des clients dans le système a un impact positif sur l'état du système.

FIGURE 3.6 –  $W_q$  en fonction de  $\xi$  pour différentes valeurs de  $\lambda$ FIGURE 3.7 –  $W_q$  en fonction de  $\xi$  pour différentes valeurs de  $\mu$

Le temps d'attente dans la file  $W_q$  décroît par la croissance de  $\xi$  (taux d'impatience). Ce qui est logique, de plus pour  $\mu = 5$  le temps de séjour dans la file est moins que dans le cas  $\mu = 2$  et le temps d'attente dans la file pour  $\lambda = 2$  est moins que dans le cas  $\lambda = 5$  (figure 3.6, 3.7).

TABLE 3.5 – Variations des paramètres de performance vs.  $q$ 

$q$	$P_i$	$P_b$	$L_s$	$L_q$	$W_s$	$W_q$	$R_a$	$R_e$
0.0	0.114306	0.885694	4.461594	3.128260	1.115398	0.782065	0.892319	0.000000
0.1	0.110831	0.889169	4.510447	3.177114	1.127612	0.794279	0.811881	0.090209
0.2	0.107054	0.892946	4.561241	3.227907	1.140310	0.806977	0.729799	0.182450
0.3	0.102918	0.897082	4.614111	3.280778	1.153528	0.820194	0.645976	0.276847
0.4	0.098357	0.901643	4.669213	3.335879	1.167303	0.833970	0.560306	0.373537
0.5	0.093290	0.906710	4.726715	3.393382	1.181679	0.848345	0.472672	0.472672
0.6	0.087628	0.912372	4.786810	3.453476	1.196702	0.863369	0.382945	0.574417
0.7	0.081274	0.918726	4.849711	3.516377	1.212428	0.879094	0.290983	0.678959
0.8	0.074147	0.925853	4.915659	3.582326	1.228915	0.895582	0.196626	0.786506
0.9	0.066204	0.933796	4.984931	3.651597	1.246233	0.912899	0.099699	0.897288

FIGURE 3.8 –  $R_a$  en fonction de  $\xi$  pour différentes valeurs de  $\mu$ FIGURE 3.9 –  $R_a$  en fonction de  $\xi$  pour différentes valeurs de  $q$ 

Par la croissance de  $\xi$  (taux d'impatience) le taux d'abandon du système  $R_a$  croit. De plus la croissance de  $R_a$  pour le cas  $\mu = 2$  est plus importante que pour le cas  $\mu = 5$  (figure 3.8). Cela est interprété par  $\xi$  lui même.

Et la croissance est aussi plus important dans le cas  $q = 0.2$  (la probabilité de rétention) que dans le cas  $q = 0.8$  (figure 3.9). Car plus la rétention des clients dans le système augmente plus le taux d'abandon diminue.

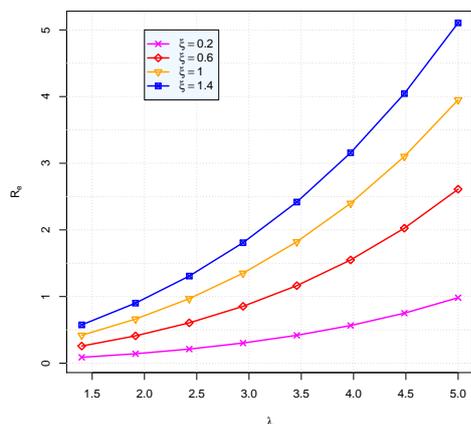


FIGURE 3.10 –  $R_e$  en fonction de  $\lambda$  pour différentes valeurs de  $\xi$

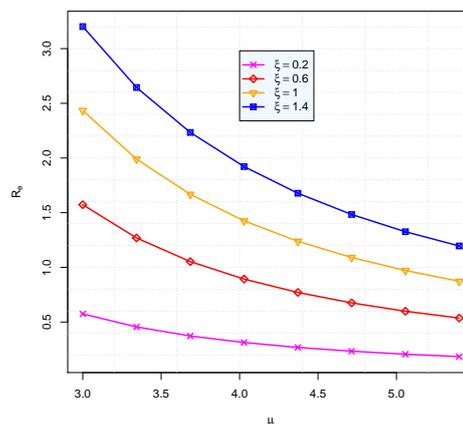


FIGURE 3.11 –  $R_e$  en fonction de  $\mu$  pour différentes valeurs de  $\xi$

Le taux de rétention croit par la croissance de  $\lambda$  et décroît par la croissance de  $\mu$ , pour le cas  $\xi = 1.4$  la croissance dans la figure (3.10) et la décroissance dans la figure (3.11) de  $R_e$  est plus important que dans le cas  $\xi = 0.2$ .

# Conclusion générale

Dans ce mémoire, on a étudié un modèle de files d'attente avec serveur unique et découragement des clients (abandon et rétention de clients). La solution stationnaire et les différents paramètres de performance sont également donnés. Certains modèles de files d'attente sont dérivés comme un cas spécial de ce modèle.

L'analyse du modèle est limitée à capacité finie. Le cas de capacité infinie du modèle peut également être étudié. En outre, le modèle peut être résolu dans un état transitoire. L'analyse économique du modèle peut également être mis à l'étude. La même idée peut être étendue à certains modèles de files d'attente non-markovien.

# Bibliographie

- [1] Abou El-Ata, M. O. and Hariri, A. M. A. The  $M/M/c/N$  queue with balking and renegeing. Computers and Operations Research. 19 (1992), No. 13 713-716.
- [2] Ammar, S.I., El-Sherbiny, A.A., and Al-Seedy, R.O., "A matrix approach for the transient solution of an  $M/M/1/N$  queue with discouraged arrivals and renegeing", International Journal of Computer Mathematics, 89 (2012) 482-491.
- [3] Ancker Jr., C. J., and Gafarian, A. V., "Some Queuing Problems with Balking and Renegeing I", Operations Research, 11 (1963) 88-100.
- [4] Ancker, Jr., C. J., and Gafarian, A. V., "Some Queuing Problems with Balking and Renegeing. II.", Operations Research, 11 (1963) 928-937.
- [5] Anisimov, V. V., Zakusilo, O. K., and Donchenko, V. S. 1987. Elements of Queueing Theory and Asymptotic System Analysis. Vishcha Shkola, Kiev (in Russian).
- [6] Artalejo. J. R and A. Gomez-Corral. Channel idle periods in computer and telecommunication systems with customer retrials. Telecommunication Systems, 24, 29-46, (2003).
- [7] Borovkov, A. A. Stochastic Processes in Queueing Theory. Springer-Verlag, Berlin (Russian original : Nauka, Moscow),(1972).
- [8] Bouchentouf. A. A, M. Kadi, A. Rabhi 2013 Analysis of Two Heterogeneous Server Queueing Model With Balking, Renegeing and Feedback. Mathematical Sciences And Applications E-Notes Volume 2 No. 2 pp. 10-21 (2013)
- [9] Claudie Chabriac. Processus stochastiques et modélisation. (2012-2013).
- [10] Claudie Hasseforder CHABRIAC , Eléments de Théorie des files d'attente, page05, Janvier 2008.
- [11] Courtois, P.J., and Georges, J., "On a Single Server Finite Capacity Queueing Model with State dependent Arrival and Service Process", Operations Research. 19 (1971) 424-435.
- [12] Erlang, A. K. Probability and Telephone Calls. Nyt Tidsskrift Matematik Series B, 20,33 - 39. (1909).

- [13] Garnett. O, A. Mandelbaum, and M. Reiman. Designing a call center with impatient customers. *Manufacturing Service Operations Management*, 4(3) : 208-227, June 2002.
- [14] Gomez-Corral. A. 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).
- [15] Gross, D. and C.M, Harris, *Fundamentals of Queueing Theory*, Willey, New York, 1975 1984.
- [16] Hadidi, N., "Busy periods of queues with state dependent arrival and service rates", *Journal of Applied Probability*, 11 (1974) 842-848.
- [17] Haight, F. A., "Queueing with Reneging", *Metrika*, 2 (1959)186-197.
- [18] 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.
- [19] Kapodistria, S., "The  $M/M/1$  Queue with Synchronized Abandonments", *Queueing Systems*, 68 (2011) 79-109.
- [20] Kashyap, B. R. K., and Chaudhry, M. L. 1988. *An Introduction to Queueing Theory*. Aarkay, Calcutta, India.
- [21] 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).
- [22] Khintchine, A. Y 1969. *Mathematical Methods in the Theory of Queueing*. Second edition, Hafner Publishing Company, New York (First edition : Griffin, London, 1960 ; Russian original : 1955). 859-866.
- [23] Kumar, R. and Sharma, S.K. An  $M/M/1/N$  Queueing Model with Retention of renege customers and Balking, *American Journal of Operational Research*. 2(2012), No. 1, 1-5.
- [24] Kumar, R.A., "Catastrophic-cum-Restorative Queueing Problem with Correlated Input and Impatient Customers", *International Journal of Agile Systems and Management*, 5 (2012) 122-131.
- [25] Kumar, R., and Sharma, S.K., "Managing congestion and revenue generation in supply chains facing customer impatience", *Inventi Impact : Supply Chain Logistics*, 2012 (2012) 13-17.
- [26] Kumar, R., and Sharma, S.K., " $M/M/1/N$  queueing system with retention of renege customers", *Pakistan Journal of Statistics and Operation Research*, 8 (2012) 859-866.

- [27] K.B, GK, Techniques de modélisation : Méthodes analytiques.
- [28] Libman, L. and Orda, A. (2002). Optimal retrial and timeout strategies for accessing network resources. *IEEE/ACM Transactions on Networking*, 10(4), 551- 564.
- [29] Lubacz. J. 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).
- [30] 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.
- [31] Morse, P.M., *Queues, Inventories and Maintenance*, Wiley, New York, (1968)
- [32] Moulay Hachemi, *Files d'attente et applications*, page 12, 2014/2015.
- [33] Natvig, B., "On a Queuing Model Where Potential Customers Are Discouraged by Queue Length", *Scandinavian Journal of Statistics*, 2 (1975) 34-42.
- [34] Newell, G. F. 1982. *Applications of Queueing Theory*. Second edition, Chapman and Hall, London (First edition : 1971).
- [35] Nobel. R., A Discrete-Time Retrial Queueing Model with One Server, (China), *INFORMS/APS Conference*, June (2004).
- [36] Petit. M., *Introduction à la modélisation des réseaux*, page 21, 26 Octobre 2010.
- [37] Raynolds, J.F., "The stationary solution of a multi-server queueing model with discouragement", *Operations research*, 16 (1968) 64-71.
- [38] Robert, E. Reneging phenomenon of single channel queues. *Mathematics of Operations Research* 4(2) 162-178. (1979).
- [39] Rubino. G., *Processus Stochastiques*, Février 2006.
- [40] Saaty, T. L. *Elements of queueing theory with applications*. New York : McGraw Hill. (1961).
- [41] Velenik. Y., *Probabilités et Statistique*, université de Genève, Version du (24 mai 2012), pp 195-202
- [42] 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) (2002) 169-180.
- [43] Wang, K., Li, N., and Jiang, Z., "Queueing System with Impatient Customers : A Review", 2010 *IEEE International Conference on Service Operations and Logistics and Informatics*. 15-17 July, 2010, Shandong, (2010) 82-87.

- [44] Xiong, W., and Altıok, T., "An approximation for multi-server queues with deterministic reneging times", *Annals of Operations Research*, 172 (2009) 143-151.
- [45] Yves Caumel. *Probabilités et processus stochastiques*. Springer. Verlag France, (2011).
- [46] Zakhar Kabluchko, *Stochastic Processes (Stochastik II)*, University of Ulm Institute of Stochastics, (2013-2014).