

N° d'ordre : 19/2023.C/MT

République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
Université des Sciences et de la Technologie Houari Boumediène

Faculté des Mathématiques



THESE DE DOCTORAT

Présentée pour l'obtention du **grade de DOCTEUR**

En : Mathématiques

Spécialité : Probabilités Statistiques et Applications

Par : MEZIANI Sara

Sujet

Modèles de Diffusion pour Files d'Attente avec Rappels

Soutenue publiquement, le 15/05/2023 , devant le jury composé de :

Hamdi Fayçal	Professeur	USTHB, Algérie	Président
Kernane Tewfik	Professeur	USTHB, Algérie	Directeur de Thèse
Berdjoudj Louiza	Professeure	Université de Bejaia	Examinatrice
Boualem Mohamed	Professeur	Université de Bejaia	Examinateur
Taleb Samira	Maitre de Conférences A	USTHB, Algérie	Examinatrice
El Maroufy Hamid	Professeur	FSTBM, Maroc	Invité
Medkour Tarek	Professeur	ENSIA, Sidi-Abdellah	Invité

Remerciements

La réalisation de cette thèse a été possible grâce au concours de plusieurs personnes à qui je voudrais présenter mes remerciements, mon respect et ma gratitude.

Je voudrais tout d'abord adresser toute ma reconnaissance au directeur de cette thèse, Monsieur **KERNANE Tewfik**. Je le remercie de m'avoir encadré, orienté, aidé et conseillé.

J'adresse mes sincères remerciements à tous les professeurs, intervenants et toutes les personnes qui par leurs écrits, leurs conseils et leurs critiques ont guidé mes réflexions et ont accepté de me rencontrer et de répondre à mes questions durant mes recherches.

Je remercie mes très chers parents qui ont toujours été là pour moi. Je remercie mes sœurs, mon frère et ma grand mère pour leurs encouragements.

Je désire aussi remercier mon mari et ma belle famille pour leur confiance et leur soutien inestimable.

Enfin, je tiens à témoigner toute ma gratitude à mes amis et collègues qui m'ont apporté leur soutien moral et intellectuel tout au long de ma démarche.

Résumé

Ce travail s'intéresse à l'étude des modèles de files d'attente avec rappels dans deux directions principales. Dans la première partie de cette contribution, le modèle M/G/1 avec une distribution générale des temps de rappels a été traité dans le cadre des processus de Markov déterministes par morceaux connus dans la littérature par les PDMPs (Piece-wise Deterministic Markov Processes). Cette nouvelle approche a permis de déterminer le générateur infinitésimal ainsi que les martingales associées au PDMP modélisant le système considéré. Ces résultats ont ensuite été utilisés pour calculer le nombre moyen de clients bloqués dans l'orbite en régime transitoire. Dans la seconde partie, nous sommes intéressés à l'approximation par les processus de diffusion dans deux modèles de files d'attente M/M/1 avec une distribution exponentielle des temps de rappels. Le premier modèle avec une source finie de clients a été étudié en utilisant la méthode de convergence du générateur infinitésimal. Tandis que le second modèle avec une source de clients infinie a été analysé en se basant sur les propriétés asymptotiques du processus associé au nombre de clients dans l'orbite.

Abstract

This work is concerned with the study of queuing models with retrials in two main directions. In the first part of this contribution, the M/G/1 model with a general retrial times distribution was treated within the framework of piecewise deterministic Markov processes known in the literature as PDMPs. This new approach made it possible to determine the infinitesimal generator as well as the martingales associated with the PDMP modeling the system under consideration. These results were further used to calculate the mean number of customers blocked in the orbit in the transient regime. In the second part, we were interested in the approximation by diffusion processes in two M/M/1 queueing models with an exponential distribution of retrial times. The first model with a finite source of customers was studied using the convergence of the infinitesimal generator method. While the second model with an infinite source of customers was analyzed on the basis of the asymptotic properties of the process associated with the number of customers in the orbit.

Table des matières

Introduction	7
1 Processus Stochastiques	10
1.1 Chaînes de Markov	13
1.1.1 Chaînes de Markov à temps discret	13
1.1.2 Chaînes de Markov à temps continu	17
1.2 Processus de naissance et de mort	19
1.3 Processus de Poisson	21
1.4 Processus de renouvellement	22
1.5 Martingales	25
1.6 Processus de Markov déterministes par morceaux	27
1.6.1 Chaîne de Markov associée	27
1.6.2 Générateur infinitésimal associé	28
1.6.3 Martingale associée	29
1.7 Processus de diffusion	31
1.7.1 Exemples de processus de diffusion	32
1.7.2 Représentation d'Itô d'un processus de diffusion	34
1.7.3 Approximation et simulation	34
2 Files d'Attente	36
2.1 Files d'attente simples	36
2.1.1 Formalisme mathématique d'une file d'attente simple	37
2.1.2 Étude de la file M/M/1	39
2.2 Files d'attente avec rappels	41
2.2.1 Modèle de file d'attente avec rappels avec un seul serveur	42
2.2.2 Modèle de file d'attente avec rappels et plusieurs serveurs	49
3 Étude des files d'attente dans le cadre d'un PDMP	54
3.1 Files d'attente simples en tant que PDMPs	54
3.1.1 File M/G/1	54
3.1.2 File GI/G/1	56

3.1.3	File M/G/c	57
3.2	Étude de la file M/G/1 avec une distribution générale des temps de rappels dans le cadre d'un PDMP	59
3.2.1	File M/G/1 avec politique de rappels classique	59
3.2.2	File M/G/1 avec politique de rappels constante	66
4	Approximation par les processus de diffusion dans les files d'attente	73
4.1	Approximation dans un trafic intense pour des files Markoviennes simples	74
4.1.1	Détermination de l'intervalle \mathbb{E}_c	74
4.1.2	Calcul des paramètres de diffusion	75
4.1.3	Application pour une file M/M/1 simple	76
4.2	Approximation dans un trafic intense pour une file périodique	77
4.3	Approximation dans les files d'attente avec une distribution exponentielle des temps de rappels	81
4.3.1	File M/M/c dans le cas de taux de rappels faible	81
4.3.2	File M/M/1 avec une politique de rappels classique et une source finie de clients	84
4.3.3	File M/M/1 avec une politique de rappels constante	87
	Conclusion	91
	Bibliographie	93

Introduction

Les modèles de file d'attente avec rappels sont spécifiés par la caractéristique suivante : un client arrivant de l'extérieur qui trouve tous les serveurs occupés rejoint une zone d'attente virtuelle appelée orbite. Par conséquent, il devient un client bloqué au lieu de quitter le système. Les clients bloqués dans l'orbite sont censés retenter leur chance pour accéder à la zone de service après un laps de temps aléatoire. Les intervalles de temps entre ces tentatives sont appelés temps de rappels. Les systèmes de files d'attente avec des temps de rappels sont considérés comme une partie spéciale de la théorie des files d'attente qui a un large champ d'applications dans les réseaux informatiques et de communication, dans les problèmes de production et de contrôle de qualité ainsi que dans les systèmes de fabrications flexibles.

Les échanges téléphoniques dans un centre d'appels sont une application bien connue des files d'attente avec rappels dans la littérature. Ce problème a été abordé par Fayolle [28] dans le cas de la politique de rappels constante en analysant la distribution stationnaire des états du systèmes et la distribution du temps de séjour pour une file M/M/1. Dans le travail de Artalejo [3], une analyse exhaustive du système de la file M/G/1 avec rappels et vacances du serveur a été réalisée. De plus, l'auteur a montré que la distribution du nombre de clients dans le système se décompose en trois variables qui représentent la durée des vacances, la politique de rappels et la file M/G/1 ordinaire sans vacances ni rappels, respectivement. La politique de rappels constante a également été introduite pour analyser des systèmes d'attente avec plusieurs serveurs [5, 65]. Par ailleurs, le taux de rappels peut dépendre du nombre de clients bloqués dans l'orbite. Les files d'attente avec rappels possédant cette caractéristique, dite politique de rappels classique, ont été largement étudiées dans la littérature [25, 26]. En considérant cette politique de rappels, la propriété de décomposition stochastique de la file M/G/1 avec rappels a été déterminée dans [1]. Cette étude a été étendue au cas d'une file $M^X/G/1$ avec deux phases de service hétérogène et des vacances de Bernoulli dans [9]. Dans les travaux de Gupta et Kumar [35, 36], le régime stationnaire de la file M/M/1 avec rappels a été étudié en supposant différents types de vacances du serveur. Une étude similaire a été précédemment réalisée pour ce modèle en supposant qu'une interruption peut être produite par le client pendant le service [43]. Les systèmes de files d'attente avec plusieurs serveurs sont également analysés en considérant la politique de rappels classique [57, 64].

Tous les travaux cités ci-dessus ont supposé la distribution exponentielle pour les temps de rappels. Cependant, dans une partie majeure de cette thèse, nous considérons une distribution générale des temps de rappels. La recherche dans cette direction est très compliquée, en particulier lorsqu'il s'agit de la politique de rappels classique. La difficulté de modéliser un modèle aussi complexe par un processus de Markov, en tenant compte des temps de rappels écoulés ou résiduels pour chaque client en orbite qui peut comprendre un très grand nombre de clients bloqués, est à l'origine de cette complexité. Néanmoins, plusieurs travaux dans la littérature ont abordé ce problème, notamment celui de Kapyrin [45]. Bien que Falin [24] ait décliné son approche, ce travail reste la première tentative de généraliser la distribution des temps de rappels dans le cadre d'une étude analytique de la file M/G/1 avec rappels. En outre, de nombreuses approximations et simulations sont présentées, non seulement pour des modèles à un seul serveur [78, 2], mais aussi pour des modèles à plusieurs serveurs [8, 74, 75].

L'analyse des systèmes d'attente avec une distribution générale des temps de rappels demeure un problème de recherche ambitieux dans la théorie des files d'attente. Dans la première partie de notre contribution, nous avons abordé ce problème en traitant le modèle de la file M/G/1 avec une distribution générale des temps de rappels dans le cadre des processus de Markov déterministes par morceaux (PDMPs) introduits par Davis [16, 17]. Ces processus ont déjà servi à analyser des systèmes de files d'attente ainsi que des modèles épidémiques [6, 10, 32]. Cette nouvelle approche nous a permis d'étudier minutieusement la dynamique du modèle considéré en régime transitoire, après avoir défini les caractéristiques des PDMPs correspondants dans le cas de la politique de rappels classique ainsi que le cas de la politique de rappels constante. Ensuite, en utilisant les générateurs infinitésimaux des PDMPs, nous avons déterminé les martingales associées qui ont été ultérieurement exploitées pour le calcul du nombre moyen de clients bloqués dans l'orbite en régime transitoire.

Le nombre de clients dans l'orbite est une quantité de performance essentielle dans l'analyse des files d'attente avec rappels. En effet, avoir des informations sur cette quantité en régime permanent permet de trouver des solutions pour plusieurs problèmes importants rencontrés fréquemment dans les systèmes qui assurent la conception et la maintenance des données. Pour cette raison, la seconde partie de cette contribution a été menée dans ce sens. En premier lieu, nous avons approximé le processus modélisant la file M/M/1 avec rappels et une source finie de clients par un processus de diffusion en spécifiant ses paramètres infinitésimaux par la méthode de convergence du générateur infinitésimal (voir Fuchs [29]). Puis, en deuxième lieu, nous nous sommes intéressés uniquement au processus déterminant le nombre de clients dans l'orbite dans la file M/M/1 avec rappels. L'analyse de ce processus dans un trafic intense a mené à une approximation par un processus de diffusion. Afin de réaliser ces études d'approximation, nous avons considéré une distribution exponentielle pour les temps de rappels. Cette hypothèse est nécessaire

pour la conservation de la propriété de Markov. Une étude numérique a également été réalisée afin de montrer l'efficacité des résultats obtenus.

La suite de cette thèse est organisée comme suit. Dans le premier chapitre, nous présentons les processus stochastiques auxquels nous avons recours dans cette étude. Le deuxième chapitre est consacré à l'étude des files d'attente. Nous présentons d'abord le formalisme mathématique des files d'attente simples. Ensuite, nous étudions les modèles de files d'attente avec rappels. La modélisation des files d'attente simples ainsi que des files d'attente avec rappels par les PDMPs est établie dans le troisième chapitre. Dans le quatrième chapitre, nous nous orientons vers l'approximation par les processus de diffusion dans les modèles de files d'attente simples et avec rappels. Nous terminons cette thèse par une conclusion.

Chapitre 1

Processus Stochastiques

Un processus stochastique ou processus aléatoire décrit l'évolution de tout phénomène produit par le hasard dans le temps. La théorie est basée sur les variables aléatoires qui interviennent dans le calcul classique des probabilités afin de mesurer l'issue d'une expérience, une épreuve ou un événement aléatoire. Mathématiquement, un processus aléatoire noté $\{X(t), t \in T\}$ est une famille de variables aléatoires désignées par le paramètre t et définies sur un même espace de probabilités (Ω, \mathcal{F}, P) . Pour chaque instant $t \in T \subset \mathbb{R}^+$, la variable aléatoire $X(t)$ représente l'état du processus. Un processus stochastique est toujours défini sur un espace des états noté \mathbb{E} qui représente l'ensemble de toutes les valeurs possibles pour la variable aléatoire $X(t)$. Ainsi, pour définir un processus stochastique, il faut :

- un espace des temps T ;
- un espace des états \mathbb{E} ;
- une famille de variables aléatoires $\{X(t), t \in T\}$ définies sur un même espace des probabilités (Ω, \mathcal{F}, P) et à valeurs dans l'espace des états \mathbb{E} .

En outre, la caractérisation de l'ensemble des temps T et l'espace des états \mathbb{E} permet de déterminer les classes des processus stochastiques.

- Si T est un ensemble infini dénombrable alors le processus $\{X(t), t \in T\}$ est dit à temps discret et il est noté $\{X_n, n \in T\}$.
- Si T est un intervalle ou un ensemble d'intervalles alors le processus $\{X(t), t \in T\}$ est dit à temps continu.
- Si \mathbb{E} est un espace fini ou dénombrable alors le processus $\{X(t), t \in T\}$ est dit à espace des états discret.
- Si \mathbb{E} est un espace continu alors le processus $\{X(t), t \in T\}$ est dit à espace des états continu.

Des relations de dépendance sont aussi présentes entre les variables aléatoires, ce qui permet de distinguer les processus stochastiques de manière plus générale.

- **Processus croissant** : Un processus $\{X(t), t \in T\}$ est dit croissant si

$$X(t_1) < X(t_2), \text{ p.s.} \quad \text{pour tout } 0 < t_1 < t_2.$$

- **Processus à accroissements indépendants** : Un processus $X(t), t \in T$ est dit à accroissements indépendants si pour tout $0 < t_1 < t_2 < t_3$

$$P(X(t_3) - X(t_2) = x, X(t_2) - X(t_1) = x') = P(X(t_3) - X(t_2) = x) \times P(X(t_2) - X(t_1) = x').$$

- **Processus à accroissements stationnaires** : Un processus $\{X(t), t \in T\}$ est dit à accroissements stationnaires si pour tout $0 < t_1 < t_2$

$$P(X(t_2) - X(t_1) = x) = P(X(t_2 - t_1) = x).$$

C'est-à-dire $X(t_2) - X(t_1)$ est égale en loi à $X(t_2 - t_1)$.

- **Processus homogène** : Un processus $\{X(t), t \in T\}$ est dit homogène dans le temps si pour tout $0 < t_1 < t_2$

$$P(X(t_1 + t_2) - X(t_1) = x) = P(X(t_2) = x).$$

C'est-à-dire, la loi de la variable $X(t_1 + t_2) - X(t_1)$ ne dépend que de t_2 .

- **Processus régulier** : Un processus $\{X(t), t \in T\}$ à espace des états continu est dit régulier si toutes ses versions $\{\tilde{X}(t), t \in T\}$ ont des trajectoires $t \rightarrow X(t, \omega) (\omega \in \Omega)$ continues possédant les mêmes propriétés de régularité. L'existence de ces versions est garantie si un tel processus satisfait certaines contraintes sur les moments de ses accroissements qui ont été énoncées dans le théorème 1.1 ci-dessous.

Théorème 1.1 (Théorème de Kolmogorov-Čentsov). *Soit $\{X(t), t \in T\}$ un processus dont l'intervalle des temps $T \subset \mathbb{R}$ et l'espace des états est un espace métrique complet (\mathbb{E}, d) . S'il existe des constantes $a, b, C > 0$ vérifiant pour tout $t_1, t_2 \in T$:*

$$E [d(X(t_1), X(t_2))^a] \leq C |t_1 - t_2|^{1+b}.$$

Alors il existe une variation $\{\tilde{X}(t), t \in T\}$ de $\{X(t), t \in T\}$ qui est un processus continu, c'est-à-dire

- (i) $\{\tilde{X}(t)\}$ est un processus dont les trajectoires sont presque sûrement des fonc-

tions continues ;

(ii) Pour tout $t \geq 0$, $P(X(t) = \tilde{X}(t)) = 1$.

De plus, les trajectoires de $\{\tilde{X}(t)\}$ sont localement höldériennes d'exposant γ , tel que $0 < \gamma < b/a$:

$$d(\tilde{X}(t_1, \omega), \tilde{X}(t_2, \omega)) \leq C_\gamma(\omega) |t_1 - t_2|^\gamma \quad \text{pour tout } t_1, t_2 \in T.$$

Nous nous intéressons dans ce chapitre aux processus stochastiques Markoviens qui sont très présents dans l'étude des files d'attente. Lorsque l'hypothèse markovienne est vérifiée, c'est-à-dire l'état futur d'un système peut être déduit de son passé récent, les processus de Markov sont généralement utilisés pour modéliser l'état du système. Afin de donner la définition formelle de la propriété de Markov, nous présentons les définitions suivantes.

Définition 1.1 (Filtration). *Une filtration est une famille de σ -algèbres $(\mathcal{F}_t)_{t \in T}$ telle que $\mathcal{F}_{t_1} \subset \mathcal{F}_{t_2}$ pour tout $t_1 < t_2$.*

Définition 1.2 (Processus adapté). *Le processus $\{X(t), t \in T\}$ est dit adapté à la filtration \mathcal{F}_t , si pour tout $t \in T$, $X(t)$ est une variable aléatoire \mathcal{F}_t -mesurable.*

Définition 1.3 (Propriété de Markov). *Un processus stochastique $\{X(t), t \in T\}$, défini sur un espace des états \mathbb{E} et adapté à la filtration \mathcal{F}_t , vérifie la propriété de Markov si pour tout $A \in \mathbb{E}$ et $t_1, t_2 \in T$ avec $t_1 < t_2$*

$$P(X(t_2) \in A | \mathcal{F}_{t_1}) = P(X(t_2) \in A | X(t_1)).$$

Dans le cas où \mathbb{E} est un espace discret, cette propriété s'exprime comme suit. Pour les instants $0 \leq t_0 < t_1 < \dots < t_n < t_{n+1}$ et les états $i_0, i_1, \dots, i_{n-1}, i, j \in \mathbb{E}$, nous avons

$$P\{X(t_{n+1}) = j | X(t_0) = i_0, \dots, X(t_n) = i\} = P\{X(t_{n+1}) = j | X(t_n) = i\}.$$

Dans ce cas, le processus est dit sans mémoire comme cette propriété signifie que l'état actuel résume l'historique du processus et il est susceptible d'influencer son état futur.

Nous présentons dans ce chapitre les processus de Markov qui seront principalement utilisés dans cette thèse, notamment le processus de Poisson, les processus de naissance et de mort, les processus de Markov déterministes par morceaux et les processus de diffusion.

1.1 Chaînes de Markov

Lorsque l'espace des états est discret, les processus stochastiques possédant la propriété de Markov sont souvent appelés chaînes de Markov.

1.1.1 Chaînes de Markov à temps discret

Définition 1.4 (Chaîne de Markov à temps discret). *Un processus stochastique $\{X_n, n \in \mathbb{N}\}$ défini sur espace des états \mathbb{E} discret est dit chaîne de Markov à temps discret si pour tout $n \geq 0$ et les états $i_0, i_1, \dots, i_{n-1}, i, j$, nous avons*

$$P(X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \dots, X_1 = i_1, X_0 = i_0) = P(X_{n+1} = j | X_n = i).$$

Propriétés de chaînes de Markov à temps discret

- **Absence de mémoire** : Une chaîne de Markov à temps discret est un processus sans mémoire, c'est-à-dire, l'état du processus à l'instant $n + 1$ ne dépend que de son état précédent à l'instant n et non de ses états aux temps antérieurs.
- **Probabilités de transition en une étape** : La probabilité de transition d'un état initial i à l'état final j entre les instants n et $n + 1$ est donnée par :

$$p_{ij}(n) = P(X_{n+1} = j | X_n = i) \quad \text{pour tout } i, j \in \mathbb{E}.$$

- **Homogénéité dans le temps** : Une chaîne de Markov est dite homogène dans le temps si les probabilités de transition en une étape ne dépendent pas du temps n :

$$p_{ij}(n) = P(X_{n+1} = j | X_n = i) = P(X_1 = j | X_0 = i) = p_{ij} \quad \text{pour tout } n \geq 0.$$

- **Matrice de transition** : Les probabilités de transition en une étape p_{ij} , $i, j \in \mathbb{E}$ constituent une matrice carrée dite matrice de transition qui a la forme suivante

$$\mathbf{P} = \begin{pmatrix} p_{00} & p_{01} & \cdots & p_{0j} & \cdots \\ p_{10} & p_{11} & \cdots & p_{1j} & \cdots \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ p_{i0} & p_{i1} & \cdots & p_{ij} & \cdots \\ \vdots & \vdots & \ddots & \vdots & \vdots \end{pmatrix}.$$

La matrice \mathbf{P} est une matrice stochastique, c'est-à-dire, $\sum_{j \geq 0} p_{ij} = 1, i \in \mathbb{E}$.

- **Distribution d'une chaîne de Markov :** La chaîne de Markov $\{X_n, n \in \mathbb{N}\}$ est complètement déterminée à partir de sa loi de probabilité

$$\pi(n) = \pi(0) \times \mathbf{P}^{(n)},$$

où $\pi(n) = \{\pi_i(n), i \in \mathbb{E}\} = \{P(X_n = i), i \in \mathbb{E}\}$, $\mathbf{p}(0) = \{P(X_0 = i), i \in \mathbb{E}\}$ est la loi initiale de la chaîne et $\mathbf{P}^{(n)}$ représente la matrice de transition en n étapes :

$$\mathbf{P}^{(n)} = \begin{pmatrix} p_{00}^{(n)} & p_{01}^{(n)} & \cdots & p_{0j}^{(n)} & \cdots \\ p_{10}^{(n)} & p_{11}^{(n)} & \cdots & p_{1j}^{(n)} & \cdots \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ p_{i0}^{(n)} & p_{i1}^{(n)} & \cdots & p_{ij}^{(n)} & \cdots \\ \vdots & \vdots & \ddots & \vdots & \vdots \end{pmatrix}$$

pour tout $i, j \in \mathbb{E}$ avec $p_{ij}^{(n)} = P(X_{m+n} = j | X_m = i) = P(X_n = j | X_0 = i)$ pour tout $m, n \geq 0$ est la probabilité de passage de l'état i à l'état j en n étapes pour.

- **Équations de Chapman Kolmogorov :** Pour $n \geq 0, m \geq 0$ et $i, j \in \mathbb{E}$, nous avons

$$p_{ij}^{(n+m)} = \sum_{k \in \mathbb{E}} p_{ik}^{(n)} p_{kj}^{(m)}.$$

L'écriture matricielle est donnée par

$$\mathbf{P}^{(n+m)} = \mathbf{P}^{(n)} \mathbf{P}^{(m)}.$$

Ce qui permet de calculer la matrice des probabilités de n passages $\mathbf{P}^{(n)}$ comme suit :

$$\mathbf{P}^{(n)} = \mathbf{P}^{(n-1)} \mathbf{P} = \mathbf{P}^{n-1} \mathbf{P} = \mathbf{P}^n.$$

Ainsi, \mathbf{P}^n est la n^{ime} puissance de la matrice de transition en une seule étape \mathbf{P} .

- **Classification des états :** Un état peut être
 - **Accessible :** Un état j est dit accessible à partir de l'état i s'il existe $n \in \mathbb{N}$ tel que $p_{ij}^{(n)} > 0$.
 - **Communicant :** Deux états i et j communiquent s'ils sont accessibles l'un à l'autre. La relation de communication entre deux états est notée $i \longleftrightarrow j$ et est

une relation d'équivalence, c'est-à-dire, elle est : réflexive ($i \longleftrightarrow i$), symétrique (si $i \longleftrightarrow j$ alors $j \longleftrightarrow i$) et transitive (si $i \longleftrightarrow k$ et $k \longleftrightarrow j$ alors $i \longleftrightarrow j$). L'espace des états \mathbb{E} est souvent partagé en plusieurs classes d'équivalence selon les relations qui existent entre les états. Ainsi, si deux états appartiennent à la même classe alors ils sont dits équivalents entre eux. Sinon, si ils appartiennent à deux classes différentes, ils sont dits non équivalents.

- **Périodique** : Soit $d(i) = PGCD\{n \geq 1 : p_{ii}^{(n)} > 0\}$ la période de l'état i . Un état i est dit apériodique si $d(i) = 1$, sinon il est dit périodique. De plus, si $i \longleftrightarrow j$ alors $d(i) = d(j)$.
- **Récurrent** : Un état i est dit récurrent si

$$\sum_{n \geq 0} p_{ii}^{(n)} = \infty.$$

Cette expression est équivalente à

$$\sum_{n \geq 1} f_{ii}^{(n)} = 1$$

avec $f_{ij}^{(n)} = P(X_n = j, X_{n-1} \neq j, X_{n-2} \neq j, \dots, X_1 \neq j | X_0 = i)$ est la probabilité de la première visite de l'état j partant de l'état i après n transitions. D'où, $f_{ii}^{(n)}$ est dite la probabilité de premier retour à l'état i en n étapes.

Soit $m_i = \sum_{n \geq 1} n f_{ii}^{(n)}$ le temps moyen de retour à l'état i . Alors, un état récurrent est dit de récurrence positive si $m_i < \infty$ et il est dit de récurrence nulle si $m_i = \infty$.

- **Transitoire** : Si

$$\sum_{n \geq 0} p_{ii}^{(n)} < \infty,$$

alors l'état i est dit transitoire. C'est-à-dire, il est non récurrent.

- **Ergodique** : Un état est dit ergodique s'il est apériodique et de récurrence positive.
- **Absorbant** : Si $p_{ii} = 1$ alors l'état i est absorbant. En d'autres termes, un état est dit absorbant si la chaîne de Markov ne peut pas quitter cet état une fois qu'elle y est rentrée.

Notons que dans une classe d'équivalence, si un état possède une propriété (récurrent, transitoire, périodique, ...), alors tous les états appartenant à la même classe possèdent aussi cette propriété.

- **chaîne de Markov irréductible** : Une chaîne de Markov est dite irréductible si elle est défini sur un espace des états \mathbb{E} qui forme qu'une seule classe d'équivalence. Autrement dit, si tous les états communiquent.
- **chaîne de Markov absorbante** : Si la chaîne de Markov comprend au moins un état absorbant qui soit accessible de n'importe quel autre état alors la chaîne est dite absorbante.
- **Distribution des états d'une chaîne de Markov homogène en régime permanent** : En pratique, le régime permanent est généralement atteint après un nombre fini de transitions.
 - **Distribution limite** : À l'équilibre ($n \rightarrow \infty$), une distribution limite des états $\pi = \{\pi_j = \lim_{n \rightarrow \infty} p_{ij}^{(n)}, j \in \mathbb{E}\}$ existe si la probabilité que la chaîne de Markov atteigne l'état j est indépendante de son état initial i .
 - **Distribution stationnaire** : Si la distribution limite d'une chaîne de Markov à temps discret existe, alors elle est stationnaire et est déterminée par

$$\pi = \pi \mathbf{P}^n.$$

Le vecteur des probabilités limites π est la solution de ce système sous la condition $\sum_{j \in \mathbb{E}} \pi_j = 1$. Cette distribution est unique et est donnée par :

$$\pi_j = \begin{cases} \frac{1}{m_j} & \text{si } j \text{ est de récurrence positif;} \\ 0 & \text{si } j \text{ est de récurrence nulle ou transitoire.} \end{cases}$$

Pour tout $j \in \mathbb{E}$, la probabilité π_j peut être interprétée comme la proportion du temps que la chaîne de Markov a passé dans l'état j .

- **Chaîne ergodique** : Une chaîne de Markov à temps discret est dite ergodique si $\pi_j > 0$ pour tout $j \in \mathbb{E}$.

1.1.2 Chaînes de Markov à temps continu

Définition 1.5 (Chaîne de Markov à temps continu). *Un processus stochastique $\{X(t), t \geq 0\}$ défini sur un espace des états \mathbb{E} au plus dénombrable est dit Chaîne de Markov à temps continu si pour tout $t', s \geq 0$ et les états $i, j, x(s) \in \mathbb{E}$, on a*

$$P(X(t+t') = j | X(t) = i, X(s) = x(s); 0 \leq s \leq t) = P(X(t+t') = j | X(t) = i).$$

Propriétés de chaînes de Markov à temps continu

- **Absence de mémoire** : L'état futur $X(t+t')$ de la Chaîne ne dépend que de son état présent $X(t)$ et non de ses états antérieurs $X(s), 0 \leq s \leq t$.
- **Probabilités de transition** : La probabilité de transition de l'état initial i vers un état j entre les instants t et t' est donnée par

$$p_{ij}(t, t') = P(X(t+t') = j | X(t) = i); \quad t, t' \geq 0.$$

- **Homogénéité dans le temps** : Une chaîne de Markov à temps continu est homogène dans le temps si

$$P(X(t+t') = j | X(t) = i) = P(X(t) = j | X(0) = i); \quad 0 \leq t \leq t'.$$

- **Matrice de transition** : La matrice contenant les probabilités de transition $p_{ij}(t, t')$ est donnée par

$$\mathbf{P}(t, t') = \begin{pmatrix} p_{00}(t, t') & p_{01}(t, t') & \cdots & p_{0j}(t, t') & \cdots \\ p_{10}(t, t') & p_{11}(t, t') & \cdots & p_{1j}(t, t') & \cdots \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ p_{i0}(t, t') & p_{i1}(t, t') & \cdots & p_{ij}(t, t') & \cdots \\ \vdots & \vdots & \ddots & \vdots & \vdots \end{pmatrix}.$$

Avec $\sum_{j \in \mathbb{E}} p_{ij}(t, t') = 1$ pour tout $i \in \mathbb{E}$.

- **Équations de Chapman Kolmogorov** : Pour tout $t, t' \in \mathbb{R}_+$, on a

$$p_{ij}(t+t') = \sum_{k \in \mathbb{E}} p_{ik}(t) p_{kj}(t').$$

L'écriture matricielle est

$$\mathbf{P}(t+t') = \mathbf{P}(t)\mathbf{P}(t').$$

- **Temps de séjour dans un état :** Soit T_i la variable aléatoire désignant le temps de séjour de la chaîne de Markov $\{X(t), t \geq 0\}$ dans un état $i \in \mathbb{E} \subseteq \mathbb{N}$. Autrement dit, T_i est le temps passé par la chaîne de Markov $\{X(t), t \geq 0\}$ dans l'état i avant de faire une transition vers l'état j ($j \neq i$). Comme la chaîne possède la propriété de Markov alors la variable T_i vérifie pour tout $i \in \mathbb{E}$ et $t, t' \geq 0$

$$P(T_i > t + t' | T_i > t) = P(T_i > t').$$

Cela signifie que T_i possède la propriété d'absence de mémoire associée à la loi exponentielle. Ainsi, la variable aléatoire T_i suit une loi exponentielle de paramètre $\lambda_i > 0$. En outre, la loi exponentielle sert à modéliser des événements sans mémoires caractérisés comme suit. Sachant qu'un phénomène a duré déjà t unités de temps, la probabilité qu'il dure au moins $t + t'$ unités de temps est la même que la probabilité de durer au moins t' unités de temps à partir du moment initial qu'il a commencé de se produire.

- **Probabilités de transition instantanées :** La propriété de Markov entraîne que l'état suivant j que la chaîne de Markov visitera à partir de l'état i est indépendant de la variable aléatoire T_i . De plus, cette transition se produit dans une période de temps très petite. Ainsi, la chaîne de Markov à temps continu passe d'un état i vers un autre état j avec une probabilité instantanée notée p_{ij} . Les probabilités p_{ij} constituent une matrice de probabilités instantanées notée $\mathbf{P} = (p_{ij})_{i,j \in \mathbb{E}}$ telle que $p_{ii} = 0$ et $\sum_{j \in \mathbb{E}} p_{ij} = 1$.
- **Taux de transition :** Pour tout $i \neq j$, les quantités $q_{ij} = \lambda_i p_{ij}$ sont appelées les taux de transition de la chaîne de Markov à temps continu.
- **Générateur infinitésimal :** Les taux de transition q_{ij} constituent une matrice dite le générateur infinitésimal de la chaîne de Markov à temps continu. Cette matrice est notée $\mathbf{Q} = (q_{ij})_{i,j \in \mathbb{E}}$ avec $\sum_{j \in \mathbb{E}} q_{ij} = 0$ pour tout $i \in \mathbb{E}$. D'où,

$$q_{ii} = - \sum_{j \neq i} q_{ij} = -\lambda_i \sum_{j \neq i} p_{ij} = -\lambda_i.$$

Ainsi, la matrice \mathbf{Q} est représentée comme suit.

$$\mathbf{Q} = \begin{pmatrix} -\lambda_0 & \lambda_0 p_{01} & \cdots & \lambda_0 p_{0i} & \cdots \\ \lambda_1 p_{10} & -\lambda_1 & \cdots & \lambda_1 p_{1i} & \cdots \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \lambda_i p_{i0} & \lambda_i p_{i1} & \cdots & -\lambda_i & \cdots \\ \vdots & \vdots & \ddots & \vdots & \vdots \end{pmatrix}.$$

1.2 Processus de naissance et de mort

Les processus de naissance et de mort sont fréquemment rencontrés dans l'étude des files d'attente. Cette classe des processus possède la propriété de Markov et se caractérise par l'apparition et la disparition, autrement dit, la naissance et la mort d'un individu au sein d'une population en suivant une certaine loi de probabilité.

Définition 1.6 (Processus de naissance et de mort). *Un processus de naissance et de mort est un processus $\{X(t), t > 0\}$ de Markov à temps continu, homogène et défini sur un espace des états discret $\mathbb{E} \subset \mathbb{N}$. À partir d'un état $n \in \mathbb{E}$, les seules transitions possibles pour $\{X(t), t > 0\}$ sont :*

- $n \rightarrow n + 1$: une seule naissance avec un taux moyen $\lambda_n > 0$ pour tout $n \geq 0$;
- $n \rightarrow n - 1$: une seule mort avec un taux moyen $\mu_n > 0$ pour tout $n \geq 1$ avec $\mu_0 = 0$.

Propriétés de Processus de naissance et de mort

- **Probabilités de transition** : Pour un processus de naissance et de mort, un seul évènement (soit une naissance ou une mort) peut se produire dans un intervalle de temps infinitésimal $[t, t + h]$. Par conséquent, les probabilités de transition d'une seule étape $p_{i,j}(h) = P(X(t+h) = j | X(t) = i)$ vérifient les conditions suivantes :
 - (i) $p_{n,n+1}(h) = \lambda_n h + o(h)$ pour tout $n \geq 0$;
 - (ii) $p_{n,n-1}(h) = \mu_n h + o(h)$ pour tout $n \geq 1$;
 - (iii) $p_{n,n}(h) = 1 - (\lambda_n + \mu_n)h + o(h)$ pour tout $n \geq 0$ avec $\mu_0 = 0$.
- **Matrice de transition** : La matrice contenant les probabilités de transition dans

un intervalle de temps infinitésimal $p_{ij}(h)$ est donnée par

$$\mathbf{P}(h) = \begin{pmatrix} 1 - \lambda_0 h & \lambda_0 h & 0 & \cdots & \cdots & \cdots \\ \mu_1 h & 1 - (\lambda_1 + \mu_1) h & \lambda_1 h & 0 & \cdots & \cdots \\ 0 & \mu_2 h & 1 - (\lambda_2 + \mu_2) h & \lambda_2 h & 0 & \cdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots \end{pmatrix}.$$

- **Générateur infinitésimal** : Les taux de naissance et de mort constituent le générateur infinitésimal suivant :

$$\mathbf{Q} = \begin{pmatrix} -\lambda_0 & \lambda_0 & 0 & \cdots & \cdots & \cdots \\ \mu_1 & -(\lambda_1 + \mu_1) & \lambda_1 & 0 & \cdots & \cdots \\ 0 & \mu_2 & -(\lambda_2 + \mu_2) & \lambda_2 & 0 & \cdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots \end{pmatrix}.$$

- **Équations de Chapman Kolmogorov** : Soit $p_n(t) = P(X(t) = n)$ la probabilité que le processus de naissance et de mort soit dans l'état n au temps t . D'après les probabilités de transition, les équations de Chapman-Kolmogorov s'écrivent comme suit :

$$\begin{cases} p_0(t+h) = (1 - \lambda_0 h)p_0(t) + \mu_1 h p_1(t) + o(h); \\ p_n(t+h) = \lambda_{n-1} h p_{n-1}(t) + (1 - (\lambda_n + \mu_n) h)p_n(t) + \mu_{n+1} h p_{n+1}(t) + o(h) \text{ pour tout } n \geq 1. \end{cases} \quad (1.1)$$

- **Distribution stationnaire** : Soit $p_n = \lim_{t \rightarrow \infty} p_n(t)$ la probabilité que le processus de naissance et de mort soit dans l'état n à l'équilibre. Alors,

$$p_n = \prod_{i=1}^n \frac{\lambda_{i-1}}{\mu_i} p_0; \quad (1.2)$$

avec $p_0 = \left(1 + \sum_{n=1}^{\infty} \left(\prod_{i=1}^n \frac{\lambda_{i-1}}{\mu_i}\right)\right)^{-1}$ si la série $\sum_{n=1}^{\infty} \left(\prod_{i=1}^n \frac{\lambda_{i-1}}{\mu_i}\right)$ est convergente.

Les probabilités stationnaires p_n sont calculées en passant à la limite ($t \rightarrow \infty$) dans le système des équations de Chapman-Kolmogorov Eq. (1.1) et en utilisant la propriété $\sum_{n=0}^{\infty} p_n = 1$ pour déterminer la probabilité p_0 .

1.3 Processus de Poisson

Le processus de Poisson est un cas particulier des processus de comptage dits aussi les processus ponctuels (Voir Cox et Isham [14]). Ce processus est beaucoup utilisé pour modéliser l'arrivée des clients dans un système d'attente.

Définition 1.7 (Processus de comptage). *Un processus $\{N(t), t \geq 0\}$ à espace des états discret est dit un processus de comptage si :*

1. $N(0) = 0$;
2. Pour tout $t < t'$, $N(t) \leq N(t')$;
3. Pour tout $t < t'$, $N(t') - N(t)$ représente le nombre d'occurrence d'un certain évènement dans l'intervalle $]t, t']$.

Définition 1.8 (Processus de Poisson). *Un processus aléatoire $\{A(t), t \geq 0\}$ à espace des états discret est dit un processus de Poisson de taux λ ($\lambda > 0$) si :*

1. $\{A(t), t \geq 0\}$ est un processus de comptage ;
2. $\{A(t), t \geq 0\}$ est un processus à accroissements indépendants ;
3. $\{A(t), t \geq 0\}$ est un processus à accroissements stationnaires ;
4. Pour tout $t \geq 0$, la variable aléatoire $A(t)$ suit la loi de Poisson de paramètre λt :

$$P(A(t) = n) = \frac{(\lambda t)^n}{n!} e^{-\lambda t} \quad \text{pour tout } t \geq 0, n \in \mathbb{N}.$$

Propriétés du processus de Poisson

- **Loi des inter-occurrences :** Soit $S_n = T_n - T_{n-1}$ ($n \geq 0$) le temps séparant les deux évènements qui se sont produits aux instants T_{n-1} et T_n selon un processus de Poisson de paramètre λ avec $T_0 = 0$. Les variables aléatoires S_n sont indépendantes et suivent une loi exponentielle de paramètre λ .
- **Probabilités infinitésimales :** Pour tout $t \geq 0$, un processus de Poisson $\{A(t), t \geq 0\}$ d'intensité λ est caractérisé par les probabilités suivantes :
 - (i) La probabilité qu'il y ait une seule occurrence dans un petit intervalle de temps $[t, t + h]$ est proportionnelle à la longueur de cet intervalle h et le coefficient de proportionnalité est λ :

$$P(A(t+h) - A(t) = 1) = \lambda h + o(h).$$

- (ii) La probabilité qu'il y ait plus d'une occurrence dans un petit intervalle est négligeable :

$$P(A(t+h) - A(t) > 1) = o(h).$$

- (iii) La probabilité qu'il y ait aucune occurrence dans un petit intervalle de temps est déduite des deux dernières égalités :

$$P(A(t+h) - A(t) = 0) = 1 - \lambda h + o(h).$$

Notons que le processus de Poisson est le plus simple des processus de naissance et de mort, car il s'agit d'un processus de naissance pur.

Processus de Poisson périodique

Le processus de Poisson est caractérisé par un taux d'intensité λ constant. Toutefois, ce taux peut être une fonction dépendante du temps t . Dans ce cas, nous obtenons un modèle général dépendant du temps appelé processus de Poisson non-homogène (PPNH) avec une fonction d'intensité $\lambda(t) \geq 0$, pour tout $t \geq 0$ (Garrido et Lu [30]).

D'après Harrison et Lemoine [38], la définition suivante donne la forme la plus simple d'un processus de Poisson périodique .

Définition 1.9 (Processus de Poisson périodique). *Soit $\{A(t), t \geq 0\}$ un PPNH avec une fonction d'intensité $\lambda(t)$ bornée telle que $\lambda(n+t) = \lambda(t)$ pour tout $t \in [0, 1[$ et $n \in \mathbb{N}^*$. Sans perte de généralité, la période est supposée égale à 1. L'unité de temps est considéré être le jour et un point $t \in [0, 1[$ représente généralement un moment de la journée. Soit*

$$\Lambda = \int_0^1 \lambda(t) dt$$

le taux d'intensité cumulé du processus.

Le processus $A(t)$ est dit un processus de Poisson périodique avec une période d'un jour et un taux d'occurrence moyen Λ .

1.4 Processus de renouvellement

Un processus de renouvellement, noté $\{N(t), t \geq 0\}$, est un processus stochastique à temps continu défini sur un espace des états discret à valeurs positives. Ce processus dénombre les occurrences d'un certain évènement lorsque les temps d'inter-occurrences sont des variables aléatoires positives indépendantes et identiquement distribuées (i.i.d). L'analyse de remplacements des composants fournit un champs d'application important pour les processus de renouvellement. Dans un ensemble de composants, X_i représente la durée de vie du i ème composant. Le premier composant est mis en place au temps 0. Ce composant reste fonctionnel pendant X_1 unités de temps et est donc remplacé à l'instant

X_1 . Le deuxième composant assure une durée de fonctionnement de X_2 unités de temps et est ensuite renouvelé par un troisième composant au temps $X_1 + X_2$, etc. Soit S_n le moment où le n ème composant est renouvelé, alors

$$S_n = \sum_{i=1}^n X_i;$$

$$S_0 = 0.$$

Avec $(X_i)_{i \geq 1}$ est une suite de variables aléatoires positives et i.i.d de la même fonction de répartition F_X .

Définition 1.10 (Processus de renouvellement). *Un processus de comptage $\{N(t), t \geq 0\}$ décrivant le nombre de remplacements effectués dans un intervalle de temps $]0, t]$ avec*

$$N(t) = \max\{n \geq 0 : S_n \leq t\};$$

$$N(0) = 0.$$

est un processus de renouvellement.

Notons que, lorsque les variables aléatoires X_i sont exponentiellement distribuées, le processus de renouvellement est un processus de Poisson.

Propriétés des processus de renouvellement

- Taux d'un processus de renouvellement $\lambda > 0$ est donnée par

$$\lambda = \frac{1}{E(X_1)}.$$

- Pour tout entier $n \geq 0$, nous avons

$$P(N(t) \geq n) = P(S_n \leq t).$$

- **Fonction de renouvellement :** Un processus de renouvellement peut être caractérisé par sa fonction de renouvellement donnée par

$$\begin{aligned} m(t) &= E(N(t)) \\ &= \sum_{n \geq 1} P(N(t) \geq n) \\ &= \sum_{n \geq 1} P(S_n \leq t). \end{aligned}$$

Si $m(t) = \lambda t$, alors $N(t)$ est un processus de Poisson.

- **Durée de vie résiduelle moyenne :** Soit X la variable aléatoire qui représente la durée de vie d'une certaine entité. Étant donné que cette entité est âgée de t unités de temps, sa durée de vie restante après le temps t est aléatoire. L'espérance de cette durée de vie résiduelle est appelée durée de vie résiduelle moyenne. D'après Guess et al. [34], la durée de vie résiduelle moyenne est définie comme suit.

Définition 1.11 (Durée de vie résiduelle moyenne). *Soit F_X la fonction de répartition de la variable aléatoire X possédant un moment d'ordre 1 fini. Pour tout $t \geq 0$, la durée de vie résiduelle moyenne est donnée par*

$$\mu_X(t) = \begin{cases} E(X - t | X > t) & \text{si } \bar{F}(t) > 0; \\ 0 & \text{si } \bar{F}(t) = 0. \end{cases}$$

Où $\bar{F}(t) = 1 - F(t)$.

Notons que $\mu_X(t)$ peut être considérée comme une fonction de la durée de vie résiduelle moyenne comme elle est définie pour chaque instant $t \geq 0$. Lorsque $\bar{F}(t) > 0$, cette fonction peut s'écrire

$$\mu_X(t) = \frac{1}{\bar{F}(t)} \int_0^{\infty} \bar{F}(x+t) dx$$

ou encore

$$\mu_X(t) = \frac{1}{\bar{F}(t)} \int_t^{\infty} \bar{F}(u) du.$$

Notons que, dans le cas particulier où la variable aléatoire X suit la loi exponentielle de paramètre λ , nous avons

$$\begin{aligned} \mu_X(t) &= \frac{1}{e^{-\lambda t}} \int_t^{\infty} e^{-\lambda u} du \\ &= \frac{1}{\lambda} \\ &= E(X). \end{aligned}$$

Ce résultat découle de la propriété d'absence de mémoire de la loi exponentielle.

- **Comportement asymptotique :** À long terme, un processus de renouvellement se comporte comme un processus de Poisson. Ce résultat est connu sous le nom du théorème de renouvellement élémentaire.

Théorème 1.2 (Théorème élémentaire de renouvellement). *Pour un processus de renouvellement $\{N(t), t \geq 0\}$ avec des temps des inter-occurrences de moyenne*

$E(X_1)$ et de variance $Var(X_1)$, nous avons

$$\begin{aligned} E[N(t)] &\approx \frac{t}{E(X_1)} = \lambda t, \\ Var[N(t)] &\approx \frac{t Var(X_1)}{[E(X_1)]^2} = Var(X_1)\lambda^2 t \end{aligned} \tag{1.3}$$

pour de grandes valeurs de t .

1.5 Martingales

Une martingale, en mathématique, est un processus stochastique qui modélise les gains ou les pertes accumulés par un joueur au cours d'un jeu équitable. La définition mathématique d'une martingale est donnée comme suit.

Définition 1.12 (Martingale). *Un processus $\{M(t), t \in \mathbb{R}_+\}$ est une martingale par rapport à la filtration (\mathcal{F}_t) si*

1. *La variable $M(t)$ est intégrable pour tout $t \geq 0 : E(|M(t)|) < \infty ;$*
2. *$\{M(t), t \in \mathbb{R}_+\}$ est adapté à la filtration (\mathcal{F}_t) ;*
3. *$E(M(t)|\mathcal{F}_s) = M(s)$ presque sûrement pour tout $s \leq t$.*

Si la condition 3. est remplacée par $E(M(t)|\mathcal{F}_s) \leq M(s)$ alors $\{M(t), t \in \mathbb{R}_+\}$ est une sur-martingale, et si elle est remplacée par $E(M(t)|\mathcal{F}_s) \geq M(s)$ alors $\{M(t), t \in \mathbb{R}_+\}$ est une sous-martingale.

Propriétés des martingales

- $E(E(M(t)|\mathcal{F}_s)) = E(M(s))$ pour tout $s \leq t$.
- $E(M(t)) = E(M(0))$ pour tout $t \geq 0$.
- $E(M(t))$ est monotone croissante (décroissante) pour une sous (sur)-martingale.

Caractérisation des processus de Markov par des martingales

Pour un processus de Markov $\{X(t), t \geq 0\}$ défini sur l'espace des états \mathbb{E} , le générateur infinitésimal \mathcal{G} défini sur son domaine $\mathfrak{D}(\mathcal{G})$ est donné par

$$\mathcal{G}f = \lim_{t \rightarrow 0} \frac{P_t f - f}{t}, \quad \text{pour toute } f \in \mathfrak{D}(\mathcal{G}) \tag{1.4}$$

tel que pour tout $t > 0$ et $x \in \mathbb{E}$

$$\begin{aligned} P_t : \quad B(\mathbb{E}) &\longrightarrow B(\mathbb{E}) \\ P_t f(x) &\longrightarrow E_x(f(X(t))) = E(f(X(t)) | X(0) = x), \end{aligned}$$

où $B(\mathbb{E})$ est l'ensemble de toutes les fonctions $f : \mathbb{E} \longrightarrow \mathbb{R}$ mesurables et bornées avec $\|f\| = \sup_{x \in \mathbb{E}} |f(x)|$.

Le domaine $\mathfrak{D}(\mathcal{G})$ se constitue de toutes les fonctions f pour lesquelles $\mathcal{G}f$ existe.

Pour toute fonction $f \in \mathfrak{D}(\mathcal{G})$ et pour tout $t > 0$, nous avons

$$P_t f(x) = f(x) + \int_0^t \mathcal{G}P_s f(x) ds. \quad (1.5)$$

Ce résultat est connu sous le nom de la formule de Dynkin.

En utilisant l'espérance conditionnelle, cette formule s'écrit

$$E_x \left(f(X(t)) - f(x) - \int_0^t \mathcal{G}f(X(s)) ds \right) = 0.$$

Ainsi, le processus

$$M^f(t) = f(X(t)) - f(x) - \int_0^t \mathcal{G}f(X(s)) ds \quad (1.6)$$

est une martingale par rapport à $\mathcal{F}_t^X = \sigma(X(s), s \leq t)$ pour tout $x \in \mathbb{E}$.

La caractérisation des processus de Markov par les martingales est donnée par le théorème suivant.

Théorème 1.3 (Koroliuk et Limnios [56]). *Soit $\{X(t), t \geq 0\}$ un processus stochastique défini sur un espace des états \mathbb{E} et adapté à la filtration $(\mathcal{F}_t^X)_{t \geq 0}$. Soit \mathcal{G} son générateur infinitésimal défini sur son domaine $\mathfrak{D}(\mathcal{G})$. Étant donné que le processus donné par Eq. (1.6) est une martingale par rapport à \mathcal{F}_t^X pour toute fonction $f \in \mathfrak{D}(\mathcal{G})$, alors $\{X(t), t \geq 0\}$ est un processus de Markov généré par le générateur \mathcal{G} .*

D'après Koroliuk et Limnios [56], le processus $\{X(t), t \geq 0\}$ est dit la solution du problème de martingale du générateur \mathcal{G} .

1.6 Processus de Markov déterministes par morceaux

Les processus de Markov déterministes par morceaux, dits en anglais *piecewise deterministic Markov processes* (PDMP), ont été introduits par Davis [16] comme étant une famille générale des modèles stochastiques de non-diffusion. Ces processus de Markov sont caractérisés par le fait qu'entre deux sauts, le processus ne reste pas constant, mais évolue de manière déterministe. Ainsi, un PDMP se compose d'une combinaison de mouvements déterministes et de sauts aléatoires. La classe des PDMPs fournit un cadre pour l'étude des problèmes d'optimisation, notamment dans les systèmes de files d'attente [6, 17]. Comme l'a montré Davis [16] et ensuite indiqué en détail dans Davis [17], un PDMP peut être déterminé explicitement à l'aide de trois paramètres ϕ, λ, Q . Soit $\{X(t)\}$ un PDMP défini sur un espace des états \mathbb{E} qui est structuré comme suit. Soit K un ensemble dénombrable et $d : K \rightarrow \mathbb{N}$ une fonction donnée. Pour chaque $v \in K$, M_v est un ensemble ouvert de $\mathbb{R}^{d(v)}$. Alors, $\mathbb{E} = \cup_{v \in K} v \times M_v = \{(v, \xi) : v \in K, \xi \in M_v\}$. On dénote par \mathcal{E} la σ -algèbre générée par les sous-ensembles de Borel de \mathbb{E} . L'état du processus sera noté $X(t) = (v(t), \xi(t))$ et les caractéristiques ϕ, λ et Q sont définies comme suit :

- La fonction de flux $\phi_v : (\mathbb{R}_+ \times M_v) \rightarrow \mathbb{E}$.
- Le taux de sauts $\lambda : \mathbb{E} \rightarrow \mathbb{R}_+$. On suppose que cette fonction est mesurable et pour tout $x = (v, \xi) \in \mathbb{E}$, il existe $\varepsilon > 0$ tel que $\int_0^\varepsilon \lambda(\phi_v(t, \xi)) dt$ existe.
- $Q : (\mathbb{E} \cup \partial^* \mathbb{E}) \times \mathcal{E} \rightarrow [0, 1]$, avec $\partial^* \mathbb{E} = \cup_{v \in K} \partial^* M_v$ où $\partial^* M_v = \{\xi' \in \partial M_v : \phi_v(t, \xi) = \xi', \text{ pour tout } (t, \xi) \in \mathbb{R}_+ \times M_v\}$, la mesure de transition spécifiant les localisations post-saut avec $Q(x; \{x\}) = 0$. Notons que ∂M_v est la limite de M_v et $\partial^* M_v$ représente les points limites pour lesquels le flux sort de M_v .

Avec un choix commode de l'espace des états \mathbb{E} et des paramètres \mathfrak{X}, λ et Q il est possible de modéliser presque tous les processus de non-diffusion trouvés dans la littérature. Plusieurs applications importantes ont été présentées dans les travaux de Davis [16, 17].

1.6.1 Chaîne de Markov associée

Le comportement du PDMP dépend des caractéristiques ϕ, λ et Q de la manière suivante. Soit T_1 un temps de sauts de fonction de répartition

$$P_x(T_1 > t) = \begin{cases} \exp\{-\int_0^t \lambda(\phi_v(s, \xi)) ds\} & \text{si } t < t_*(x); \\ 0 & \text{si } t \geq t_*(x). \end{cases}$$

La variable déterministe $t_*(x) = \inf\{t \in \mathbb{R}_+ : \phi_v(t, \xi) \in \partial^* \mathbb{E}\}$ représente le temps jusqu'à ce que l'ensemble $\partial^* \mathbb{E}$ soit atteint à partir d'un état $x \in \mathbb{E}$. Soit Z_1 une variable aléatoire

ayant la distribution $Q(\phi_v(T_1, \xi); \cdot)$ pour tout $x \in \mathbb{E}$. En démarant d'un point $x = (v, \xi) \in \mathbb{E}$, la trajectoire de $X(t)$ pour $t \leq T_1$ est donnée par

$$X(t) = \begin{cases} (v, \phi_v(t, \xi)) & \text{for } t < T_1; \\ Z_1 & \text{for } t = T_1. \end{cases}$$

En partant de $X(T_1) = Z_1$, nous choisissons maintenant le temps d'inter-saut suivant $T_2 - T_1$ et l'emplacement après le saut $X(T_2) = Z_2$ de manière similaire. On obtient ainsi une trajectoire déterministe par morceaux ($X(t)$) avec des temps de transition T_1, T_2, \dots et des positions après transition Z_1, Z_2, \dots . La séquence de variables aléatoires $\{Z_n\}$ est la chaîne de Markov associée au processus original $\{X(t)\}$, de sorte que les résultats précédents sur les chaînes de Markov à temps discret ont été utilisés pour étudier la stabilité et l'ergodicité des PDMPs, voir [12, 18, 13]. De plus, on suppose qu'il n'existe qu'un nombre fini de sauts de $X(t)$ dans tout intervalle de temps fini (voir l'hypothèse 3.1 dans Davis [16] ou l'hypothèse 24.4 dans Davis [17]). Ainsi, si Z_0 est le point initial, alors la chaîne de Markov associée $\{Z_n, n \geq 0\}$ a pour mesure de transition $\mathcal{P} : \mathbb{E} \cup \partial^* \mathbb{E} \cup \mathcal{E} \rightarrow [0, 1]$ donnée par :

$$\begin{aligned} \mathcal{P}(x; A) &= \int_0^{t_*(x)} Q(\phi_v(s, \xi); A) \lambda(\phi_v(s, \xi)) \exp(-\Lambda(s, x)) ds \\ &\quad + Q(\phi_v(t_*(x), \xi); A) \exp(-\Lambda(t_*(x), x)). \end{aligned}$$

Où $\Lambda(t, x) = \int_0^t \lambda(\phi_v(s, \xi)) ds$ pour tout $x \in \mathbb{E}$ et $0 \leq t \leq t_*(x)$.

Notons que $\Lambda(t_*(x), x) = \infty$ lorsque $t_*(x) = \infty$.

1.6.2 Générateur infinitésimal associé

Le générateur infinitésimal est un outil indispensable dans l'étude des modèles de files d'attente à travers les PDMPs. D'après Davis [16, 17], le générateur \mathcal{G} au point $x \in \mathbb{E}$ s'écrit pour toute fonction $f \in \mathfrak{D}(\mathcal{G})$ sous la forme suivante :

$$\mathcal{G}f(x) = \sum_{k=1}^{d(v)} \frac{\partial}{\partial \xi_k} f(x) + \lambda(x) \int_{\mathbb{E}} (f(x') - f(x)) Q(x, dx').$$

Notons que $\mathfrak{D}(\mathcal{G})$ est le domaine du générateur \mathcal{G} .

Soit $\mathcal{M}(\mathbb{E})$ l'ensemble des fonctions réelles et mesurables. On définit l'ensemble

$$\partial^* \mathcal{M}(\mathbb{E}) = \left\{ f \in \mathcal{M}(\mathbb{E}) : f(x) = \lim_{t \rightarrow 0} f(\phi_v(t, \xi)) \text{ existe pour tout } x = (v, \xi) \in \partial^* \mathbb{E} \right\};$$

et la variable

$$S_*(x) = \inf \{t \in \mathbb{R}_+ : P_x(T_1 > t) = 0\}.$$

Le domaine $\mathfrak{D}(\mathcal{G})$ du générateur \mathcal{G} est constitué des fonctions $f \in \mathcal{M}(\mathbb{E})$ vérifiant :

1. La fonction $t \rightarrow f(\phi_v(t, \xi))$ est absolument continue sur $[0, S_*(x)[$ pour tout $x = (v, \xi) \in \mathbb{E}$.

2. La condition limite

$$f(x) = \int_{\mathbb{E}} f(y) Q(dy; x); \quad x \in \partial^* \mathbb{E}.$$

3. $\mathcal{B}f \in L_1^{loc}(p)$, où dans l'intervalle $]T_k, Tk - 1]$

$$\begin{aligned} \mathcal{B}f(x, s, \omega) &= f(x) - f(X(s^-, \omega)) \\ &= f(x) - f(\phi_v(s, \xi)); \end{aligned}$$

tel que $\mathcal{B}f \in L_1^{loc}(p)$ si pour $n = 1, 2, \dots$

$$\mathbb{E}_x \left(\sum_{T_i \leq n} |f(X(T_k)) - f(X(T_{k-}))| \right) < \infty;$$

avec $X(T_{k-1}) = (v, \xi)$ et $p(t, A) = \sum_{i=1}^{\infty} \mathbf{I}_{(t \geq T_i)} \mathbf{I}_{(X(T_i) \in A)}$ pour tout $A \in \mathbb{E}$.

1.6.3 Martingale associée

Dans cette section, nous allons présenter un résultat essentiel sur les PDMPs.

Proposition 1.1 (Davis [17]). *Le processus*

$$M^f(t) = f(X(t)) - f(X(0)) - \int_0^t \mathcal{G}f(X(s)) ds, \quad (1.7)$$

défini pour toute fonction bornée $f \in \mathfrak{D}(\mathcal{G})$, est une martingale par rapport à la filtration naturelle du PDMP $\{X(t)\}$ donnée par $\mathcal{F}_t = \sigma\{X(s), s \leq t\}$.

Preuve 1. — Pour tout $t \geq 0$, $M^f(t)$ est \mathcal{F}_t -mesurable. D'où, $M^f(t)$ est adaptée à la filtration $(\mathcal{F}_t)_{t \geq 0}$.

— Pour tout $s \leq t$, nous avons d'après Eq. (1.7)

$$E_x(M^f(t) - M^f(s)|\mathcal{F}_s) = E_x(f(X(t))|\mathcal{F}_s) - f(X(s)) - E_x\left(\int_s^t \mathcal{G}f(X(u))du|\mathcal{F}_s\right).$$

D'après Davis [17], la propriété de Markov peut également s'exprimer comme suit.

$$E_x(f(X(s+t))|\mathcal{F}_s) = E_z(f(X(t))|X(s) = z),$$

pour $f \in B(\mathbb{E})$, $x \in \mathbb{E}$ et $s, t \geq 0$.

Ainsi,

$$\begin{aligned} E_x(f(X(t))|\mathcal{F}_s) &= E_z(f(X(t-s))|X(s) = z) \\ &= P_{t-s}f(X(s)), \end{aligned}$$

et

$$\begin{aligned} E_x\left(\int_s^t \mathcal{G}f(X(u))du|\mathcal{F}_s\right) &= \int_s^t E_x(\mathcal{G}f(X(u))|\mathcal{F}_s) du \\ &= \int_s^t P_{u-s}\mathcal{G}f(X(s))du. \end{aligned}$$

D'après la partie 3. de la proposition 14.10 dans Davis [17], nous avons

$$E_x\left(\int_s^t \mathcal{G}f(X(u))du|\mathcal{F}_s\right) = \int_s^t \mathcal{G}P_{u-s}f(X(s))du.$$

En posant le changement de variable $r = u - s$, nous obtenons

$$\int_s^t \mathcal{G}P_{u-s}f(X(s))du = \int_0^{t-s} \mathcal{G}P_r f(X(s))dr.$$

D'où,

$$E_x(M^f(t) - M^f(s)|\mathcal{F}_s) = P_{t-s}f(X(s)) - f(X(s)) - \int_0^{t-s} \mathcal{G}P_r f(X(s))dr.$$

En s'identifiant à Eq. (1.5), nous avons

$$E_x(M^f(t) - M^f(s)|\mathcal{F}_s) = 0,$$

ce qui établit la propriété de martingale.

Dans la littérature, cette propriété est généralement appelée "propriété de martingale" ou "condition de martingale" associée au générateur infinitésimal. Cette propriété signifie que si la fonction utilisée pour calculer le générateur infinitésimal est elle-même une martingale, alors le générateur calculé sera nul [10, 15, 16, 17]. En effet, cela est dû à la définition et aux propriétés des générateurs infinitésimaux. Lorsque la fonction f est une martingale, cela signifie que son espérance conditionnelle future est égale à sa valeur actuelle, compte tenu des informations disponibles jusqu'à ce moment. Mathématiquement, cela se traduit par $E(f(X(t))|\mathcal{F}_t) = f(X(t))$, pour tout $s \geq t$. En conséquence, si la fonction utilisée pour calculer le générateur est déjà une martingale, cela signifie que les termes décrivant les variations futures de la martingale sont déjà équilibrés par l'espérance conditionnelle, ce qui conduit à un générateur nul.

Il convient de noter que toutes les martingales calculées par la suite dans le chapitre 3 sont adaptées à la filtration naturelle du PDMP considéré.

1.7 Processus de diffusion

Un processus de diffusion est un processus de Markov à espace des états continu avec des trajectoires presque sûrement continues. Ce processus modélise mathématiquement le phénomène de diffusion en physique qui est la trajectoire d'une particule immergée dans un fluide en mouvement et subissant des déplacements aléatoires provoqués par des collisions avec d'autres particules. La définition mathématique d'un processus de diffusion unidimensionnelle stipule qu'un processus de Markov $\{X(t), t \geq 0\}$, défini sur un espace des états continu \mathbb{E} et admettant la fonction $p(x', t'|x, t)$ comme la probabilité de transition de l'état x à l'instant t vers l'état x' à l'instant t' , vérifie l'équation différentielle partielle (EDP)

$$\frac{\partial}{\partial t} p(x', t'|x, t) = -\frac{\partial}{\partial x'} a(x', t) p(x', t'|x, t) + \frac{1}{2} \frac{\partial^2}{\partial x'^2} b(x', t) p(x', t'|x, t).$$

Cette équation est dite équation de Fokker-Planck qui est exactement l'équation de Kolmogorov progressive d'un processus Markovien à espace des états continu [29].

Les paramètres infinitésimaux $a(x, t)$ et $b(x, t)$ sont des fonctions continues en x et en t définies comme suit pour tout $h > 0$:

$$a(x, t) = \lim_{h \rightarrow 0} \frac{1}{h} E \{X(t+h) - X(t) | X(t) = x\}; \quad (1.8)$$

$$b(x, t) = \lim_{h \rightarrow 0} \frac{1}{h} \text{Var} \{X(t+h) - X(t) | X(t) = x\}. \quad (1.9)$$

La fonction $a(x, t)$ est appelée le paramètre de dérive et la fonction $b(x, t)$ est dite le paramètre de diffusion ou encore la moyenne et la variance infinitésimales, respectivement.

Le générateur infinitésimal d'un processus de diffusion $X(t)$, agissant sur toute fonction réelle et continue f pour tout $x \in \mathbb{E}$, s'écrit sous la forme suivante :

$$\mathcal{G}f(x) = a(x, t)f'(x) + \frac{1}{2}b(x, t)f''(x).$$

Pour un processus de diffusion homogène dans le temps, le générateur ne dépend pas de t . Cela signifie que les fonctions a et b sont indépendantes de t .

Dans ce travail, nous allons considérer les processus de diffusion comme étant des limites d'autres processus de Markov décrivant des systèmes de files d'attente avec rappel.

1.7.1 Exemples de processus de diffusion

En théorie des probabilités et en statistique, il existe plusieurs processus de diffusion de base qu'on reconnaît d'après la forme particulière de leurs paramètres de dérive et de diffusion.

Mouvement Brownien

Le mouvement Brownien dit aussi processus de Wiener est l'un des processus de diffusion les plus connus qui intervient fréquemment en mathématiques pures et appliquées, en économie, en finance quantitative, en biologie évolutive et en physique. Ce processus possède les propriétés suivantes :

- $W(0) = 0$.
- Le processus $\{W(t)\}$ a des accroissements indépendants et stationnaires.
- Pour tout intervalle de temps $]t, t[$, la variable aléatoire $W(t') - W(t)$ admet la distribution normale $\mathcal{N}(0, \tilde{t})$ où $\tilde{t} = t_2 - t_1$.
- Le processus Brownien vérifie l'équation de Fokker-Planck

$$\frac{\partial}{\partial t} p(x', t' | x, t) = \frac{1}{2} \frac{\partial^2}{\partial x'^2} p(x', t' | x, t).$$

- Si le processus $\{W(t)\}$ vérifie

$$\frac{\partial}{\partial t} p(x', t' | x, t) = -\mu \frac{\partial}{\partial x'} p(x', t' | x, t) + \frac{\sigma^2}{2} \frac{\partial^2}{\partial x'^2} p(x', t' | x, t),$$

alors il est dit processus Brownien avec dérive μ et variance infinitésimale σ^2 .

Mouvement Brownien régularisé

Le mouvement Brownien régularisé est un processus de Wiener défini sur un espace des états \mathbb{E} de dimension finie avec des frontières réfléchissantes c.à.d les frontières $\partial^*\mathbb{E}$ sont sujettes à un mécanisme de refoulement. Dans le domaine de la recherche opérationnelle, \mathbb{E} est généralement non-négatif. Il a été montré que les mouvements Browniens régularisés décrivent les modèles de files d'attente soumis à un trafic intense (voir Harrison [37]). Ils ont été proposés pour la première fois par Kingman [54], puis prouvés par Iglehart et Whitt [?, 42].

Considérons un mouvement Brownien unidimensionnel $\{Z(t)\}$ démarrant de 0 et limité à des valeurs positives (une seule barrière réfléchissante à 0) avec une dérive $a(x, t) = \mu$ et une variance $b(x, t) = \sigma^2$. La distribution de transition correspondante à $Z(t)$ est donnée, pour tout $t \geq 0$, par

$$P(Z(t) \leq z) = \Phi\left(\frac{z - \mu t}{\sigma t^{1/2}}\right) - e^{2\mu z/\sigma^2} \Phi\left(\frac{-z - \mu t}{\sigma t^{1/2}}\right);$$

avec $\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-t^2/2} dt$; $z \in \mathbb{R}$ est la fonction de distribution cumulative de la distribution normale centrée et réduite $\mathcal{N}(0, 1)$.

Si $\mu < 0$, la distribution stationnaire de $Z(t)$ lorsque $t \rightarrow \infty$ coïncide avec la distribution exponentielle de paramètre $-2\mu/\sigma^2$. Ainsi,

$$P(Z \leq z) = 1 - e^{2\mu z/\sigma^2}.$$

Processus d'Ornstein-Uhlenbeck

Le processus d'Ornstein-Uhlenbeck est un processus stationnaire de Gauss-Markov, ce qui signifie que c'est un processus gaussien, un processus de Markov et qu'il est homogène dans le temps. Les applications de ce processus sont principalement apparues en mathématiques financières et dans les sciences physiques. Mathématiquement, un processus d'Ornstein-Uhlenbeck $\{U(t)\}$ vérifie l'équation de Fokker-Planck

$$\frac{\partial}{\partial t} p(u', t' | u, t) = \mu \frac{\partial}{\partial u'} u' p(u', t' | u, t) + \frac{\sigma^2}{2} \frac{\partial^2}{\partial u'^2} p(u', t' | u, t).$$

Ainsi, pour tout $u \in \mathbb{E}$ et $t > 0$

$$a(u, t) = -\mu u \quad \text{et} \quad b(u, t) = \sigma^2.$$

Par ailleurs, le paramètre de dérive du processus d'Ornstein-Uhlenbeck peut être exprimé par une constante supplémentaire θ de sorte que l'équation de Fokker-Planck s'écrit

$$\frac{\partial}{\partial t} p(u', t' | u, t) = \mu \frac{\partial}{\partial u'} (\theta - u') p(u', t' | u, t) + \frac{\sigma^2}{2} \frac{\partial^2}{\partial u'^2} p(u', t' | u, t).$$

1.7.2 Représentation d'Itô d'un processus de diffusion

Un processus de diffusion $\{X(t)\}$ est caractérisé par sa moyenne infinitésimale $a(x, t)$ et sa variance infinitésimale $b(x, t) = \sigma^2(x, t)$ avec $\sigma(x, t)$ est le coefficient de diffusion. Le paramètre de dérive $a(x, t)$ décrit le taux de variation instantané de l'espérance conditionnelle des accroissements et le paramètre de diffusion $b(x, t)$ indique le taux de changement instantané de la covariance conditionnelle des accroissements. Dans le cas où les paramètres $a(x, t)$ et $\sigma(x, t)$ sont continus par rapport à t , satisfont pour tout $x, x' \in \mathbb{E}$, $t \in T (T \subseteq [t_0, \infty[)$ et $C > 0$ une constante la condition de Lipschitz

$$\| a(x, t) - a(x', t) \| + \| \sigma(x, t) - \sigma(x', t) \| \leq C \| x_1 - x_2 \| \quad (1.10)$$

et vérifient pour tout $x \in \mathbb{E}$, $t \in T$ et $D > 0$ une constante la condition de non-explosion

$$\| a(x, t) \|^2 + \| \sigma(x, t) \|^2 \leq D(1 + \| x \|^2), \quad (1.11)$$

le processus $\{X(t)\}$ peut être représenté par l'équation différentielle stochastique (EDS) d'Itô comme suit

$$dX(t) = a(x, t)dt + \sigma(x, t)dW(t), \quad X(t_0) = x_0.$$

Où $\{W(t)\}$ est le processus Brownien.

1.7.3 Approximation et simulation

Les EDS sont généralement compliquées, voire impossibles à résoudre de manière explicite. Néanmoins, il existe plusieurs méthodes numériques permettant d'obtenir des solutions approximatives. La méthode la plus utilisée pour approximer un processus de diffusion $\{X(t), 0 \leq t \leq T\}$ représenté sous forme d'un processus d'Itô est la méthode d'Euler. Cette technique permet d'obtenir un échantillon de trajectoires du processus approximé en suivant les étapes suivantes :

Le schéma d'Euler

1. Fixer une valeur initiale $X(0) = x_0$;
2. Discrétiser l'intervalle $[0, T]$ pour obtenir des instants isolés $0 = t_0, t_1, \dots, t_N = T$

en posant l'incrément de temps $\Delta t = t_{i+1} - t_i = \frac{1}{N}$;

3. Poser $\tilde{X}(0) = x_0$ et simplifier les notations : $\tilde{X}(t_i) = \tilde{X}_i$ et $W(t_i) = W_i$;
4. Calculer d'une manière itérative

$$\tilde{X}_{i+1} = \tilde{X}_i + a(x_i, t_i)(t_{i+1} - t_i) + \sigma(x_i, t_i)(W_{i+1} - W_i),$$

pour $i = 0, 1, \dots, N - 1$;

5. Pour tout $t \in [t_i, t_{i+1}[$, poser

$$\tilde{X}(t) = \tilde{X}_i + \frac{t - t_i}{t_{i+1} - t_i}(\tilde{X}_{i+1} - \tilde{X}_i).$$

Le processus $\{\tilde{X}(t)\}$ obtenu par le schéma d'Euler est une approximation du processus de diffusion $\{X(t), 0 \leq t \leq T\}$.

Définition 1.13 (Ordre de convergence). *Une approximation à temps discret \tilde{X}_δ d'un processus à temps continu X , avec δ est l'accroissement du temps maximal de la discrétisation, est dit d'ordre γ de convergence forte vers X si pour toute date ultérieure fixe T , nous avons*

$$E|\tilde{X}_\delta(T) - Y(T)| \leq C\delta^\gamma, \text{ pour tout } \delta < \delta_0;$$

avec $\delta_0 > 0$ et C une constante indépendante de δ .

D'après Iacus et Stefano [41], les approximations par le schéma d'Euler sont d'ordre $\gamma = \frac{1}{2}$ de convergence forte.

Chapitre 2

Files d'Attente

Une file d'attente est un groupe d'unités (personnes, machines, tâches, etc.), appelées clients, qui attendent pour effectuer une certaine demande supérieure à la capacité de l'offre. Généralement, cette offre répond à un besoin qui se manifeste sous la forme d'un service. Une file d'attente peut se former au cinéma, à la poste ou même à l'hôpital lorsqu'il s'agit des phénomènes sociaux. Ces files d'attente générées par des individus sont généralement soumises à des politiques d'ordonnement et de service des clients afin d'assurer le bon déroulement de la queue. Néanmoins, le stockage des programmes informatiques avant leur traitement et la gestion des avions pendant le décollage ou l'atterrissage sont également des exemples de files d'attente. L'introduction des mathématiques issues du domaine des probabilités pour l'étude des phénomènes d'attente a donné naissance à la théorie des files d'attente, qui vise à trouver des solutions optimales pour la gestion des files d'attente. La première tentative de modélisation mathématique des files d'attente a été faite par l'ingénieur danois Erlang en 1917 lors de ses travaux sur la gestion du réseau téléphonique de Copenhague entre 1909 et 1920 [19, 20]. Cette théorie a connu un véritable développement grâce aux contributions des mathématiciens Khintchine, Kendall, Pollaczek et Kolmogorov [50, 46, 69, 55].

Ce chapitre est constitué de deux parties principales. Dans la première partie, nous allons rappeler les éléments essentiels de description d'une file d'attente simple ainsi que les objectifs de la théorie des files d'attente. La seconde partie sera consacré à l'étude du phénomène de rappels dans les file d'attente.

2.1 Files d'attente simples

Une file d'attente simple appelée aussi une station se compose d'une unité de service qui contient un ou plusieurs serveurs et d'un espace d'attente. Ainsi, un système de file d'attente est défini comme suit. Les clients qui réclament un service arrivent de l'extérieur, se dirigent éventuellement vers la file d'attente, reçoivent le service et quittent ensuite la station.

2.1.1 Formalisme mathématique d'une file d'attente simple

L'observation récurrente des phénomènes d'attente rencontrés dans la vie réelle a poussé les chercheurs à créer un formalisme mathématique qui permet de caractériser et ensuite d'analyser les systèmes d'attente quantitativement. Ce formalisme est connu dans la littérature par la théorie des files d'attente qui constitue un champ d'application important pour les processus stochastiques.

Caractéristiques d'une file d'attente

Un système de file d'attente peut être complètement caractérisé par les composantes suivantes.

- (i) **Le processus des arrivées** : Cette composante est relative à la distribution des inter-arrivées des clients. Les clients peuvent se présenter dans le système un par un ou par groupe.
- (ii) **Processus de service** : Cette composante est associée à la distribution des temps de service des clients.
- (iii) **Nombre de serveurs** : Une file d'attente peut comporter un ou plusieurs serveurs.
- (iv) **Capacité du système** : Le nombre maximal de clients qui se trouvent simultanément dans la queue et dans les serveurs représente la capacité du système.
- (v) **Source des clients** : La population d'où proviennent les clients peut être finie ou infinie.
- (vi) **Discipline du service** : L'ordre d'accès des clients au service peut être :
 - **FCFS** (First Come First Served) : Le premier arrivé est le premier servi ;
 - **LCFS** (Last Come First Served) : Le dernier clients arrivé est le premier servi ;
 - **FIFO** (First In First Out) : Le premier client arrivé est le premier qui quitte le système ;
 - **LIFO** (Last In First Out) : Le dernier client arrivé est le premier qui quitte le système ;
 - **Random** (Aléatoire) : Les clients accèdent au service aléatoirement.

Notation de Kendall

La notation constituée de la séquence des six symboles $A/B/m/K/L/Z$ introduite par Kendall permet de symboliser les différents paramètres caractérisant une file d'attente. Chaque symbole est réservé pour une des caractéristiques (i)-(vi) mentionnées précédemment dans l'ordre. Pour les codes A et B, les symboles des lois les plus rencontrées ainsi que leurs significations sont résumés dans le tableau suivant.

Symbole de la distribution	Signification
M	Loi exponentielle (Markovienne)
G	Loi générale
GI	Loi générale indépendante
PH	Loi de type phase

TABLE 2.1 – Symboles des distributions de probabilités.

Le code m indique le nombre de serveurs dans le système, K égale au nombre de serveurs m plus le nombre de clients en attente, L indique le nombre total de clients qui ont un potentiel de parcourir le système et Z indique la discipline de service.

Les trois derniers paramètres sont généralement omis lorsqu'ils sont donnés par $K = \infty$, $L = \infty$ et Z en FIFO.

Stabilité des systèmes de files d'attente

Étant donné une file d'attente, on peut étudier l'existence ou non d'un régime stationnaire. Dans les systèmes à capacité infinie, cette étude consiste à déterminer la condition pour laquelle le processus stochastique modélisant le système d'attente considéré est stable. Par conséquent, sa loi de probabilité, après un certain temps lorsque le régime transitoire est terminé, peut être donnée par des formules explicites ou en termes de fonctions génératrices. Les files d'attente fermées, c'est-à-dire les files à capacité finie, sont toujours stables. Cependant, on peut également s'intéresser à l'analyse du régime transitoire afin de décrire l'évolution du système qui n'a pas encore atteint son état stable. Les lois de probabilités, dans ce cas, sont données en fonction du temps. Les méthodes de calcul utilisées afin de décider de la stabilité des systèmes d'attente sont diverses. Pour les files d'attente caractérisées par des processus qui ne sont pas Markoviens, les méthodes les plus utilisées sont :

- **Méthode de la chaîne de Markov incluse** : la chaîne de Markov incluse est obtenue en observant un processus non-Markovien uniquement dans les instants d'occurrence des transitions, comme par exemple les instants de fin de service dans une file d'attente. En d'autres termes, les probabilités de transition sont les seules

à être prises en compte dans l'étude du processus et les temps de séjour dans les états sont complètement ignorés.

- **Méthode de la variable supplémentaire :** cette méthode consiste à affecter à chaque variable non markovienne, qui constitue le processus étudié, une variable supplémentaire qui détermine soit le temps écoulé, soit le temps résiduel associé qui permet de rendre le processus markovien.

Toutefois, ces méthodes sont moins efficaces dans l'analyse des systèmes d'attente avec une distribution générale des temps de rappels dans le cas d'une politique de rappels classique. Ceci est dû au fait que le processus modélisant le système doit tenir compte du temps écoulé ou résiduel de chaque client en orbite qui peut contenir un très grand nombre de clients bloqués. Dans Kernane [47], l'auteur a présenté plusieurs techniques qui ont été employées pour déterminer les conditions de stabilité et d'instabilité de différents systèmes d'attente avec des temps de rappels qui suivent une distribution générale. Dans ce travail, nous introduisons une nouvelle technique, basée sur la modélisation par les PDMPs, afin d'étudier le régime transitoire d'une file M/G/1 avec des temps de rappels distribués selon une loi générale.

Mesures de performance d'une file d'attente

L'objectif de la théorie des files d'attente est de générer un certain nombre de résultats qui permettent de décider de l'efficacité d'un système d'attente. Ces résultats sont des quantités dites mesures de performance. Pour un système d'attente, il est essentiel de déterminer :

- Le nombre moyen de clients dans le système ;
- Le nombre moyen de clients dans la file ;
- Le temps moyen d'attente ;
- La probabilité de blocage : Pour les files à serveur unique, on peut s'intéresser à la distribution de l'état du serveur. La probabilité que le serveur soit occupé, appelée la probabilité de blocage et généralement notée p_1 , est égale à l'intensité du trafic écoulé (the intensity of the carried traffic) selon Falin et Templeton [26]. Dans le cas où aucun client n'est perdu, l'intensité du trafic écoulé est équivalente à l'intensité du trafic offert (the intensity of the offered traffic) qui est a priori égale au nombre moyen de clients entrants dans le système multiplié par le moment d'ordre 1 de la distribution du service. La probabilité que le serveur soit libre est donnée par $p_0 = 1 - p_1$.

2.1.2 Étude de la file M/M/1

La file d'attente M/M/1 est décrite comme suit. Les clients arrivent dans le système un par un selon un processus de Poisson de paramètre λ . Le système de la file comporte

un seul serveur opérationnel dont les durées de service sont des variables aléatoires i.i.d d'une même loi exponentielle de paramètre μ . Le processus des arrivées et le processus de service sont supposés être indépendants.

Pour chaque instant $t \geq 0$, la variable aléatoire $N(t)$ représente le nombre de clients dans le système. Dans un intervalle de temps petit $[t, t + h]$, nous pouvons constater les transitions suivantes :

$$N(t + h) = \begin{cases} N(t) + 1 & \text{une arrivée dans le système avec la probabilité } \lambda h; \\ N(t) - 1 & \text{un départ d'un client du système avec la probabilité } \mu h. \end{cases}$$

Avec $N(t) > 0$.

Grâce aux propriétés fondamentales du processus des arrivées et du processus de service, le processus $\{N(t), t \geq 0\}$ est un processus de naissance et de mort avec un taux de naissance $\lambda_n = \lambda$ et un taux de mort $\mu_n = \mu$ pour tout $n \geq 0$.

Soit $p_n(t) = P(N(t) = n)$ la probabilité d'avoir n clients dans le système à l'instant t . Le système des équations de Chapman-Kolmogorov en régime transitoire s'écrit :

$$\begin{cases} p'_0(t) = -\lambda p_0(t) + \mu p_1(t) & \text{pour } n = 0; \\ p'_n(t) = -(\lambda + \mu)p_n(t) + \lambda p_{n-1}(t) + \mu p_{n+1}(t) & \text{pour } n \geq 1. \end{cases}$$

Soit $p_n = \lim_{t \rightarrow \infty} p_n(t)$ la probabilité d'avoir n clients dans le système à l'équilibre. Le système des équations de Chapman-Kolmogorov en régime stationnaire s'écrit :

$$\begin{cases} \lambda p_0 = \mu p_1 & \text{pour } n = 0; \\ (\lambda + \mu)p_n = \lambda p_{n-1} + \mu p_{n+1} & \text{pour } n \geq 1. \end{cases} \quad (2.1)$$

Afin de résoudre les équations ci-dessus, nous utilisons la fonction génératrice

$$P(z) = \sum_{n=0}^{\infty} z^n p_n, \quad |z| \leq 1.$$

Ainsi, les équations (2.1) s'écrivent

$$P(z) = \sum_{n=0}^{\infty} z^n p_0 \rho^n \quad \text{si } |z\rho| \leq 1,$$

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

D'où, la condition de stabilité du système est donnée par $\rho < 1$.

Par identification, on trouve $p_n = p_0 \rho^n$ pour tout $n \geq 0$. La probabilité p_0 est déter-

minée en utilisant la condition de normalisation $\sum_{n=0}^{\infty} p_n = 1$.

Finalement, la distribution stationnaire du nombre de clients dans le système est donnée par

$$p_n = (1 - \rho)\rho^n, \quad n \geq 0. \quad (2.2)$$

Notons que la distribution stationnaire donnée par Eq. (2.2) peut être déterminée directement en utilisant le résultat sur la distribution stationnaire d'un processus de naissance et de mort (voir Eq.(1.2)).

Les mesures de performance sont calculées comme suit :

(a) Le nombre moyen de clients dans le système :

$$\begin{aligned} E(N) &= \sum_{n=0}^{\infty} np_n \\ &= (1 - \rho) \sum_{n=0}^{\infty} n\rho^n \\ &= \frac{\rho}{1 - \rho}. \end{aligned} \quad (2.3)$$

(b) Le nombre moyen de clients dans la file :

$$\begin{aligned} E(K) &= \sum_{n=0}^{\infty} np_{n+1} \\ &= (1 - \rho) \sum_{n=1}^{\infty} n\rho^{n+1} \\ &= \frac{\rho^2}{1 - \rho}. \end{aligned}$$

(b) La probabilité du blocage :

$$p_1 = \rho.$$

2.2 Files d'attente avec rappels

Le traitement des appels entrants dans un centre d'appels est un phénomène particulier dans la théorie des files d'attente et ses applications. Cette particularité se manifeste par le fait que chaque fois qu'un appelant reçoit un signal occupé, il essaie à nouveau d'obtenir un canal libre. Dans ce cas, aucun appel entrant n'est perdu. Toutefois, les appels qui ne parviennent pas à obtenir un canal libre à la première tentative sont réintroduits dans le centre d'appels avec un flux d'entrée différent du flux d'entrée initial. Dans la théorie des files d'attente, ce phénomène complexe est modélisé par une classe de files d'attente

appelées files avec rappels qui tiennent compte à la fois du flux d'entrée initial et du flux d'entrée secondaire. Les files d'attente avec rappels sont également utilisées pour modéliser d'autres systèmes tels que les systèmes informatiques et de télécommunication ainsi que l'atterrissage des avions dans une piste d'aéroport. D'autres exemples d'applications de files d'attente avec rappels ont été évoqués en détail dans Phung-Duc [68].

2.2.1 Modèle de file d'attente avec rappels avec un seul serveur

Le modèle de file d'attente avec rappels à serveur unique est décrit comme suit. Les clients arrivent de l'extérieur un par un selon un flux d'entrée initial. Si le serveur est libre lorsqu'un client arrive, ce dernier est servi immédiatement et quitte le système. Autrement, si le serveur est occupé, le client entrant rejoint une zone d'attente virtuelle appelée orbite et devient un client bloqué. Les clients bloqués tentent à nouveau leur chance d'être servis jusqu'à ce qu'ils trouvent le serveur inoccupé. Ces tentatives représentent le flux entrant secondaire et suivent différentes politiques de rappels. La politique de rappels est dite classique dans le cas où les clients bloqués rappellent simultanément pour obtenir le service. Dans le cas contraire, où les clients rappellent pour le service un par un avec un temps de rappels de l'orbite qui peut être en cours même si le serveur est occupé, la politique de rappels est dite constante. Le flux d'entrée secondaire qui représente le taux de rappels dépend du nombre de clients dans l'orbite lorsqu'il s'agit de la politique de rappels classique alors que dans le cas de la politique de rappels constante, le taux de rappels est indépendant du nombre de clients bloqués.

Les files avec un seul serveur et des temps de rappels exponentiellement distribués sont largement élaborées dans la littérature. Dans Fayolle [28], les échanges téléphoniques dans un centre d'appels sont modélisés par une file $M/M/1$ avec politique de rappels constante. Dans ce travail, l'auteur a analysé la distribution des états du système ainsi que la distribution du temps de séjour. Cette analyse a ensuite été généralisée au cas d'une file $M/G/1$ dans Farahmand [27] et à celui d'un système $G/M/1$ dans Lillo [59]. Dans le travail de Kim et al. [51], l'analyse du système $G/M/1$ est devenue plus simple par rapport à Lillo [59] en utilisant la méthode de l'analyse matricielle. La politique de rappels constante a également été introduite dans des modèles de file d'attente complexes. Dans Li et Wang [58], les auteurs ont étudié les stratégies de l'adhésion et de l'abandon des clients dans une file d'attente $M/M/1$ avec politique de rappels constante et un serveur sujet aux catastrophes. Le comportement stratégique a été précédemment analysé par Zhou et al. [79] dans une file d'attente avec politique de rappels constante en supposant que le serveur devient actif lorsque le nombre de clients dans l'orbite atteint un certain seuil. De même, les files d'attente avec un seul serveur et politique de rappels classique ont été amplement étudiées dans la littérature [25, 26]. De plus, les vacances du serveur ainsi que les classes de clients sont aussi considérés dans différents modèles de files d'attente

avec politique de rappels classique [35, 36].

L'utilisation des modèles de files d'attente avec une distribution exponentielle des temps de rappels dans la modélisation des systèmes informatiques et de télécommunication a rencontré quelques difficultés comme les temps de rappels ne sont pas toujours conformes à une distribution exponentielle. Cependant, les recherches sur les files avec une distribution générale des temps de rappels, en particulier le modèle M/G/1 avec politique de rappels classique est très restreinte. La première tentative d'analyse de ce modèle a été faite par Kapyrin [45], mais Falin [24] a montré par la suite que son approche était incorrecte. Dans le travail de Yang et Templeton [78], une méthode d'approximation a été proposée pour calculer le nombre de clients dans le système en régime stationnaire en appliquant la propriété de décomposition. Artalejo et Gómez-Corral [2] ont pu également développer une méthode d'approximation pour calculer la distribution limite du système en se basant sur la propriété de décomposition de la file M/G/1 avec une distribution générale des temps de rappels et politique de rappels classique. Dans Kernane [48], la stabilité du système a été étudiée pour ce modèle en supposant que le processus des temps de service est stationnaire et ergodique. Les conditions de stabilité et d'instabilité ont été déterminées par l'auteur pour les différentes politiques de rappels. De plus, dans le travail de Nekrasova [66], le modèle M/G/1 avec une distribution générale des temps de rappels et politique de rappels classique a été étudié en supposant qu'il y ait plusieurs classes de clients dans le système. Par ailleurs, la politique de rappels constante a été précédemment considérée par Gómez-Corral [31] qui a analysé le modèle M/G/1 avec une distribution générale des temps de rappels en régime stationnaire ainsi qu'en régime transitoire. Ultérieurement, plusieurs travaux sur des modèles plus complexes ont supposé la politique de rappels constante avec une distribution générale des temps de rappels [80, 44].

2.2.1.1 File M/G/1 avec une distribution exponentielle des temps de rappels

Le modèle considéré est une file d'attente à capacité infinie et à un serveur unique. Les clients arrivent dans le système selon un processus de Poisson de taux λ . Les durées de service sont i.i.d. avec la même fonction de répartition générale F . Après un temps aléatoire dans l'orbite, les clients bloqués répètent leurs tentatives pour capter le service. Les temps de rappels se produisent selon un processus de Poisson de taux ν . Les inter-arrivées, les périodes de service et les temps de rappel sont supposés être mutuellement indépendants.

Cas de la politique de rappels classique

Le modèle M/G/1 avec politique de rappels classique a été minutieusement étudié dans la littérature [25, 26]. Dans ce modèle, l'état du système à l'instant t est décrit par le processus $\{(C(t), N(t)), t \geq 0\}$ où $C(t)$ représente l'état du serveur ($C(t) = 1$ ou 0

suivant que le serveur est occupé ou libre) et $N(t) \in \mathbb{N}$ est le nombre de clients dans l'orbite. Comme les durées de service ne sont pas exponentiellement distribuées, alors le processus $\{(C(t), N(t)), t \geq 0\}$ n'est pas Markovien. Par conséquent, ce modèle a été analysé en utilisant la méthode de la variable supplémentaire dans Falin [26].

Soient

$\varepsilon(t)$: Le temps de service écoulé dans le cas $C(t) = 1$;

$L_F(s) = \int_0^\infty e^{-sx} dF(x)$: La transformée de Laplace-stieltjes de la fonction F ;

$\beta_k = (-1)^k L_F^{(k)}(0)$: Le moment d'ordre k de la fonction F ;

$k(z) = \sum_0^\infty k_n z^n = L_F(\lambda(1-z))$: La fonction génératrice du nombre d'arrivées durant le temps de service d'un client, avec $k_n = \int_0^\infty \frac{(\lambda x)^n}{n!} \exp(-\lambda x) dF(x)$.

La distribution conjointe de l'état du serveur et le nombre de clients dans l'orbite en régime stationnaire est données par

$$p_{0n} = \lim_{t \rightarrow \infty} P\{C(t) = 0, N(t) = n\};$$

$$p_{1n}(x) = \lim_{t \rightarrow \infty} \frac{d}{dx} P\{C(t) = 1, \varepsilon(t) < x, N(t) = n\}.$$

Les fonctions génératrices correspondantes sont

$$P_0(z) = \lim_{n=0}^\infty z^n p_{0n};$$

$$P_1(z, x) = \lim_{n=0}^\infty z^n p_{1n}(x).$$

Sous la condition $\lambda\beta_1 < 1$ d'existence d'un régime stationnaire, la distribution conjointe est donnée en terme de fonctions génératrices partielles par

$$P_0(z) = (1 - \rho) \exp \left\{ \frac{\lambda}{\nu} \int_1^z \frac{1 - k(u)}{k(u) - u} du \right\};$$

$$P_1(z, x) = \lambda \frac{1 - z}{k(z) - z} P_0(z) [1 - F(x)] e^{-(\lambda - \lambda z)x}.$$

Où $\rho = \lambda\beta_1$.

En posant

$$P_1(z) = \int_0^\infty P_1(z, x) dx$$

avec

$$\int_0^{\infty} [1 - F(x)] e^{-(\lambda - \lambda z)x} dx = \frac{1 - L_F(\lambda(1 - z))}{\lambda(1 - z)},$$

nous obtenons

$$P_1(z) = \frac{1 - k(z)}{k(z) - z} P_0(z).$$

À l'aide des fonctions génératrices $P_0(z)$ et $P_1(z)$, les différentes caractéristiques de performance du système en régime stationnaire ont été déterminées comme suit :

- (a) La distribution stationnaire du nombre de clients dans l'orbite N en terme de fonction génératrice :

$$\begin{aligned} P(z) &= P_0(z) + P_1(z) \\ &= \frac{1 - z}{k(z) - z} (1 - \rho) \exp \left\{ \frac{\lambda}{\nu} \int_1^z \frac{1 - k(u)}{k(u) - u} du \right\}. \end{aligned}$$

En particulier, la moyenne et la variance du nombre de clients dans l'orbite sont donnés par :

$$E(N) = \frac{\lambda^2}{1 - \rho} \left[\frac{\beta_1}{\nu} + \frac{\beta_2}{2} \right];$$

$$\begin{aligned} Var(N) &= \frac{\lambda^3 \beta_3}{3(1 - \rho)} + \frac{\lambda^3 \beta_2}{2\nu(1 - \rho)^2} \\ &\quad + \frac{\lambda^4 \beta_2^2}{4(1 - \rho)^2} + \frac{\lambda \rho}{\nu(1 - \rho)} \\ &\quad + \frac{\lambda^2 \beta_2}{2(1 - \rho)}. \end{aligned}$$

- (b) La distribution du nombre de clients dans le système K en terme de fonction génératrice :

$$\begin{aligned} Q(z) &= P_0(z) + zP_1(z) \\ &= (1 - \rho) \frac{(1 - z)k(z)}{k(z) - z} \exp \left\{ \frac{\lambda}{\nu} \int_1^z \frac{1 - k(u)}{k(u) - u} du \right\}. \end{aligned}$$

En particulier, la moyenne et la variance du nombre de clients dans système sont donnés par :

$$E(K) = \rho + \frac{\lambda^2}{1 - \rho} \left[\frac{\beta_1}{\nu} + \frac{\beta_2}{2} \right];$$

$$\begin{aligned} Var(K) &= \rho - \rho^2 + \frac{\lambda^3 \beta_3}{3(1-\rho)} + \frac{\lambda^4 \beta_2^2}{4(1-\rho)^2} + \frac{\lambda^2 \beta_2}{2(1-\rho)} \\ &+ \lambda^2 \beta_2 + \frac{\lambda^3 \beta_2}{2\nu(1-\rho)^2} + \frac{\lambda \rho}{\nu(1-\rho)}. \end{aligned}$$

(c) La probabilité que le serveur soit occupé p_1 :

$$p_1 = P_1(1) = \rho.$$

Le cas particulier des durées de service exponentiellement distribuées (i.e. $F(x) = 1 - \exp^{-\mu x}$) a été également considéré. Sous la condition de stabilité $\frac{\lambda}{\mu} < 1$, la distribution conjointe $P_{in} = \lim_{t \rightarrow +\infty} P\{C(t) = i, N(t) = n\}$ du processus $(C(t), N(t))$ est donnée par

$$P_{0n} = \frac{\rho^n}{n! \nu^n} \prod_{i=0}^{n-1} (\lambda + i\nu) p_{00}, \quad n \geq 1;$$

$$P_{1n} = \frac{\rho^{n+1}}{n! \nu^n} \prod_{i=1}^n (\lambda + i\nu) p_{00}, \quad n \geq 0;$$

où $p_{00} = (1 - \rho)^{\frac{\lambda}{\nu} + 1}$ avec $\rho = \frac{\lambda}{\mu}$.

Cette distribution est obtenue en suivant deux méthodes : la méthode d'élimination des probabilités p_{1n} des équations de Chapman Kolmogorov décrivant les transitions du système en régime stationnaire et la méthode d'introduction des fonctions génératrices $P_i(z) = \sum_{n=0}^{\infty} z^n P_{in}$ pour résoudre le système d'équations d'équilibre

$$(\lambda + n\nu)p_{0n} = \mu p_{1n}, \quad n \geq 0;$$

$$(\lambda + \mu)p_{1n} = \lambda p_{0n} + (n+1)\nu p_{0,n+1} + \lambda p_{1,n-1}, \quad n \geq 0.$$

Ainsi, les fonctions génératrices correspondantes sont

$$P_0(z) = (1 - \rho) \left(\frac{1 - \rho}{1 - \rho z} \right)^{\frac{\lambda}{\nu}};$$

$$P_1(z) = \rho \left(\frac{1 - \rho}{1 - \rho z} \right)^{\frac{\lambda}{\nu} + 1}.$$

Comme précédemment, les diverses quantités de performance en régime stationnaire sont obtenues à l'aide des fonctions génératrices $P_0(z)$ et $P_1(z)$.

(a) La distribution stationnaire du nombre de clients dans l'orbite en terme de fonction

génératrice

$$P(z) = (1 + \rho - \rho z) \left(\frac{1 - \rho}{1 - \rho z} \right)^{\frac{\lambda}{\nu} + 1}.$$

En particulier, la moyenne et la variance du nombre de clients dans l'orbite sont données par :

$$E(N) = \frac{\rho(\lambda + \rho\nu)}{(1 - \rho)\nu};$$

$$Var(N) = \frac{\rho(\lambda + \rho\nu + \rho^2\nu - \rho^3\nu)}{(1 - \rho)^2\nu}.$$

(b) La distribution stationnaire du nombre de clients dans le système en terme de fonction génératrice :

$$Q(z) = \left(\frac{1 - \rho}{1 - \rho z} \right)^{\frac{\lambda}{\nu} + 1}.$$

En particulier, la moyenne et la variance du nombre de clients dans le système sont données par :

$$E(K) = \frac{\rho(\lambda + \nu)}{\nu(1 - \rho)};$$

$$Var(K) = \frac{\rho(\lambda + \nu)}{\nu(1 - \rho)^2}.$$

(c) la probabilité du blocage :

$$p_1 = \rho = \frac{\lambda}{\mu}.$$

Cas de la politique de rappels constante

Le premier travail qui a considéré la politique de rappels constante est celui de Fayolle [28] en analysant un modèle de file d'attente de type M/M/1. Dans ce travail, il a été supposé que l'orbite lance des rappels selon un processus de Poisson de taux θ afin de trouver le serveur libre. Si à la fin d'un temps de rappels le serveur est inoccupé, alors un seul client bloqué est servi et quitte le système. Le client bloqué qui reçoit le service peut être choisi aléatoirement ou celui qui est en tête de la file dans l'orbite. Ce modèle ainsi décrit est stable si $\frac{\lambda}{\mu} < \frac{\theta}{\lambda + \theta}$ (Voir Fayolle [28] et Artalejo et al. [5]). Contrairement au cas de la politique de rappels classique, la condition de stabilité dans le cas de la politique de rappels constante dépend du taux de rappels ν . L'état du système est décrit par le processus $\{(C(t), N(t)), t \geq 0\}$ défini sur l'espace des états $\{0, 1\} \times \mathbb{N}$. La distribution

conjointe $p_{in} = \lim_{t \rightarrow +\infty} P\{C(t) = i, N(t) = n\}$ s'écrit comme suit :

$$p_{0n} = \frac{\lambda}{\lambda + \theta(1 - \delta_{0,n})}(1 - \rho)\rho^n; \quad (2.4)$$

$$p_{1n} = \frac{\lambda}{\mu}(1 - \rho)\rho^n. \quad (2.5)$$

Avec $\rho = \frac{\lambda(\lambda + \theta)}{\mu\theta}$.

Le modèle de file d'attente avec politique de rappels constante de type M/G/1 a été présenté dans Kernane [48] comme cas particulier et il a été démontré que ce modèle est stable si $\lambda\beta_1 < \frac{\theta}{\lambda + \theta}$.

Autres politiques de rappels

Il existe d'autres politiques de rappels que nous allons présenter dans ce qui suit.

- **La politique de contrôle** : La politique de rappels de contrôle, qui est une variante de la politique de rappels constante, stipule que l'orbite peut générer un temps de rappels juste après la fin d'un temps de service afin de trouver le serveur inactif. Si le temps de rappels se termine avant une arrivée externe, un client bloqué de l'orbite est servi. La politique de rappels de contrôle peut également être mise en place en mettant en file d'attente les clients dans l'orbite selon leur ordre d'arrivée et seul le client qui est en tête de cette file d'attente est autorisé à rappeler pour le service selon un processus de Poisson de taux ν . D'après Kernane [48], la file M/G/1 avec une distribution exponentielle des temps de rappels et politique de rappels de contrôle est stable si $\lambda\beta_1 < \frac{\nu}{\lambda + \nu}$.
- **La politique versatile** : La politique versatile aborde le phénomène de rappels dans son cadre le plus général : Afin de trouver le serveur libre, l'orbite lance un processus de Poisson de taux θ indépendant du processus des arrivées et chaque client en orbite génère des tentatives répétées selon un processus de Poisson de taux ν en se comportant indépendamment des autres clients en orbite et du processus d'arrivée externe. Ainsi, la politique versatile est la combinaison de la politique constante et de la politique classique. Le modèle M/G/1 avec la politique de rappels versatile a été introduit dans la littérature par Artalejo et Gomez-corrall [4] et ensuite analysé par Kernane et Aissani [49]. Il a été démontré que ce modèle est stable si l'une des deux conditions suivantes est vérifiée :
 - (i) $\theta > 0, \nu = 0$ et $\lambda\beta_1 < \frac{\theta}{\lambda + \theta}$;
 - (ii) $\theta \geq 0, \nu > 0$ et $\lambda\beta_1 < 1$.

2.2.1.2 File M/G/1 avec une distribution générale des temps de rappels

Dans ce modèle, les temps de rappels $\{R_n\}$ sont des variables aléatoires i.i.d d'une même fonction de répartition G avec une espérance finie $E(R_1)$. La transformée de Laplace-stieltjes correspondante est $L_G(s) = \int_0^{\infty} e^{-sx} dG(x)$. Nous présentons dans ce qui suit les conditions de stabilité obtenues dans Kernane [48] pour les différentes politiques de rappels.

- (i) **Politique de rappels classique** : $\lambda E(S_1) < 1$.
- (ii) **Politique de rappels constante** : $\lambda E(S_1) < \frac{1 - L_G(\lambda)}{\lambda E(R_1)}$.
- (iii) **Politique de rappels de contrôle** : $\lambda E(S_1) < L_G(\lambda)$.
- (iv) **Politique de rappels versatile** : $\lambda E(S_1) < 1$.

Notons que la suite $\{S_n\}$ représente les temps de service successifs qui est supposée être stationnaire et ergodique avec $0 < E(S_n) < \infty$.

2.2.2 Modèle de file d'attente avec rappels et plusieurs serveurs

Dans cette section, nous nous intéressons au modèle de file d'attente M/M/c avec rappels décrit comme suit. Les clients arrivent dans le système selon un processus de Poisson de taux λ . Si un client arrivant de l'extérieur trouve au moins un serveur libre, il occupe un serveur et quitte le système une fois le service terminé. Dans le cas contraire, si tous les serveurs sont occupés, il rejoint l'orbite. Après un certain temps aléatoire dans l'orbite, chacun de ces clients ou l'orbite lance des rappels jusqu'à ce qu'après une ou plusieurs tentatives il trouve un serveur libre. Le client bloqué qui réussit à rejoindre un serveur inactif quitte le système après la fin de son service. Les durées de service sont des variables aléatoires exponentiellement distribuées avec le paramètre μ et les temps de rappels suivent une loi exponentielle de paramètre ν . Il est supposé que les inter-arrivées, les durées de service et les temps de rappels sont mutuellement indépendants.

L'état du système à l'instant t peut être décrit par le processus $\{(C(t), N(t)), t \geq 0\}$ où $C(t)$ est le nombre de serveurs occupés et $N(t)$ est le nombre de clients bloqués en orbite à l'instant t . Sous les hypothèses ci-dessus, le processus $\{(C(t), N(t)), t \geq 0\}$ est un processus de Markov défini sur l'espace des états $\mathbb{E} = \{0, 1, \dots, c\} \times \mathbb{N}$.

Cas de la politique de rappels classique

Le modèle considéré est celui décrit dans Falin [26] qui suppose que $\mu = 1$. Les taux de transition infinitésimaux $q_{(in)(jm)}$ du processus $\{(C(t), N(t)), t \geq 0\}$ sont donnés par

(i) pour $0 \leq i \leq c - 1$

$$q_{(in)(jm)} = \begin{cases} \lambda & \text{si } (j, m) = (i + 1, n); \\ i & \text{si } (j, m) = (i - 1, n); \\ n\nu & \text{si } (j, m) = (i + 1, n - 1); \\ -(\lambda + i + n\nu) & \text{si } (j, m) = (i, n); \\ 0 & \text{sinon.} \end{cases}$$

(ii) pour $i = c$

$$q_{(cn),(j,m)} = \begin{cases} \lambda & \text{si } (j, m) = (c, n + 1); \\ c & \text{si } (j, m) = (c - 1, n); \\ -(\lambda + c) & \text{si } (j, m) = (c, n); \\ 0 & \text{sinon.} \end{cases}$$

Pour la file considérée, les taux de transition à partir d'un point $(i, n) \in \mathbb{E}$ dépendent de la seconde coordonnée n . De ce fait, l'analyse de ce modèle ainsi que la détermination des mesures de performance s'avère compliquée. D'après Falin [26], la condition de stabilité de ce système est $\lambda < c$. Cette condition est obtenue en utilisant un critère basé sur les dérivées moyennes pour déterminer la condition d'ergodicité suffisante et à partir d'un point de vue pratique, qui dit que le nombre moyen de serveurs occupés doit être inférieur au nombre total des serveurs disponibles, pour déterminer la condition d'ergodicité nécessaire (voir Falin [26] pour plus de détails).

Formules explicites des caractéristiques de performance

Pour le modèle considéré, nous nous intéressons essentiellement au nombre moyen de clients dans l'orbite $E(N)$ et la probabilité de blocage $B = \lim_{t \rightarrow \infty} P\{C(t) = c\}$. Les distributions stationnaires ne peuvent pas être déterminées analytiquement dans le cas général ($c > 2$).

Le cas $c = 2$

La distribution conjointe du nombre de serveurs occupés et du nombre de clients dans l'orbite en régime stationnaire

$$\begin{aligned} p_{0n} &= \lim_{t \rightarrow \infty} P\{C(t) = 0, N(t) = n\}, \\ p_{1n} &= \lim_{t \rightarrow \infty} P\{C(t) = 1, N(t) = n\}, \\ p_{2n} &= \lim_{t \rightarrow \infty} P\{C(t) = 2, N(t) = n\} \end{aligned}$$

est donnée par

$$\begin{aligned}
p_{0n} &= \frac{\lambda^n}{n! \nu^n} \prod_{k=0}^{n-1} \frac{(\lambda + k\nu)^2 + k\nu}{2 + 3\lambda + 2\nu + 2k\nu} p_{00}; \\
p_{1n} &= (\lambda + n\nu) \frac{\lambda^n}{n! \nu^n} \prod_{k=0}^{n-1} \frac{(\lambda + k\nu)^2 + k\nu}{2 + 3\lambda + 2\nu + 2k\nu} p_{00}; \\
p_{2n} &= (1 + \lambda + (n + 1)\nu) \frac{\lambda^n}{n! \nu^n} \prod_{k=0}^n \frac{(\lambda + k\nu)^2 + k\nu}{2 + 3\lambda + 2\nu + 2k\nu} p_{00};
\end{aligned}$$

avec

$$\begin{aligned}
p_{00}^{-1} &= \sum_{n=0}^{\infty} \frac{\lambda^n}{n! \nu^n} \prod_{k=0}^{n-1} \frac{(\lambda + k\nu)^2 + k\nu}{2 + 3\lambda + 2\nu + 2k\nu} \\
&\quad \times \left[1 + \lambda + n\nu + \frac{(1 + \lambda + (n + 1)\nu)((\lambda + n\nu)^2 + n\nu)}{2 + 3\lambda + 2(n + 1)\nu} \right].
\end{aligned}$$

est la probabilité initiale obtenue à l'aide de de la condition de normalisation $\sum_{n=0}^{\infty} (p_{0n} + p_{1n} + p_{2n}) = 1$.

Les fonctions génératrices partielles correspondantes $P_i(z) = \sum_{n=0}^{\infty} z^n p_{in}$ sont

$$\begin{aligned}
P_0(z) &= F\left(a, b, c; \frac{\lambda z}{2}\right) p_{00}; \\
P_1(z) &= \left[\lambda F\left(a, b, c; \frac{\lambda z}{2}\right) + \frac{\lambda^3}{2 + 3\lambda + 2\nu} \right. \\
&\quad \left. \times F\left(a + 1, b + 1, c + 1; \frac{\lambda z}{2}\right) \right] p_{00}; \\
P_2(z) &= \left[\frac{\lambda^2}{2 - \lambda z} F\left(a, b, c; \frac{\lambda z}{2}\right) + \frac{\lambda^3(\lambda z - 1)}{(2 + 3\lambda + 2\nu)(2 - \lambda z)} \right. \\
&\quad \left. \times F\left(a + 1, b + 1, c + 1; \frac{\lambda z}{2}\right) \right] p_{00};
\end{aligned}$$

où

$$\begin{cases} a = \frac{2\lambda + 1 + \sqrt{4\lambda + 1}}{2\nu}; \\ b = \frac{2\lambda + 1 - \sqrt{4\lambda + 1}}{2\nu}; \\ c = \frac{2 + 3\lambda + 2\nu}{2\nu}; \end{cases}$$

et

$$p_{00} = \frac{2 - \lambda}{(2 + \lambda)F(a, b, c; \frac{\lambda}{2}) + \frac{\lambda^3}{2+3\lambda+2\nu}F(a + 1, b + 1, c + 1; \frac{\lambda}{2})};$$

avec

$$F(a, b, c; x) = \sum_{n=0}^{\infty} \frac{x^n}{n!} \prod_{k=0}^{n-1} \frac{(a+k)(b+k)}{c+k}$$

est la fonction hypergéométrique.

En particulier, la fonction génératrice du nombre de clients dans l'orbite

$$P(z) = \left[\left(1 + \lambda + \frac{\lambda^2}{2 - \lambda z} \right) F \left(a, b, c; \frac{\lambda z}{2} \right) + \frac{\lambda^3}{2 + 3\lambda + 2\nu} \right. \\ \left. \times \left(z - 1 + \frac{1}{2 - \lambda z} \right) F \left(a + 1, b + 1, c + 1; \frac{\lambda z}{2} \right) \right] p_{00}.$$

Les expressions explicites des caractéristiques de performance sont données par

(a) Le nombre moyen de clients dans l'orbite

$$E(N) = \frac{1 + \nu}{\nu} \times \frac{\lambda^3 + (\lambda^2 - 2\lambda + 2)g}{(2 - \lambda)(2 + \lambda + g)},$$

avec

$$g = \frac{\lambda^3}{2 + 3\lambda + 2\nu} \times \frac{F(a + 1, b + 1, c + 1; \frac{\lambda}{2})}{F(a, b, c; \frac{\lambda}{2})}.$$

(b) La probabilité de blocage

$$B = \frac{\lambda^2 + (\lambda - 1)g}{2 + \lambda + g}.$$

Cas général

En régime stationnaire, la distribution du nombre de serveurs occupés $p_i = \lim_{t \rightarrow \infty} P(C(t) = i)$ est donnée par

$$p_i = \frac{(\lambda + r_{i-1}) \times \cdots \times (\lambda + r_0)}{i!} p_0, \quad 0 \leq i \leq c; \\ p_0 = \left(\sum_{i=0}^c \frac{(\lambda + r_{i-1}) \times \cdots \times (\lambda + r_0)}{i!} \right)^{-1}.$$

Où $r_i = \frac{\nu E_i(N)}{p_i} = \lim_{t \rightarrow \infty} E(\nu N(t) | C(t) = i)$ est le taux de rappel étant donné que le nombre de serveurs occupés est égal à i .

Les caractéristiques de performance sont données explicitement par

(a) Le nombre moyen de clients dans l'orbite

$$E(N) = \frac{1 + \nu}{\nu} \times \frac{\lambda + \lambda^2 - E((C(t))^2)}{c - \lambda}.$$

(b) Le nombre moyen de serveurs occupés $E(C) = \lim_{t \rightarrow \infty} E(C(t))$

$$E(C) = \lambda.$$

Chapitre 3

Étude des files d'attente dans le cadre d'un PDMP

La représentation par les PDMPs a été adoptée pour décrire différents modèles stochastiques tels que les modèles épidémiologiques dans Gómez-Corral et López-García [32] et Clancy [10] ainsi que les modèles de files d'attente dans Breuer [6]. Ce chapitre est consacré à l'étude des files d'attente en tant que PDMPs. Dans un premier temps, nous présentons quelques modèles de files d'attente simples modélisés par les PDMPs. Puis, dans un second temps, nous mettrons en place le formalisme des PDMPs pour étudier les modèles de files d'attente avec une distribution générale des temps de rappels.

3.1 Files d'attente simples en tant que PDMPs

Dans Davis [16, 17], plusieurs exemples ont été traité en utilisant la méthode de la variable supplémentaire afin de les transformer en des modèles Markoviens modélisables par des PDMPs. Nous nous intéressons dans cette section uniquement aux modèles de files d'attente.

3.1.1 File M/G/1

Les clients arrivent dans le système selon un processus de Poisson de taux λ . Soit T_i l'instant d'arrivée du i ème client et $Y_i > 0$ est la durée de service requise pour chaque client arrivant au temps $T_i, i \geq 1$. Les durées de service Y_i sont i.i.d d'une même loi générale F . Le processus des arrivées et les temps de service sont supposés être indépendants. La charge du service totale jusqu'à l'instant $t > 0$ est donnée par

$$L(t) = L(0) + \sum_{i \geq 1} Y_i \mathbf{I}_{(t \geq T_i)};$$

où $L(0) \geq 0$ est la charge du service initiale lorsque $t = 0$.

Nous définissons également l'équation

$$V(t) = L(t) - \int_0^t \mathbf{I}_{(V(s) > 0)} ds$$

qui représente le temps d'attente virtuel, c'est-à-dire le temps qu'un client arrivant à l'instant t doit attendre pour être servi.

L'état du système à l'instant t est décrit par le processus $\{X(t), t \geq 0\} = \{(v(t), \xi(t)), t \geq 0\}$ tel que $v(t) \in \{0, 1\}$ représente l'état de la file et $\xi(t) = V(t)$. Le processus $\{X(t)\}$ est un PDMP défini sur l'espace des états $\mathbb{E} = (\{0\} \times \{0\}) \cup (\{1\} \times]0, \infty[)$ possédant la limite $\partial^* \mathbb{E} = \{1\} \times \{0\}$. Lorsque la file est vide ($v(t) = 0$), un temps d'attente exponentiel est généré pour une nouvelle arrivée (c'est dû à la propriété d'absence de mémoire de la loi exponentielle). Autrement, si la file est occupée ($v(t) = 1$), le temps d'attente virtuel diminue à un temps unitaire entre les sauts.

Le générateur de ce processus peut être déterminé comme suit. Partant d'un point initial $(1, v)$ et après un temps infinitésimal h , le processus $\{X(t)\}$ passe vers l'état $(1, v - h)$ avec la probabilité $(1 - \lambda h)$ et avec la probabilité λh il passe vers l'état $(1, v + Y - h)$ où Y est une variable aléatoire possédant la même distribution que F .

Ainsi,

$$E_{(1,v)} [f(X(h)) - f(1, v)] = (1 - \lambda h)f(1, v - h) + \lambda h \int_0^\infty f(1, v + y - h) dF(y) - f(1, v) + o(h).$$

En divisant par h nous trouvons

$$\frac{1}{h} [E_{(1,v)} f(X(h)) - f(1, v)] = \frac{1}{h} [f(1, v - h) - f(1, v)] + \lambda \left[\int_0^\infty f(1, v + y - h) dF(y) - f(1, v - h) \right] + o(1).$$

Comme $h \rightarrow 0$ nous obtenons

$$\mathcal{G}f(1, v) = -\frac{\partial}{\partial v} f(1, v) + \lambda \int_0^\infty [f(1, v + y) - f(1, v)] dF(y).$$

De la même façon, en partant du point initial $(0, 0)$, nous obtenons

$$\mathcal{G}f(0, 0) = \lambda \int_0^{\infty} [f(1, y) - f(0, 0)] dF(y).$$

Le domaine $\mathfrak{D}(\mathcal{G})$ est constitué des fonctions $f(v, v)$ qui sont différentiables par rapport à v pour tout v et v , qui satisfont la condition aux limites

$$f(1, 0) = f(0, 0);$$

et vérifient les conditions d'intégrabilité

$$\left| \int_0^{\infty} f(1, v + y) dF(y) - f(1, v) \right| < \infty;$$

$$\left| \int_0^{\infty} f(1, y) dF(y) - f(0, 0) \right| < \infty.$$

3.1.2 File GI/G/1

Nous considérons une file d'attente illimitée avec un seul serveur. Les temps des inter-arrivées sont i.i.d d'une même loi générale c.à.d les clients arrivent selon un processus de renouvellement. Pour obtenir un modèle de Markov nous allons inclure la variable aléatoire $\tau(t)$ désignant le temps écoulé depuis la dernière arrivée. Les durées de service sont i.i.d d'une même loi générale de fonction de répartition F . Les inter-arrivées et les durées de service sont mutuellement indépendantes. Le processus aléatoire décrivant cette file d'attente est donné par $\{X(t), t \geq 0\} = \{(v(t), \xi(t)), t \geq 0\}$ défini sur l'espace des états $\mathbb{E} = (\{0\} \times [0, \infty[) \cup (\{1\} \times [0, \infty[\times]0, \infty[)$ qui admet la limite $\partial^* \mathbb{E} = \{1\} \times [0, \infty[\times \{0\}$. Dans le cas où la file est vide ($v(t) = 0$), $\xi(t) = \tau(t)$. Dans le cas contraire i.e. la file est occupée ($v(t) = 1$), $\xi(t) = (\xi_1(t), \xi_2(t)) = (\tau(t), V(t))$ tel que $V(t)$ est le temps d'attente virtuel.

À partir d'un état initial $(1, \tau, v)$ et après un temps infinitésimal h , le processus $\{X(t)\}$ passe vers l'état $(1, \tau + h, v - h)$ avec la probabilité $(1 - \lambda(\tau)h)$ et avec la probabilité $\lambda(\tau)h$ il passe vers l'état $(1, \tau', v + Y - h)$ où $\tau' \in [0, h]$ et Y est une variable aléatoire possédant la même distribution que F . Dans ce cas, le générateur infinitésimal s'écrit comme suit :

$$\mathcal{G}f(1, \tau, v) = \frac{\partial}{\partial \tau} f(1, \tau, v) - \frac{\partial}{\partial v} f(1, \tau, v) + \lambda(\tau) \int_0^{\infty} [f(1, 0, v + y) - f(1, \tau, v)] dF(y).$$

D'autre part, si le processus démarre d'un état initial $(0, \tau)$ alors le générateur infinitésimal est donné par

$$\mathcal{G}f(0, \tau) = \frac{\partial}{\partial \tau} f(0, \tau) + \lambda(\tau) \int_0^{\infty} [f(1, 0, y) - f(0, \tau)] dF(y).$$

Le domaine $\mathfrak{D}(\mathcal{G})$ est constitué des fonctions f qui sont différentiables par rapport à τ et v pour tout v, τ et v , qui satisfont la condition aux limites

$$f(1, \tau, 0) = f(0, \tau, 0);$$

et vérifient les conditions d'intégrabilité

$$\left| \int_0^{\infty} f(1, 0, v + y) dF(y) - f(1, \tau, v) \right| < \infty;$$

$$\left| \int_0^{\infty} f(1, 0, y) dF(y) - f(0, \tau) \right| < \infty.$$

3.1.3 File M/G/c

Dans ce modèle, les clients arrivent dans le système un par un selon un processus de Poisson de taux λ . Les durées de service sont des variables aléatoires i.i.d de même distribution générale F . Les c serveurs sont identiques et servent les clients dans leur ordre d'arrivée. Comme la file d'attente est soumise à des arrivées individuelles de Poisson, la probabilité que deux services ou plus soient achevés est négligeable dans un intervalle de temps exponentiel. Ainsi, lorsqu'il y a des clients dans la queue, un seul serveur peut être inactif à la fois. En outre, le processus des arrivées et le processus de service sont supposés être indépendants.

L'état du système à l'instant t est donné par le PDMP $\{X(t), t \geq 0\} = \{(C(t), \mathbf{Y}(t)), t \geq 0\}$ défini sur l'espace des états $\mathbb{E} = \mathbb{N} \times \mathbb{R}_+^c$. Pour chaque état $(n, \mathbf{y}) = (n, y_1, \dots, y_c) \in \mathbb{E}$, la première composante n représente le nombre de clients en attente et les composantes y_i représentent le temps de service résiduel dans le i ème serveur. Dans le cas où le i ème serveur est libre, $y_i = 0$.

Pour tout $x = (n, \mathbf{y}) \in \mathbb{E}$ et $t \in \mathbb{R}_+$, la fonction de flux $\phi(t, x)$ est définie par

$$\phi(t, x) = (n, (y_1 - t)^+, \dots, (y_c - t)^+);$$

avec $(y_i - t)^+ = \max(0, y_i - t)$.

De plus, le temps écoulé jusqu'à ce que le serveur suivant soit inactif est donné par

$$t_*(\mathbf{y}) = \begin{cases} \min\{y_i : 1 \leq i \leq c, y_i > 0\} & \text{pour } \mathbf{y} \neq \mathbf{0}; \\ \infty & \text{pour } \mathbf{y} = \mathbf{0}. \end{cases}$$

Dans ce qui suit, deux mesures de transition Q_1 et Q_2 seront introduites pour capturer les transitions qui peuvent se produire. La mesure de transition Q_1 est associée au processus des arrivées et est donnée par

$$Q_1(x; \{m\} \times A) = \begin{cases} \delta_{m,n+1} \mathbf{1}_A(\mathbf{y}) & \text{pour } n \geq 1, \prod_{i=1}^c y_i > 0; \\ \delta_{m,n} \prod_{j \neq i} \mathbf{1}_{A_j}(y_j) P_F(A_i) & \text{pour } i = \min\{l : y_l = 0\}; \end{cases}$$

pour tout $x \in \mathbb{E}$ et $A = A_1 \times \dots \times A_c$ avec $A_i \in \mathcal{B}_{\mathbb{R}_+}$ où $\mathcal{B}_{\mathbb{R}_+}$ est la σ -algèbre des ensembles de Borel sur l'intervalle \mathbb{R}_+ .

Notons que $\delta_{m,n}$ est le symbole de Kronecker et le dernier cas dans la définition de Q_1 est à prendre en compte uniquement dans le cas $n = m = 0$. Dans ce cas, la mesure Q_1 reflète la transition $(0, y_1, \dots, y_{i-1}, 0, y_{i+1}, \dots, y_c) \rightarrow (0, y_1, \dots, y_{i-1}, y_i, y_{i+1}, \dots, y_c)$ relative à un nouveau client externe entrant dans le système qui rejoint immédiatement le premier serveur inactif. Le premier cas est relatif à la transition $(n, y_1, \dots, y_c) \rightarrow (n+1, y_1, \dots, y_c)$ correspondante au cas où le nouveau client entrant trouve tous les serveurs occupés et rejoint ainsi la file d'attente.

La mesure de transition Q_2 est associée au cas où l'un des serveurs devient inactif. S'il y a des clients en attente dans la queue, il commencera immédiatement à servir un nouveau client. Ceci fait référence à la transition $(n, y_1, \dots, y_{i-1}, 0, y_{i+1}, \dots, y_c) \rightarrow (n-1, y_1, \dots, y_{i-1}, y_i, y_{i+1}, \dots, y_c)$. Ainsi, la mesure de transition Q_2 est donnée par

$$Q_2(x; \{m\} \times A) = \begin{cases} \delta_{m,n-1} \times \prod_{j \neq i} \mathbf{1}_{A_j}(y_j) P_F(A_i) & \text{pour } n \geq 1, y_i = 0; \\ \delta_{m,n} \times \mathbf{1}_A(\mathbf{y}) & \text{pour } n = 0. \end{cases}$$

D'après Breuer[6], la représentation par le PDMP $\{X(t), t \geq 0\} = \{(C(t), \mathbf{Y}(t)), t \geq 0\}$ défini sur $\mathbb{E} = \mathbb{N} \times \mathbb{R}_+^c$ est légèrement imprécise comme le cas $n = 0$ ne correspond pas à un saut. En posant $\mathbb{E} = (\mathbb{N} \times \mathbb{R}_+^c) \cup (\{0\} \times \cup_{k=1}^{c-1} \mathbb{R}_+^k)$, le processus $X(t)$ se conforme exactement à la spécification de Davis [17]. Ainsi, les sauts correspondants au cas de la file vide ($n = 0$) et au moins un temps de service non achevé ($\cup_{k=1}^{c-1} \mathbb{R}_+^k$) peuvent également être capturés.

3.2 Étude de la file M/G/1 avec une distribution générale des temps de rappels dans le cadre d'un PDMP

Dans cette section, nous traitons la file d'attente M/G/1 avec une distribution générale des temps de rappels comme étant un PDMP au lieu de l'étudier en utilisant les méthodes classiques telles que la chaîne de Markov incluse et la méthode de la variable supplémentaire. Cette nouvelle approche permet d'étudier la dynamique d'un modèle aussi complexe de manière plus approfondie ainsi que de déterminer les quantités de performance en régime transitoire à travers les martingales associées, ce qui est difficile à élaborer même pour les files d'attente Markoviennes simples. En effet, la modélisation par les PDMPs a été introduite pour analyser plusieurs modèles dans la littérature notamment les systèmes de files d'attente [6] et les modèles épidémiques [10, 32]. Cependant, la similitude entre la dynamique du modèle SIR avec une distribution générale des durées d'infection et celle de la file d'attente M/G/1 avec des temps de rappels distribués selon une loi générale, qui est clairement mise en évidence par Gómez-Corral et López-García [32], nous a le plus motivé à réaliser les travaux dans Meziani et Kernane [62, 61].

Dans cette section, nous commençons par modéliser la file M/G/1 avec une distribution générale des temps de rappels par un PDMP dans le cas d'une politique de rappels classique et dans le cas d'une politique de rappels constante. Ensuite, pour chaque modèle, nous présentons les martingales associées. Enfin, nous calculons le nombre moyen de clients bloqués dans l'orbite en régime transitoire en utilisant les martingales déterminées précédemment.

3.2.1 File M/G/1 avec politique de rappels classique

Le modèle considéré est la file M/G/1 avec des temps de rappels distribués selon une loi générale. Les clients arrivent dans le système selon un processus de Poisson de taux λ . Chaque client entrant qui trouve le serveur occupé rejoint l'orbite. Les durées de service sont i.i.d. et ont la même fonction de distribution générale F . Après une période de temps aléatoire dans l'orbite, chaque client bloqué répète sa tentative pour accéder au service. Les temps de rappel sont i.i.d. et ont la même fonction de distribution générale G . Les inter-arrivées, les durées de service et les temps de rappel sont supposés être mutuellement indépendants. Le modèle étudié peut être représenté comme un PDMP de la manière suivante. Soit $\{X(t) = (v(t), \xi(t)); t \in \mathbb{R}_+\}$, où $v(t) = (C(t), N(t))$ et $\xi(t) = (Y(t), \mathbf{R}(t))$, le PDMP décrivant l'état du système à l'instant t . La composante $C(t)$ représente l'état du serveur ($C(t)$ est égal à 1 ou 0 selon que le serveur est occupé ou libre), $N(t)$ est le nombre de clients bloqués, $Y(t)$ est le temps de service résiduel et $\mathbf{R}(t) = (R_1(t), R_2(t), \dots, R_{N(t)}(t))$ où $R_k(t)$ désigne la durée restante du temps de rappel du k ième client bloqué avec $1 \leq k \leq N(t)$. Le processus considéré est défini sur

l'espace des états $\mathbb{E} = (\mathbb{E}_0 \cup \partial^*\mathbb{E}_0) \cup (\mathbb{E}_1 \cup \partial^*\mathbb{E}_1)$, où $\mathbb{E}_0 = \{(0, n, 0, r_1, r_2, \dots, r_n), n \in \mathbb{N}, 0 < r_1 < r_2 < \dots < r_n\}$, $\partial^*\mathbb{E}_0 = \{(0, n, 0, 0, r_2, \dots, r_n), n \in \mathbb{N}^*, 0 < r_2 < \dots < r_n\}$, $\mathbb{E}_1 = \{(1, n, y, r_1, r_2, \dots, r_n), n \in \mathbb{N}, y > 0, 0 < r_1 < r_2 < \dots < r_n\}$ et $\partial^*\mathbb{E}_1 = \{(1, n, 0, r_1, r_2, \dots, r_n), n \in \mathbb{N}, 0 < r_1 < r_2 < \dots < r_n\}$. Cette formulation particulière de l'espace des états \mathbb{E} nous permet de spécifier facilement la mesure de transition $Q(x; \cdot)$ de $\{X(t)\}$ et nous pouvons voir qu'elle est entièrement adéquate à notre cas où tous les clients bloqués réessaient de capturer le service simultanément.

Définissons la fonction de flux ϕ :

$$\phi(t, x) = \begin{cases} (0, n, 0, (r_1 - t)^+, (r_2 - t)^+, \dots, (r_n - t)^+) & \text{pour } x \in \mathbb{E}_0; \\ (1, n, (y - t)^+, (r_1 - t)^+, (r_2 - t)^+, \dots, (r_n - t)^+) & \text{pour } x \in \mathbb{E}_1. \end{cases}$$

Nous donnons en outre quelques notations pour une utilisation ultérieure. Soit $\mathcal{B}_{(a,b)}$ la σ -algèbre des ensembles de Borel sur l'intervalle $]a, b[$ et β_n la σ -algèbre de Borel sur l'ensemble $F^{(n)} = \{(u_1, \dots, u_n) \in]0, \infty[^n : u_1 < \dots < u_n\}$.

Dans le même esprit que Gómez-Corral et López-García [32] et Breuer[6], afin de décrire plus nettement les sauts qui peuvent se produire, nous allons introduire trois mesures de transition Q_1 , Q_2 et Q_3 qui reflèteront les transitions associées au processus des arrivées, à l'achèvement du service et au délogement d'un client bloqué de l'orbite, respectivement.

Pour les états $x \in \mathbb{E}_0$, la mesure de transition $Q_1(x; \cdot)$ est donnée par

$$Q_1(x; \{1\} \times \{n\} \times A \times B) = P_F(A) \mathbf{1}_B(r_1, \dots, r_n),$$

où $A \in \mathcal{B}_{(0,\infty)}$ et $B \in \beta_n$.

La mesure de transition $Q_1(x; \cdot)$ capte la transition de l'état $(0, n, 0, r_1, \dots, r_n)$ vers l'état $(1, n, y, r_1, \dots, r_n)$ relative à un nouveau client externe entrant dans le système qui rejoint immédiatement le serveur inactif.

Pour les états $x \in \partial^*\mathbb{E}_0$, $A \in \mathcal{B}_{(0,\infty)}$ et $B \in \beta_{n-1}$ la mesure de transition $Q_3(x; \cdot)$ est donnée par

$$Q_3(x; \{1\} \times \{n-1\} \times A \times B) = P_F(A) \mathbf{1}_B(r_2, \dots, r_n).$$

Il s'agit de la transition de l'état $(0, n, 0, 0, r_2, \dots, r_n)$ vers l'état $(1, n-1, y, r'_1, \dots, r'_{n-1})$ qui se produit lorsqu'un client bloqué rejoint le serveur inactif.

De la même manière, nous déterminons les mesures de transition $Q_1(x; \cdot)$ et $Q_2(x; \cdot)$ pour les états $x \in \mathbb{E}_1$ et $x \in \partial^*\mathbb{E}_1$, respectivement. La mesure de transition $Q_1(x; \cdot)$ est

spécifiée comme suit :

(i) Pour $A \in \mathcal{B}_{(0,\infty)}$, $B \in \mathcal{B}_{(0,r_1)}$ et les ensembles $C \in \beta_n$,

$$Q_1(x; \{1\} \times \{n+1\} \times A \times B \times C) = \mathbf{1}_A(y)P_G(B)\mathbf{1}_C(r_1, \dots, r_n).$$

(ii) Pour $A \in \mathcal{B}_{(0,\infty)}$, les ensembles $B \in \beta_k$, $C \in \mathcal{B}_{(r_k, r_{k+1})}$ et les ensembles $D \in \beta_{n-k}$,

$$Q_1(x; \{1\} \times \{n+1\} \times A \times B \times C \times D) = \mathbf{1}_A(y)\mathbf{1}_B(r_1, \dots, r_k)P_G(C)\mathbf{1}_D(r_{k+1}, \dots, r_n).$$

(iii) Pour $A \in \mathcal{B}_{(0,\infty)}$, les ensembles $B \in \beta_n$ et $C \in \mathcal{B}_{(r_n, \infty)}$,

$$Q_1(x; \{1\} \times \{n+1\} \times A \times B \times C) = \mathbf{1}_A(y)\mathbf{1}_B(r_1, \dots, r_n)P_G(C).$$

La mesure de transition $Q_1(x; \cdot)$ reflète la transition de l'état $(1, n, y, r_1, \dots, r_n)$ vers l'état $(1, n+1, y, r'_1, \dots, r'_n, r'_{n+1})$ correspondante au cas où un client entrant trouve le serveur occupé et rejoint l'orbite. Ainsi, il devient un client bloqué et un nouveau temps de rappel résiduel généré à partir de G doit être ajouté au vecteur (r_1, \dots, r_n) à sa position appropriée, ce qui donne un nouveau vecteur des temps de rappel résiduels $(r'_1, \dots, r'_n, r'_{n+1})$ avec $r'_1 < \dots < r'_n < r'_{n+1}$.

Enfin, la mesure de transition $Q_2(x; \cdot)$ associée à la transition de l'état $(1, n, 0, r_1, \dots, r_n)$ vers l'état $(0, n, 0, r_1, \dots, r_n)$, lorsque le serveur devient inactif, est donnée par

$$Q_2(x; \{0\} \times \{n\} \times \{0\} \times A) = \mathbf{1}_A(r_1, \dots, r_n),$$

où $A \in \beta_n$.

Notons que, pour notre processus, le taux des sauts est exactement le taux d'arrivée λ . Par conséquent, nous avons $\Lambda(t, x) = \int_0^t \lambda(\phi(s, x))ds = \lambda t$.

Martingales associées

Nous présentons dans cette section les martingales associées au PDMP qui modélise le système de file d'attente avec rappel considéré en utilisant son générateur infinitésimal.

Théorème 3.1 (Meziani et Kernane [62]). *Pour $0 \leq z_1 \leq 1, 0 \leq z_2 \leq 1, \gamma > 0, \delta \geq 0$ et $n \geq 0$ la fonction*

$$z_1^{C(t)} z_2^{N(t)} e^{-\gamma Y(t)} e^{-\delta \sum_{k=1}^{N(t)} R_k(t)} e^{\theta_{C(t)}(t)} \quad (3.1)$$

avec

$$\theta_{C(t)}(t) = \begin{cases} (\lambda - n\delta)t - \ln\{\lambda z_2 \varphi_R(\delta) + \gamma\} \\ + \ln\{\lambda z_1 \varphi_S(\delta)[e^{-(\lambda z_2 \varphi_R(\delta) + \gamma)t} - 1] + \lambda z_2 \varphi_R(\delta) + \gamma\} & \text{pour } C(t) = 0; \\ (\lambda - \lambda z_2 \varphi_R(\delta) - n\delta - \gamma)t & \text{pour } C(t) = 1. \end{cases} \quad (3.2)$$

est une martingale pour les états $x \in \mathbb{E}_{C(t)}$, où

$$\varphi_S(\gamma) = \int_0^\infty e^{-\gamma y} dF(y); \gamma > 0 \quad \text{et} \quad \varphi_R(\delta) = \int_0^\infty e^{-\delta r} dG(r); \delta \geq 0.$$

Preuve. Le générateur infinitésimal du processus $\{(X(t), t)\}$, agissant sur une fonction $f(i, n, y, r_1, \dots, r_n, t) \in \mathfrak{D}(\mathcal{G})$, est donné par

$$\begin{aligned} \mathcal{G}f(0, n, 0, r_1, r_2, \dots, r_n, t) &= \lambda \int_0^\infty [f(1, n, y, r_1, r_2, \dots, r_n, t) - f(0, n, 0, r_1, r_2, \dots, r_n, t)] dF(y) \\ &\quad - \sum_{k=1}^n \frac{\partial}{\partial r_k} f(0, n, 0, r_1, r_2, \dots, r_n, t) + \frac{\partial}{\partial t} f(0, n, 0, r_1, r_2, \dots, r_n, t). \end{aligned} \quad (3.3)$$

$$\begin{aligned} \mathcal{G}f(1, n, y, r_1, r_2, \dots, r_n, t) &= \lambda \int_0^\infty [f(1, n+1, y, r_1, r_2, \dots, r_n, r_{n+1}, t) \\ &\quad - f(1, n, y, r_1, r_2, \dots, r_n, t)] dG(r_{n+1}) \\ &\quad - \frac{\partial}{\partial y} f(1, n, y, r_1, r_2, \dots, r_n, t) - \sum_{k=1}^n \frac{\partial}{\partial r_k} f(1, n, y, r_1, r_2, \dots, r_n, t) \\ &\quad + \frac{\partial}{\partial t} f(1, n, y, r_1, r_2, \dots, r_n, t). \end{aligned} \quad (3.4)$$

Où $\mathfrak{D}(\mathcal{G})$ est le domaine du générateur \mathcal{G} qui comporte les fonctions $f(i, n, y, r_1, \dots, r_n, t)$ qui sont différentiables par rapport à y, r_1, \dots, r_n et t pour tout $i, n, y, r_1, \dots, r_n, t$, qui satisfont les conditions limites

$$f(0, n, 0, 0, r_2, \dots, r_n, t) = \int_0^\infty f(1, n-1, y, r'_1, \dots, r'_{n-1}, t) dF(y); \quad (3.5)$$

$$f(1, n, 0, r_1, \dots, r_n, t) = f(0, n, 0, r_1, \dots, r_n, t); \quad (3.6)$$

où $(r'_1, \dots, r'_{n-1}) = (r_2, \dots, r_n)$ et qui vérifient les conditions d'intégrabilité

$$\left| \int_0^\infty f(1, n, y, r_1, \dots, r_n, t) dF(y) - f(0, n, 0, r_1, \dots, r_n, t) \right| < \infty; \quad (3.7)$$

$$\left| \int_0^\infty f(1, n+1, y, r_1, \dots, r_n, r_{n+1}, t) dG(r_{n+1}) - f(1, n, y, r_1, \dots, r_n, t) \right| < \infty. \quad (3.8)$$

Définissons maintenant la fonction

$$f(i, n, y, r_1, r_2, \dots, r_n, t) = z_1^i z_2^n e^{-\gamma y} e^{-\delta \sum_{k=1}^n r_k} e^{\theta_i(t)}, \quad (3.9)$$

où $\theta_i(t)$ est donné par Eq.(3.2) pour $C(t) = i$, $i \in \{0, 1\}$.

En substituant Eq.(3.9) dans Eq.(3.3) pour $i = 0$, nous obtenons

$$\begin{aligned} \mathcal{G}f(0, n, 0, r_1, r_2, \dots, r_n, t) &= \lambda \int_0^\infty [z_1 e^{-\gamma y} e^{\theta_1(t) - \theta_0(t)} - 1] dF(y) + \sum_{k=1}^n \delta + \theta'_0(t) \\ &= \lambda z_1 e^{\theta_1(t) - \theta_0(t)} \varphi_S(\gamma) - \lambda + n\delta + \theta'_0(t) \\ &= \lambda z_1 \frac{(\lambda z_2 \varphi_R(\delta) + \gamma) e^{-(\lambda z_2 \varphi_R(\delta) + \gamma)}}{\lambda z_1 \varphi_S(\gamma) [e^{-(\lambda z_2 \varphi_R(\delta) + \gamma)} - 1] + \lambda z_2 \varphi_R(\delta) + \gamma} \varphi_S(\gamma) \\ &\quad - \lambda + n\delta + \lambda - n\delta \\ &\quad - \lambda z_1 \varphi_S(\gamma) \frac{(\lambda z_2 \varphi_R(\delta) + \gamma) e^{-(\lambda z_2 \varphi_R(\delta) + \gamma)}}{\lambda z_1 \varphi_S(\gamma) [e^{-(\lambda z_2 \varphi_R(\delta) + \gamma)} - 1] + \lambda z_2 \varphi_R(\delta) + \gamma} \\ &= 0. \end{aligned}$$

De même, nous substituons Eq. (3.9) dans Eq. (3.4) pour $i = 1$.

$$\begin{aligned} \mathcal{G}f(1, n, y, r_1, r_2, \dots, r_n, t) &= \lambda \int_0^\infty [z_2 e^{-\delta r_{n+1} - 1}] dG(r_{n+1}) + \gamma + n\delta + \theta'_1(t) \\ &= \lambda z_2 \varphi_R(\delta) - \lambda + \gamma + n\delta + \theta'_1(t) \\ &= \lambda z_2 \varphi_R(\delta) - \lambda + \gamma + n\delta + \lambda - \lambda z_2 \varphi_R(\delta) - n\delta - \gamma \\ &= 0. \end{aligned}$$

Ainsi, en vertu de la propriété du générateur infinitésimal, Eq.(3.1) est une martingale pour le processus $\{X(t)\}$. \square

Notons que $\theta_{C(t)}(t)$ a été déterminée en résolvant le système des équations différentielles

ordinaires (ODEs)

$$\begin{cases} \lambda z_1 \varphi_S(\gamma) e^{\theta_1(t)} - (\lambda - n\delta) e^{\theta_0(t)} + \theta'_0(t) e^{\theta_0(t)} = 0; \\ \lambda z_2 \varphi_R(\delta) - \lambda + \gamma + n\delta + \theta'_1(t) = 0; \end{cases}$$

avec $\theta_i(0) = 0$ pour $i \in \{0, 1\}$. Ce système est obtenu en remplaçant Eq. (3.9) dans $\mathcal{G}f(0, n, 0, r_1, r_2, \dots, r_n, t) = 0$ et $\mathcal{G}f(1, n, y, r_1, r_2, \dots, r_n, t) = 0$.

Théorème 3.2 (Meziani et Kernane [62]). *Soient τ_0 et τ_1 les temps d'arrêt qui mettent fin à la période d'inactivité du serveur (quand le processus reste dans \mathbb{E}_0) et à la période d'occupation du serveur (quand le processus reste dans \mathbb{E}_1), respectivement.*

Les processus

$$f(C(t), N(t), Y(t), \mathbf{R}(t), t) - f(0, N(0), 0, \mathbf{R}(0), 0) - \int_0^t \mathcal{G}f(0, N(s), 0, \mathbf{R}(s), s) ds \quad (3.10)$$

et

$$f(C(t), N(t), Y(t), \mathbf{R}(t), t) - f(1, N(0), Y(0), \mathbf{R}(0), 0) - \int_0^t \mathcal{G}f(1, N(s), Y(s), \mathbf{R}(s), s) ds \quad (3.11)$$

définis pour $t \in [0, \tau_0]$ et $t \in [0, \tau_1]$, respectivement, sont des martingales pour toute fonction $f \in \mathcal{D}(\mathcal{G})$.

Preuve. En exploitant la proposition 1.1 ainsi que le générateur infinitésimal Eq.(3.3)-Eq.(3.4) le résultat est immédiatement obtenu. \square

Il convient de noter que selon le traitement de notre modèle comme étant un PDMP, les temps d'arrêt τ_0 et τ_1 sont donnés par les variables aléatoires $\min(A(t), R_1(t))$ et $Y(t)$, respectivement. La composante $A(t)$ correspond au temps d'inter-arrivée qui suit une distribution exponentielle de paramètre λ .

Nombre moyen de clients dans l'orbite

Dans cette section, nous allons tirer l'espérance conditionnelle du clients dans l'orbite pendant les périodes d'inactivité et d'occupation du serveur séparément. Pour ce faire, nous utiliserons principalement le résultat du théorème 3.2 ci-dessus. En effet, les auteurs dans Dassios et Zhao [15] ont montré que cette méthode basée sur les martingales permet de calculer tous les moments de $N(t)$ sans tenir compte de la condition de stabilité (voir théorème 3.6 et théorème 3.8 dans [15]).

Théorème 3.3 (Meziani et Kernane [62]). *L'espérance conditionnelle du nombre de clients bloqués $N(t)$ sachant que $N(0) = n_0$ et $Y(0) = y_0$ (lorsque $Y(t) \in \mathbb{E}_1$) est donnée*

par :

$$E[N(t)|N(0) = n_0] = \begin{cases} n_0 + \lambda t & \text{pour } t \in [0, \tau_0]; \\ n_0 + \lambda t + \frac{1}{\mu_1}(y_0 - t - \mu_Y(t)) & \text{pour } t \in [0, \tau_1]. \end{cases}$$

Où $\mu_1 = \int_0^{\infty} y dF(y)$ est le moment d'ordre 1 de la fonction F et $\mu_Y(t)$ est l'espérance du temps de rappels résiduel pour $t \in [0, \tau_1]$.

Preuve. En posant $f(i, n, y, r_1, \dots, r_n, t) = y + n\mu_1$ et vérifiant les conditions Eq.(3.5) – Eq.(3.8) d'appartenance au domaine $\mathfrak{D}(\mathcal{G})$, nous avons

$$\mathcal{G}f(0, n, 0, r_1, \dots, r_n, t) = \lambda\mu_1 \quad \text{et} \quad \mathcal{G}f(1, n, y, r_1, \dots, r_n, t) = \lambda\mu_1 - 1.$$

D'après le théorème 3.2 ci-dessus, $Y(t) + N(t)\mu_1 - n_0\mu_1 - \int_0^t \lambda\mu_1 ds$ et $Y(t) + N(t)\mu_1 - y_0 - n_0\mu_1 - \int_0^t (\lambda\mu_1 - 1) ds$ sont des martingales. Ainsi, pour $t \in [0, \tau_0]$

$$E[N(t)|N(0) = n_0] = n_0 + \lambda t - \frac{1}{\mu_1} E[Y(t)] \quad (3.12)$$

et pour $t \in [0, \tau_1]$

$$E[N(t)|N(0) = n_0, Y(0) = y_0] = n_0 + \lambda t + \frac{1}{\mu_1} (y_0 - t - E[Y(t)]). \quad (3.13)$$

En outre, nous avons

$$E[Y(t)] = \begin{cases} 0 & \text{pour } t \in [0, \tau_0]; \\ \mu_Y(t) & \text{pour } t \in [0, \tau_1]. \end{cases}$$

En substituant $E[Y(t)]$ dans Eq.(3.12) et Eq.(3.13), le résultat est atteint. \square

Notons que l'espérance du temps de service résiduel peut être déterminé comme suit. Soit S la variable aléatoire qui représente la durée de service d'un client dans le système considéré. Le temps de service de résiduel $Y(t)$ est donné par

$$Y(t) = S - t | S > t.$$

Ainsi, d'après Guess et al. [34]

$$\begin{aligned}\mu_Y(t) &= E(Y(t)) \\ &= \frac{1}{\bar{F}(t)} \int_t^\infty \bar{F}(u) du.\end{aligned}$$

3.2.2 File M/G/1 avec politique de rappels constante

Nous considérons le même modèle précédent avec une politique de rappel différente : la politique de rappels constante. Après une période de temps aléatoire dans l'orbite, le client bloqué qui est en tête de l'orbite répète sa tentative d'obtenir le service. De ce fait, les clients bloqués rappellent pour le service suivant leur ordre d'entrée dans l'orbite. Le modèle ainsi décrit est modélisé par le PDMP $\{X(t) = (C(t), N(t)), (Y(t), R(t)); t \in \mathbb{R}^+\}$ défini sur $\mathbb{E} = (\{e_0\} \cup \mathbb{E}_0 \cup \partial^*\mathbb{E}_0) \cup (\{e_1\} \cup \mathbb{E}_1 \cup \partial^*\mathbb{E}_1)$ tel que $e_0 = (0, 0, 0, 0)$, $\mathbb{E}_0 = \{(0, n, 0, r), n \in \mathbb{N}^*, r > 0\}$, $\partial^*\mathbb{E}_0 = \{(0, n, 0, 0), n \in \mathbb{N}^*\}$, $e_1 = \{(1, 0, y, 0), y > 0\}$, $\mathbb{E}_1 = \{(1, n, y, r), n \in \mathbb{N}^*, y > 0, r > 0\}$ et $\partial^*\mathbb{E}_1 = \{(1, n, 0, r), n \in \mathbb{N}^*, r > 0\}$.

Les trois mesures de transitions Q_1, Q_2 et Q_3 qui décrivent les sauts du PDMP $\{X(t)\}$ sont données comme suit.

Pour les états e_0 et $x \in \mathbb{E}_0$, la mesure de transition Q_1 associée à l'arrivée d'un nouveau client dans le système est donnée par

$$Q_1(e_0; \{1\} \times \{0\} \times A \times \{0\}) = P_F(A);$$

$$Q_1(x; \{1\} \times \{n\} \times A \times B) = P_F(A)\mathbf{1}_B(r).$$

Où $A, B \in \mathcal{B}_{(0, \infty)}$.

Pour les états $x \in \partial^*\mathbb{E}_0$ et $A, B \in \mathcal{B}_{(0, \infty)}$, la mesure de transition Q_3 relative à un client bloqué qui rejoint le service est donnée par

$$Q_3(x; \{1\} \times \{n-1\} \times A \times B) = P_F(A)P_G(B).$$

Dans ce cas, lorsque un client quitte l'orbite pour accéder au service, un autre client bloqué sera en tête de l'orbite et un nouveau temps de rappel résiduel $r > 0$ ainsi que un nouveau temps service résiduel $y > 0$ seront générés à partir de G et F , respectivement, de sorte que la transition $(0, n, 0, 0) \rightarrow (1, n-1, y, r)$ se produise.

Pour les états e_1 et $x \in \mathbb{E}_1$, la mesure de transition Q_1 est déterminée comme suit.

$$Q_1(e_1; \{1\} \times \{1\} \times A \times B) = \mathbf{1}_A(y)P_G(B);$$

$$Q_1(x; \{1\} \times \{n+1\} \times A \times B) = \mathbf{1}_A(y)\mathbf{1}_B(r).$$

Avec $A, B \in \mathcal{B}_{(0,\infty)}$.

Enfin, la mesure Q_2 pour les états $x \in \partial^*\mathbb{E}_1$ est donnée par

$$Q_2(x; \{0\} \times \{n\} \times \{0\} \times A) = \mathbf{1}_A(r);$$

Où $A \in \mathcal{B}_{(0,\infty)}$.

Comme pour le modèle précédent, le taux de sauts est λ .

Martingales associées

Soit

$$\begin{aligned} \mathcal{G}f(0, 0, 0, 0, t) &= \lambda \int_0^\infty [f(1, 0, y, 0, t) - f(0, 0, 0, 0, t)] dF(y) \\ &\quad + \frac{\partial}{\partial t} f(0, 0, 0, 0, t); \end{aligned} \quad (3.14)$$

$$\begin{aligned} \mathcal{G}f(0, n, 0, r, t) &= \lambda \int_0^\infty [f(1, n, y, r, t) - f(0, n, 0, r, t)] dF(y) \\ &\quad - \frac{\partial}{\partial r} f(0, n, 0, r, t) + \frac{\partial}{\partial t} f(0, n, 0, r, t); \end{aligned} \quad (3.15)$$

$$\begin{aligned} \mathcal{G}f(1, 0, y, 0, t) &= \lambda \int_0^\infty [f(1, 1, y, r, t) - f(1, 0, y, 0, t)] dG(r) \\ &\quad - \frac{\partial}{\partial y} f(1, 0, y, 0, t) + \frac{\partial}{\partial t} f(1, 0, y, 0, t); \end{aligned} \quad (3.16)$$

$$\begin{aligned} \mathcal{G}f(1, n, y, r, t) &= \lambda [f(1, n+1, y, r, t) - f(1, n, y, r, t)] \\ &\quad - \frac{\partial}{\partial y} f(1, n, y, r, t) - \frac{\partial}{\partial r} f(1, n, y, r, t) \\ &\quad + \frac{\partial}{\partial t} f(1, n, y, r, t). \end{aligned} \quad (3.17)$$

le générateur infinitésimal du processus $\{(X(t), t)\}$ agissant sur les fonctions $f(i, n, y, r, t) \in \mathfrak{D}(\mathcal{G})$ qui sont différentiables par rapport à y, r et t pour tout i, n, y, r et t , qui satisfont

les conditions limites

$$f(0, n, 0, 0, t) = \int_0^\infty \int_0^\infty f(1, n-1, y, r, t) dF(y) dG(r); \quad (3.18)$$

$$f(1, n, 0, r, t) = f(0, n, 0, r, t); \quad (3.19)$$

et qui vérifient les conditions d'intégrabilité

$$\left| \int_0^\infty f(1, 0, y, 0, t) dF(y) - f(0, 0, 0, 0, t) \right| < \infty; \quad (3.20)$$

$$\left| \int_0^\infty f(1, n, y, r, t) dF(y) - f(0, n, 0, r, t) \right| < \infty; \quad (3.21)$$

$$\left| \int_0^\infty f(1, 1, y, r, t) dG(r) - f(1, 0, y, 0, t) \right| < \infty. \quad (3.22)$$

Les martingales associées au PDMP modélisant le modèle de file d'attente avec politique de rappels constante sont présentées dans les deux théorèmes suivants.

Théorème 3.4 (Meziani et Kernane [62]). *Pour $0 \leq z_1 \leq 1, 0 \leq z_2 \leq 1, \gamma > 0$ et $\delta > 0$ la fonction*

$$z_1^{C(t)} z_2^{N(t)} e^{-\gamma Y(t)} e^{-\delta R(t)} e^{\theta_{X(t)}(t)} \quad (3.23)$$

avec

$$\theta_{X(t)}(t) = \begin{cases} \lambda t + \ln(\omega_1) & \text{si } X(t) = e_0; \\ (\lambda - \delta)t + \ln(\omega_2) & \text{si } X(t) \in \mathbb{E}_0; \\ (\lambda - \gamma)t + \ln(\omega_3) & \text{si } X(t) = e_1; \\ (\lambda - \lambda z_2 - \gamma - \delta)t & \text{si } X(t) \in \mathbb{E}_0. \end{cases} \quad (3.24)$$

Où

$$\omega_1 = \frac{\gamma - \lambda z_1 \varphi_S(\gamma) \omega_3 + \lambda z_1 \varphi_S(\gamma) \omega_3 e^{-\gamma t}}{\gamma};$$

$$\omega_2 = \frac{\lambda z_2 + \gamma - \lambda z_1 \varphi_S(\gamma) + \lambda z_1 \varphi_S(\gamma) e^{-(\lambda z_2 + \gamma)t}}{\lambda z_2 + \gamma};$$

$$\omega_3 = \frac{\lambda z_2 + \delta - \lambda z_2 \hat{\varphi}_R(\delta) + \lambda z_2 \hat{\varphi}_R(\delta) e^{-(\lambda z_2 + \delta)t}}{\lambda z_2 + \delta}.$$

est une martingale pour les états $x \in \mathbb{E}$, où

$$\varphi_S(\gamma) = \int_0^\infty e^{-\gamma y} dF(y); \gamma > 0 \quad \text{et} \quad \hat{\varphi}_R(\delta) = \int_0^\infty e^{-\delta r} dG(r); \delta > 0.$$

Preuve. En définissant la fonction

$$f(i, n, y, r, t) = z_1^i z_2^n e^{-\gamma y} e^{-\delta r} e^{\theta_x(t)}, \quad (3.25)$$

où $\theta_x(t)$ est donnée par Eq.(3.24), nous avons

$$\begin{aligned} \mathcal{G}f(0, 0, 0, 0, t) &= \lambda \int_0^\infty [z_1 e^{-\gamma y} e^{\theta_{1,0,y,0}(t)} - e^{\theta_{(0,0,0,0)}(t)}] df(y) + \theta'_{(0,0,0,0)}(t) e^{\theta_{(0,0,0,0)}(t)} \\ &= \lambda z_1 \varphi_S(\gamma) \omega_3 e^{-(\gamma-\lambda)t} - \lambda \frac{e^{\lambda t}}{\gamma} \{ \gamma - \lambda z_1 \varphi_S(\gamma) \omega_3 + \lambda z_1 \varphi_S(\gamma) \omega_3 e^{-\gamma t} \} \\ &\quad + \left[\lambda - \frac{\lambda z_1 \varphi_S(\gamma) \omega_3 \gamma e^{-\gamma t}}{\gamma - \lambda z_1 \varphi_S(\gamma) \omega_3 + \lambda z_1 \varphi_S(\gamma) \omega_3 e^{-\gamma t}} \right] \\ &\quad \times \frac{e^{\lambda t}}{\gamma} \{ \gamma - \lambda z_1 \varphi_S(\gamma) \omega_3 + \lambda z_1 \varphi_S(\gamma) \omega_3 e^{-\gamma t} \} \\ &= \lambda z_1 \varphi_S(\gamma) \omega_3 e^{-(\gamma-\lambda)t} - \lambda \frac{e^{\lambda t}}{\gamma} \{ \gamma - \lambda z_1 \varphi_S(\gamma) \omega_3 + \lambda z_1 \varphi_S(\gamma) \omega_3 e^{-\gamma t} \} \\ &\quad + \lambda \frac{e^{\lambda t}}{\gamma} \{ \gamma - \lambda z_1 \varphi_S(\gamma) \omega_3 + \lambda z_1 \varphi_S(\gamma) \omega_3 e^{-\gamma t} \} - \lambda z_1 \varphi_S(\gamma) \omega_3 e^{-(\gamma-\lambda)t} \\ &= 0. \end{aligned}$$

$$\begin{aligned}
\mathcal{G}f(0, n, 0, r, t) &= \lambda \int_0^\infty [z_1 z_2^n e^{-\gamma y} e^{-\delta r} e^{\theta_{(1,n,y,r)}(t)} - z_2^n e^{-\delta r} e^{\theta_{(0,n,0,r)}(t)}] dF(y) + \delta z_2^n e^{-\delta r} e^{\theta_{(0,n,0,r)}(t)} \\
&\quad + z_2^n e^{-\delta r} \theta'_{(0,n,0,r)}(t) e^{\theta_{(0,n,0,r)}(t)} \\
&= \lambda z_1 z_2^n \varphi_S(\gamma) e^{-\delta r} e^{(\lambda - \lambda z_2^n - \gamma - \delta)t} + (\delta - \lambda) z_2^n e^{-\delta r} \omega_2 e^{(\lambda - \delta)t} \\
&\quad + z_2^n e^{-\delta r} \left\{ (\lambda - \delta) - \frac{\lambda z_1 \varphi_S(\gamma) e^{-(\lambda z_2 + \gamma)t}}{\omega_2} \right\} \times e^{(\lambda - \delta)t} \omega_2 \\
&= \lambda z_1 z_2^n \varphi_S(\gamma) e^{-\delta r} e^{(\lambda - \lambda z_2^n - \gamma - \delta)t} + (\delta - \lambda) z_2^n e^{-\delta r} \omega_2 e^{(\lambda - \delta)t} \\
&\quad + z_2^n e^{-\delta r} (\lambda - \delta) e^{(\lambda - \delta)t} \omega_2 - z_2^n e^{-\delta r} \lambda z_1 \hat{\varphi}_S(\gamma) e^{-(\lambda z_2 + \gamma)t} e^{(\lambda - \delta)t} \\
&= 0.
\end{aligned}$$

$$\begin{aligned}
\mathcal{G}f(1, 0, y, 0, t) &= \lambda \int_0^\infty [z_1 z_2 e^{-\gamma y} e^{-\delta r} e^{\theta_{(1,n,y,r)}(t)} - z_1 e^{-\gamma y} e^{\theta_{(1,0,y,0)}(t)}] dG(r) + \gamma z_1 e^{-\gamma y} e^{\theta_{(1,0,y,0)}(t)} \\
&\quad + z_1 e^{-\gamma y} \theta'_{(1,0,y,0)}(t) e^{\theta_{(1,0,y,0)}(t)} \\
&= \lambda z_1 z_2 e^{-\gamma y} \hat{\varphi}_R(\delta) e^{(\lambda - \lambda z_2 - \gamma - \delta)t} + (\gamma - \lambda) z_1 e^{-\gamma y} e^{(\lambda - \gamma)t} \omega_3 \\
&\quad + z_1 e^{-\gamma y} \left\{ (\lambda - \gamma) - \frac{\lambda z_2 \hat{\varphi}_R(\delta) e^{-(\lambda z_2 + \delta)t}}{\omega_3} \right\} e^{(\lambda - \gamma)t} \omega_3 \\
&= \lambda z_1 z_2 e^{-\gamma y} \hat{\varphi}_R(\delta) e^{(\lambda - \lambda z_2 - \gamma - \delta)t} + (\gamma - \lambda) z_1 e^{-\gamma y} e^{(\lambda - \gamma)t} \omega_3 \\
&\quad + z_1 e^{-\gamma y} (\lambda - \gamma) e^{-(\lambda z_2 + \delta)t} \omega_3 - z_1 e^{-\gamma y} \lambda z_2 \hat{\varphi}_R(\delta) e^{(\lambda - \gamma)t} \omega_3 \\
&= 0.
\end{aligned}$$

$$\begin{aligned}
\mathcal{G}f(1, n, y, n, t) &= \lambda [z_2 e^{\theta_{(1,n,y,r)}(t)} - e^{\theta_{(1,n,y,r)}(t)}] + \gamma e^{\theta_{(1,n,y,r)}(t)} + \delta e^{\theta_{(1,n,y,r)}(t)} + \theta'_{(1,n,y,r)}(t) e^{\theta_{(1,n,y,r)}(t)} \\
&= (\lambda z_2 - \lambda + \gamma + \delta) e^{(\lambda - \lambda z_2 - \gamma - \delta)t} + (\lambda - \lambda z_2 - \gamma - \delta) e^{(\lambda - \lambda z_2 - \gamma - \delta)t} \\
&= 0.
\end{aligned}$$

Ainsi, d'après la propriété du générateur infinitésimal, Eq.(3.23) est une martingale pour le processus $\{X(t)\}$. \square

Notons que $\theta_{X(t)}(t)$ est solution du système

$$\begin{cases}
\lambda z_1 \varphi_S(\gamma) e^{\theta_{x \in \mathbb{E}_1}(t)} - (\lambda - \delta) e^{\theta_{x \in \mathbb{E}_0}(t)} + \theta'_{x \in \mathbb{E}_0}(t) \theta_{x \in \mathbb{E}_0}(t) = 0; \\
\lambda z_1 \varphi_S(\gamma) e^{\theta_{x=e_1}(t)} - \lambda e^{\theta_{x=e_0}(t)} + \theta'_{x=e_0}(t) e^{\theta_{x=e_0}(t)} = 0; \\
\lambda z_2 \varphi_R(\delta) e^{\theta_{x \in \mathbb{E}_1}(t)} - (\lambda - \gamma) e^{\theta_{x=e_1}(t)} \theta'_{x=e_1}(t) e^{\theta_{x=e_1}(t)} = 0; \\
\lambda z_2 - \lambda + \gamma + \delta + \theta'_{x \in \mathbb{E}_1}(t) = 0.
\end{cases}$$

Afin d'énoncer le théorème suivant, nous définissons les variables aléatoires $\tilde{\tau}_0$ et $\tilde{\tau}_1$ comme étant les temps d'arrêt qui mettent fin à la période d'inactivité du serveur (lorsque $X(t) \in e_0 \cup \mathbb{E}_0$) et à la fin de la période d'occupation (lorsque $X(t) \in e_1 \cup \mathbb{E}_1$), respectivement.

Théorème 3.5 (Meziani et Kernane [62]). *Pour toute fonction $f \in \mathfrak{D}(\mathcal{G})$, les processus*

$$f(C(t), N(t), Y(t), R(t), t) - f(0, N(0), 0, R(0), 0) - \int_0^t \mathcal{G}f(0, N(s), 0, R(s), s) ds \quad (3.26)$$

et

$$f(C(t), N(t), Y(t), R(t), t) - f(1, N(0), Y(0), R(0), 0) - \int_0^t \mathcal{G}f(1, N(s), Y(s), R(s), s) ds \quad (3.27)$$

définis pour $t \in [0, \tilde{\tau}_0]$ et $t \in [0, \tilde{\tau}_1]$, respectivement, sont des martingales.

preuve. Le résultat est obtenu en exploitant la proposition 1.1 et le générateur infinitésimal donné par Eq.(3.14)-Eq.(3.17).

Notons que, pour $X(t) \in \{e_0\} \cup \{e_1\}$, $R(t) = 0$ lorsque $N(t) = 0$. □

Comme précédemment, les variables $\tilde{\tau}_0$ et $\tilde{\tau}_1$ peuvent être identifiées d'après le traitement du processus $\{X(t)\}$ dans le cadre d'un PDMP. De ce fait, $\tilde{\tau}_0 = \min(A(t), R(t))$ et $\tilde{\tau}_1 = Y(t)$.

Nombre moyen de clients dans l'orbite

Comme dans le cas de la politique de rappels classique, dans cette section, nous allons tirer l'espérance conditionnelle des clients bloqués en régime transitoire.

Théorème 3.6 (Meziani et Kernane [62]). *L'espérance conditionnelle du nombre de clients bloqués $N(t)$ sachant que $N(0) = n_0$, $Y(0) = y_0$ (lorsque $Y(t) \in e_1 \cup \mathbb{E}_1$) et $R(0) = r_0$ (lorsque $R(t) \in \mathbb{E}_0 \cup \mathbb{E}_1$) est donnée par*

$$E[N(t)|N(0) = n_0] = \begin{cases} 0 & \text{si } X(t) = e_0; \\ n_0 + \frac{1}{\mu_1 + \nu_1} (r_0 + (\lambda\mu_1 - 1)t - \nu_R(t)) & \text{si } X(t) \in \mathbb{E}_0; \end{cases}$$

pour $t \in [0, \tilde{\tau}_0]$ et par

$$E[N(t)|N(0) = n_0] = \begin{cases} 0 & \text{si } X(t) = e_1; \\ n_0 + \lambda t + \frac{1}{\mu_1 + \nu_1} (y_0 + r_0 - 2t - \mu_Y(t) - \nu_R(t)) & \text{si } X(t) \in \mathbb{E}_1; \end{cases}$$

pour $t \in [0, \tilde{\tau}_1]$.

Où $\nu_1 = \int_0^\infty rdG(r)$ est le moment d'ordre 1 de la fonction G et $\nu_R(t) = \frac{1}{\bar{G}(t)} \int_t^\infty \bar{G}(u) du$.

Preuve. Posons $f(i, n, y, r, t) = y + n\mu_1 + r + n\nu_1$. Les conditions (3.18) – (3.21) d'appartenance au domaine $\mathfrak{D}(\mathcal{G})$ sont vérifiées. D'où

$$\mathcal{G}f(0, n, 0, r, t) = \lambda\mu_1 - 1 \quad \text{et} \quad \mathcal{G}f(1, n, y, r, t) = \lambda(\mu_1 + \nu_1) - 2.$$

D'après le théorème 5 ci-dessus, $Y(t) + N(t)\mu_1 + R(t) + N(t)\nu_1 - n_0\mu_1 - n_0\nu_1 - r_0 - \int_0^t (\lambda\mu_1 - 1)ds$ et $Y(t) + N(t)\mu_1 + R(t) + N(t)\nu_1 - y_0 - n_0\mu_1 - r_0 - n_0\nu_1 - \int_0^t [\lambda(\mu_1 + \nu_1) - 2]ds$ sont des martingales. Ainsi, pour $t \in [0, \tilde{\tau}_0]$ et $X(t) \in \mathbb{E}_0$

$$E[N(t)|N(0) = n_0, R(0) = r_0] = n_0 + \frac{1}{\mu_1 + \nu_1} (r_0 + (\lambda\mu_1 - 1)t - E[Y(t)] - E[R(t)]) \quad (3.28)$$

et pour $t \in [0, \tilde{\tau}_1]$ et $X(t) \in \mathbb{E}_1$

$$E[N(t)|N(0) = n_0, Y(0) = y_0, R(0) = r_0] = n_0 + \lambda t + \frac{1}{\mu_1 + \nu_1} (y_0 + r_0 - 2t - E[Y(t)] - E[R(t)]). \quad (3.29)$$

En outre, nous avons

$$E[Y(t)] = \begin{cases} 0 & \text{pour } t \in [0, \tilde{\tau}_0]; \\ \mu_Y(t) & \text{pour } t \in [0, \tilde{\tau}_1]. \end{cases}$$

et

$$E[R(t)] = \begin{cases} 0 & \text{pour } X(t) \in e_0 \cup e_1; \\ \nu_R(t) & \text{pour } X(t) \in \mathbb{E}_0 \cup \mathbb{E}_1. \end{cases}$$

En substituant $E[Y(t)]$ et $E[R(t)]$ dans Eq.(3.28) et Eq.(3.29), le résultat est atteint. \square

Chapitre 4

Approximation par les processus de diffusion dans les files d'attente

Dans une file d'attente avec rappels, le nombre de clients dans l'orbite est une quantité stochastique représentée par un processus défini sur un espace des états discret qui est généralement l'ensemble des entiers naturels \mathbb{N} ou une partie de cet ensemble. Par conséquent, ce processus possède des trajectoires discontinues. En outre, si ce processus possède la propriété de Markov alors il est dit processus de Markov à sauts.

L'approximation d'un processus de Markov à sauts par un processus de diffusion consiste à approximer un tel processus de Markov par un processus avec un espace des états continu et des trajectoires presque sûrement continues tout en maintenant la propriété de Markov du processus initial. Une approximation des trajectoires discrètes par des trajectoires continues demeure justifiée car les tailles des sauts sont souvent infiniment petites par rapport à la taille totale du système modélisé. Il existe plusieurs méthodes d'approximation des processus de Markov à sauts par des processus de diffusion qui sont présentées de manière détaillée dans Fuchs [29].

Dans la théorie des files d'attente, l'analyse qui consiste à approximer le processus qui décrit le système par un processus de diffusion est connue sous le nom d'approximation dans le trafic intense (heavy traffic approximation). Ces approximations, également appelées théorèmes limites, sont réalisées sous certaines conditions sur les paramètres du modèles tels que le taux d'arrivée, le nombre de serveurs, la taille de la population, En effet, l'étude asymptotique d'un modèle de file d'attente sert, en particulier, à approximer les distributions données par des formules complexes par des distributions standards telles que la distribution normale et la distribution gamma relatives à des processus Browniens si les paramètres infinitésimaux satisfont certaines conditions.

Dans le travail de Falin [22], une file d'attente périodique a été étudiée dans un trafic intense en utilisant la méthode de convergence du générateur. Cette méthode, qui consiste à mettre en place un processus réduit en multipliant le processus à approximer par une quantité qui tends vers 0, est l'une des méthodes utilisées dans ce travail. La seconde mé-

thode considérée est basée sur l'utilisation de des propriétés asymptotiques des processus des arrivées et de service dans des files d'attente simples [40, 52, 53]. En outre, l'approximation dans le trafic intense a également été établie dans des modèles de files d'attente complexes. En supposant un faible taux de rappels dans la file M/M/c avec rappels, le nombre moyen de clients dans l'orbite a été approximé par un processus de diffusion dans Falin [23]. Dans le travail de Liu et al. [60], l'approximation par un processus de diffusion a été établie en considérant une séquence de files d'attente à double entrées.

4.1 Approximation dans un trafic intense pour des files Markoviennes simples

Nous considérons qu'il y a un trafic intense dans une file d'attente lorsque le taux d'arrivée augmente de telle sorte que la condition de stabilité du système s'approche de 1. Les approximations dans un trafic intense sont généralement établies pour le processus $\{N(t), t \geq 0\}$ décrivant le nombre de clients dans le système à l'instant t . Pour une file d'attente dépendante de l'état (c'est-à-dire la loi du processus des arrivées et la loi du processus de service dépendent de l'état du système), $\{N(t), t \geq 0\}$ est un processus de naissance et de mort défini sur un espace des états discret $\mathbb{E}_d = \{0, \dots, K\}$ ($K \leq \infty$). La dépendance est représentée par la moyenne $1/\lambda_n$ et la variance $\sigma_{a,n}$ du processus des arrivées et la moyenne $1/\mu_n$ et la variance $\sigma_{s,n}$ du processus du service. Les processus stochastiques considérés sont supposés être indépendants. D'après Kimura [52], l'idée de base des approximations par les processus de diffusion pour les files d'attente est d'approcher le processus évalué $\{N(t), t \geq 0\}$ par un processus de diffusion homogène dans le temps $\{X(t), t \geq 0\}$ avec une trajectoire continue sur un intervalle $\mathbb{E}_c = [x_0, x_K]$ ($-\infty \leq x_0 < x_K \leq +\infty$) en utilisant des propriétés asymptotiques des processus des arrivées et de service. Il a été démontré que ces approximations sont efficaces dans le cas du trafic intense [40, 52, 71].

4.1.1 Détermination de l'intervalle \mathbb{E}_c

Soient $E(t)$ le nombre de clients arrivants dans le système dans l'intervalle $]0, t]$ et $D(t)$ le nombre de clients sortants du système dans l'intervalle $]0, t]$. Alors, le nombre de clients dans le système à l'instant t est donné par

$$N(t) = N(0) + E(t) - D(t), \quad t \geq 0. \quad (4.1)$$

Soit I_n un intervalle de \mathbb{E}_c correspondant à l'évènement $P(N(t) = n), n \in \mathbb{E}_d$ tel que

$$I_0 = \{0\} \quad \text{et} \quad I_n =]x_{n-1}, x_n];$$

où $x_n \in \mathbb{E}_d$ est une séquence croissante avec

$$x_0 = 0 \quad \text{et} \quad \mathbb{E}_c = \bigcup_{n \in \mathbb{E}_d} I_n = [0, x_K[.$$

En général, les bornes de \mathbb{E}_c sont supposées être réfléchissantes (reflecting boundaries) pour que le processus de diffusion $\{X(t)\}$ soit régulier dans $[x_0, x_K]$ avec $-\infty \leq x_0 < x_K \leq +\infty$.

4.1.2 Calcul des paramètres de diffusion

Soit $dX(t) = X(t) - X(0)$ l'accroissement du processus $\{X(t)\}$ accumulé sur un l'intervalle infinitésimal $]0, t]$. Alors, pour tout $x \in \mathbb{E}_c$, le processus $\{X(t)\}$ peut être caractérisé par les limites suivantes :

$$\begin{aligned} a(x) &= \lim_{t \rightarrow 0} \frac{1}{t} E [dX(t) | X(0) = x]; \\ b(x) &= \lim_{t \rightarrow 0} \frac{1}{t} Var [dX(t) | X(0) = x]. \end{aligned} \quad (4.2)$$

Ces paramètres sont supposés être constants par morceaux, c'est-à-dire pour tout $x \in I_n$

$$\begin{aligned} a(x) &= a_n; \\ b(x) &= b_n. \end{aligned} \quad (4.3)$$

L'hypothèse des paramètres constants par morceaux est basée sur la correspondance naturelle entre les événements $N(t) = n$ et $X(t) \in I_n$. De plus, le processus possédant les paramètres Eq.(4.3) est bien défini et sa trajectoire est toujours continue.

Afin de calculer les paramètres de diffusion, nous allons faire appel au théorème élémentaire de renouvellement.

Par hypothèse $\{E(t), t \geq 0\}$ est un processus de renouvellements, alors d'après Eq.(1.3) nous avons

$$\begin{aligned} E[E(t)] &\approx \lambda_n t; \\ Var[E(t)] &\approx \lambda_n^3 \sigma_{a,n}^2 t. \end{aligned} \quad (4.4)$$

Le processus $\{D(t), t \geq 0\}$ n'est pas un processus de renouvellement mais dans le cas d'un trafic intense, le serveur sera occupé la plupart du temps. Ainsi, d'après Heyman [40], $\{D(t), t \geq 0\}$ peut être approximé par un processus $\{\tilde{D}(t), t \geq 0\}$ tel que

$$\begin{aligned} E[\tilde{D}(t)] &\approx \mu_n t; \\ Var[\tilde{D}(t)] &\approx \mu_n^3 \sigma_{s,n}^2 t. \end{aligned} \quad (4.5)$$

En remplaçant $D(t)$ par $\tilde{D}(t)$ dans Eq.(4.1) et en utilisant Eq.(4.4) et Eq.(4.5), nous obtenons les résultats approximatifs suivants

$$\frac{E[N(t)]}{t} \approx \lambda_n - \mu_n;$$

$$\frac{E[N(t)]}{t} \approx \lambda_n^3 \sigma_{a,n}^2 + \mu_n^3 \sigma_{s,n}^2.$$

Par conséquent, le processus $\{N(t), t \geq 0\}$ peut être approximé par un processus de diffusion $\{X(t), t \geq 0\}$ avec les paramètres infinitésimaux

$$\begin{aligned} a_n &= \lambda_n - \mu_n && \text{pour } n = 1, \dots, N; \\ b_n &= \lambda_n^3 \sigma_{a,n}^2 + \mu_n^3 \sigma_{s,n}^2 && \text{pour } n = 1, \dots, N. \end{aligned} \quad (4.6)$$

4.1.3 Application pour une file M/M/1 simple

Nous considérons un système d'attente où les clients arrivent un par un selon un processus de Poisson de paramètre λ . Les durées de service sont des variables aléatoires i.i.d d'une même loi exponentielle de paramètre μ . L'état du système à l'instant t est décrit par le processus $\{N(t), t \geq 0\}$ défini sur $\mathbb{E}_d = \mathbb{N}$. Sa condition de stabilité est $\rho = \frac{\lambda}{\mu} < 1$. Dans le trafic intense (c'est-à-dire lorsque $\rho \approx 1$), $\{N(t), t \geq 0\}$ peut être approximé par un processus de diffusion $\{X(t)\}$ défini sur $\mathbb{E}_c = \mathbb{R}^+$. La moyenne et la variance infinitésimales sont obtenues en posant $\lambda_n = \lambda$ et $\mu_n = \mu$ pour tout $n \in \mathbb{N}$ dans Eq.(4.6). Ainsi, les paramètres de diffusion sont donnés par

$$\begin{aligned} a_n &= \lambda - \mu && \text{pour } n \in \mathbb{N}; \\ b_n &= \lambda + \mu && \text{pour } n \in \mathbb{N}. \end{aligned} \quad (4.7)$$

Distribution stationnaire du nombre de clients dans le système

Soit $p(x, t|x_0, t_0)$ la densité de probabilité conditionnelle du processus $\{X(t)\}$ qui vérifie l'équation de fokker-Planck suivante.

$$\frac{\partial}{\partial t} p(x, t|x_0, t_0) = -(\lambda - \mu) \frac{\partial}{\partial x} p(x, t|x_0, t_0) + \frac{\lambda + \mu}{2} \frac{\partial^2}{\partial x^2} p(x, t|x_0, t_0). \quad (4.8)$$

Soit $p(x) = \lim_{t \rightarrow \infty} p(x, t|x_0, t_0)$ la distribution stationnaire du processus $\{X(t)\}$. Alors, en passant à la limite dans Eq.(4.8), nous obtenons

$$0 = -(\lambda - \mu) \frac{\partial}{\partial x} p(x) + \frac{\lambda + \mu}{2} \frac{\partial^2}{\partial x^2} p(x).$$

C'est une équation différentielle ordinaire (EDO) du second ordre dont la solution est donnée par

$$p(x) = A + Be^{2\frac{(\lambda-\mu)}{\lambda+\mu}x};$$

avec $\int_0^{\infty} p(x) = 1$, $\lim_{x \rightarrow 0} p(x) = 0$. et A, B sont des constantes réelles.

D'où

$$p(x) = \frac{2(\mu - \lambda)}{\lambda + \mu} e^{-\frac{2(\mu-\lambda)}{\lambda+\mu}x}, \quad x \geq 0.$$

C'est une densité de probabilité d'une loi exponentielle de paramètre $\frac{2(\mu-\lambda)}{\lambda+\mu}$. Ce résultat est attendu puisque le nombre de clients dans le système d'une file M/M/1 suit une loi géométrique qui est l'analogie discret de la loi exponentielle.

Notons que, dans le trafic intense, la condition de stabilité du système ρ s'approche intensivement de 1 mais elle y reste toujours strictement inférieure. Ainsi, comme $\lambda < \mu$ alors $\mu - \lambda > 0$.

Nombre moyen approximatif des clients dans le système

Soit X la variable aléatoire qui approxime le nombre de clients dans le système de la file M/M/1 à l'équilibre. Nous avons vu précédemment que X suit une loi exponentielle de paramètre $2(\mu - \lambda)/\lambda + \mu$. Pour une loi exponentielle de paramètre ν , la moyenne est donné par $1/\nu$. Ainsi,

$$E(X) = \frac{\lambda + \mu}{2(\mu - \lambda)} = \frac{1 + \rho}{2(1 - \rho)}.$$

Comme $\rho \approx 1$ alors

$$E(X) \approx \frac{2\rho}{2(1 - \rho)} = \frac{\rho}{1 - \rho} = E(N).$$

Avec N est la variable aléatoire qui représente le nombre de clients dans le système de la file M/M/1 à l'équilibre et $E(N)$ est son espérance mathématique (Voir Eq.(2.3)).

4.2 Approximation dans un trafic intense pour une file périodique

Une file périodique est un système d'attente dans lequel les clients arrivent selon un processus de Poisson périodique. Le phénomène de périodicité en file d'attente se produit généralement dans les copieurs ou les duplicateurs lors de l'arrivée de la charge de travail. Ces modèles ont été étudiés auparavant par Takacs [77], Reich [72, 73], Hasofer [39] et Harrison et Lemoine [38]. Les recherches sur les files périodiques ont montré que ces systèmes sont caractérisés par l'existence d'un régime stationnaire périodique au lieu d'un

régime stationnaire dans son sens ordinaire (Voir Falin [22]). Néanmoins, la détermination de la distribution stationnaire périodique est très compliquée et ne peut se faire qu'à l'aide d'un ordinateur.

Dans ce qui suit, nous présentons une méthode approximative par les processus de diffusion établie par Falin [22]. Cette approximation a été réalisée dans le cas d'un trafic intense en utilisant le principe de la moyenne qui stipule qu'une file périodique est approximée par la file stationnaire correspondante avec le taux d'intensité cumulé Λ sous certaines conditions.

Le processus étudié est le temps d'attente virtuel $V(t)$ dans la file $M(t)/G/1$. Les clients arrivent dans ce système selon un processus de Poisson périodique avec la fonction d'intensité $\lambda(t)$. Les clients sont servis suivant leur ordre d'arrivée et les durées de service sont des variables aléatoires i.i.d de la même distribution générale F de moyenne $\mu_1 = \int_0^\infty y dF(y)$ et de variance σ^2 finie. La condition de stabilité de ce système est $\Lambda\mu_1 < 1$ avec $\Lambda(t) = \int_0^t \lambda(s) ds$ et $\Lambda = \Lambda(1)$.

Pour effectuer l'approximation présentée dans Falin [22], l'auteur a considéré une séquence de n files d'attente $M(t)/G/1$ avec des fonctions d'intensité $\lambda_n(t)$. Les distributions de service sont supposées être indépendantes de n .

Soit $\tau(t)$ la partie fractionnelle de t . Cette variable supplémentaire est introduite afin d'obtenir un processus de Markov homogène $(V(t), \tau(t))$ défini sur $\mathbb{E} = \mathbb{R}_+ \times [0, 1]$.

En partant d'un état initial (v, t) et après un temps infinitésimal h , le processus $(V(t), \tau(t))$ passe vers l'état $(v-h, \tau+h)$ avec la probabilité $(1-\lambda(\tau)h)$, et avec la probabilité $\lambda(\tau)h$ il passe vers l'état $(v-h+Y, \tau+h)$ où Y est une variable aléatoire générée de la distribution F . Ainsi, le générateur infinitésimal associé s'écrit

$$\mathcal{G}f(v, \tau) = \frac{\partial}{\partial \tau} f(v, \tau) - \frac{\partial}{\partial v} f(v, \tau) + \lambda(\tau) \int_0^\infty [f(v+y, \tau) - f(v, \tau)] dF(y), \text{ pour } v > 0, 0 < \tau < 1.$$

Pour le point initial $(0, \tau)$, le générateur infinitésimal est donné par

$$\mathcal{G}f(0, \tau) = \frac{\partial}{\partial \tau} f(0, \tau) + \lambda(\tau) \int_0^\infty [f(y, \tau) - f(0, \tau)] dF(y), \text{ pour } 0 < \tau < 1.$$

Le domaine $\mathfrak{D}(\mathcal{G})$ est constitué des fonctions f qui sont différentiables par rapport à v et τ et qui satisfont la condition limite

$$f(v, 1) = f(v, 0).$$

Cette condition signifie que le processus est étudié sur une période égale à 1.

Soit $\left(\frac{V(nt)}{\sqrt{n}}, \tau(nt)\right)$ le processus réduit défini sur l'espace $\mathbb{E} = \mathbb{R}_+ \times [0, 1]$. Le générateur infinitésimal associé est donné par

$$\begin{aligned}
\mathcal{G}_n f(v, \tau) &= n\lambda(\tau) \int_0^\infty \left[f\left(v + \frac{y}{\sqrt{n}}, \tau\right) - f(v, \tau) \right] dF(y) + n \frac{\partial}{\partial \tau} f(v, \tau) - \sqrt{n} \frac{\partial}{\partial v} f(v, \tau) + o(1); \\
&= n\lambda(\tau) \int_0^\infty \left[\frac{y}{\sqrt{n}} \frac{\partial}{\partial v} f(v, \tau) \right] dF(y) + n \frac{\partial}{\partial \tau} f(v, \tau) - \sqrt{n} \frac{\partial}{\partial v} f(v, \tau) + o(1); \\
&= \sqrt{n}\lambda(\tau) \frac{\partial}{\partial v} f(v, \tau) \mu_1 + \frac{1}{2} \lambda(\tau) \frac{\partial^2}{\partial v^2} f(v, \tau) \mu_2 - \frac{1}{2} \lambda(\tau) \frac{\partial^2}{\partial v^2} f(v, \tau) \mu_2 + n \frac{\partial}{\partial \tau} f(v, \tau) \\
&\quad - \sqrt{n} \frac{\partial}{\partial v} f(v, \tau) + o(1); \\
&= \sqrt{n}\lambda(\tau) \frac{\partial}{\partial v} f(v, \tau) \mu_1 + \frac{1}{2} \lambda(\tau) \frac{\partial^2}{\partial v^2} f(v, \tau) \mu_2 + n \frac{\partial}{\partial \tau} f(v, \tau) - \sqrt{n} \frac{\partial}{\partial v} f(v, \tau) + o(1);
\end{aligned}$$

pour $v > 0$ et $0 < \tau < 1$ avec $\mu_2 = \int_0^\infty y^2 dF(y)$, et

$$\mathcal{G}_n f(0, \tau) = \sqrt{n}\lambda(\tau) \frac{\partial}{\partial v} f(0, \tau) \mu_1 + \frac{1}{2} \lambda(\tau) \frac{\partial^2}{\partial v^2} f(0, \tau) \mu_2 + n \frac{\partial}{\partial \tau} f(0, \tau) + o(1).$$

La condition de différentiabilité et la condition limite ne changent pas pour le domaine $\mathfrak{D}(\mathcal{G}_n)$.

Pour toute fonction de classe C^2 (c'est-à-dire sa dérivée première est continue et dérivable et sa dérivée seconde est continue), définissons la fonction

$$f_n(v, \tau) = f(v) + \frac{1}{\sqrt{n}} f'(v) g(\tau) + \frac{1}{n} f''(v) h(\tau);$$

où les fonctions $g(\tau)$ et $h(\tau)$ seront déterminées ultérieurement.

Pour ces fonctions, le générateur infinitésimal est alors donné par

$$\begin{aligned}
\mathcal{G}_n f_n(v, \tau) &= \sqrt{n} f'(v) [\lambda(\tau) \mu_1 - 1 + g'(\tau)] + f''(v) \left[(\lambda(\tau) \mu_1 - 1) g(\tau) + \frac{1}{2} \lambda(\tau) \mu_2 + h'(\tau) \right] + o(1); \\
\mathcal{G}_n f_n(0, \tau) &= \sqrt{n} f'(0) [\lambda(\tau) \mu_1 + g'(\tau)] + f''(0) \left[\lambda(\tau) \mu_1 g(\tau) + \frac{1}{2} \lambda(\tau) \mu_2 + h'(\tau) \right] + o(1);
\end{aligned}$$

et la condition limite devient

$$g(1) = g(0), \quad h(1) = h(0).$$

Dans le trafic intense (c'est-à-dire lorsque $\lim_{n \rightarrow \infty} \Lambda_n = \frac{1}{\mu_1}$), la convergence vers un processus de diffusion ne se fera que sur la première coordonnée du processus $(V(t), \tau(t))$. Les fonctions g et h doivent vérifier les équations

$$\lambda(\tau)\mu_1 - 1 + g'(\tau) = \frac{1}{\sqrt{n}}c_1; \quad (4.9)$$

$$(\lambda(\tau)\mu_1 - 1)g(\tau) + \frac{1}{2}\lambda(\tau)\mu_2 + h'(\tau) = c_2. \quad (4.10)$$

Pour obtenir une approximation indépendante de τ , les constantes c_1 et c_2 seront déterminées à partir des conditions limites.

L'équation (4.9) permet d'obtenir

$$g(\tau) = \tau - \Lambda(\tau)\mu_1 + \frac{1}{\sqrt{n}}c_1\tau + g(0).$$

De la condition limite $g(1) = g(0)$ nous obtenons

$$c_1 = -\sqrt{n}(1 - \Lambda_n\mu_1).$$

Ainsi,

$$g(\tau) = \rho_n\tau - \Lambda(\tau)\mu_1 + g(0);$$

avec $\rho_n = \Lambda_n\mu_1$.

Ce qui implique

$$h(\tau) = c_2\tau + \frac{1}{2}\Lambda_\tau\mu_2 + \int_0^\tau (1 - \lambda(s)\mu_1)g(s)ds + h(0).$$

À partir de la condition limite $h(1) = h(0)$, on obtient

$$c_2 = \frac{\mu_2}{2\mu_1}.$$

D'après Falin[22], $\lim_{n \rightarrow \infty} -\sqrt{n}(1 - \rho_n) = -m + o(1)$. Par conséquent, en utilisant les fonctions $g(\tau)$ et $h(\tau)$, nous obtenons

$$\lim_{n \rightarrow \infty} \mathcal{G}_n f_n(v, \tau) = -mf'(v) + \frac{\mu_2}{2\mu_1}f''(v) = \mathcal{G}_0 f(v).$$

En suivant la même démarche pour le point $(0, \tau)$, on montre que les $f \in \mathfrak{D}(\mathcal{G}_t)$ doivent vérifier la condition $f(0) = 0$. Ceci implique que le générateur \mathcal{G}_0 correspond à un processus de Wiener régularisé à l'origine avec une moyenne infinitésimale m et une variance infinitésimale $\mu_2/2\mu_1$.

4.3 Approximation dans les files d'attente avec une distribution exponentielle des temps de rappels

Le nombre de clients bloqués dans l'orbite est une quantité de performance importante pour les files d'attente avec rappels. Dans le cas d'une distribution des temps de rappels exponentielle, plusieurs travaux ont présenté des formules explicites pour cette quantité [26, 28, 5]. Tandis que dans le cas d'une distribution des temps de rappels générale, la majorité des travaux ont porté que sur des études approximatives [8, 74, 75]. Cependant, dans le travail de Meziani et Kernane [62], les auteurs ont pu tirer une expression explicite du nombre moyen de clients bloqués dans l'orbite en régime transitoire pour la file M/G/1 avec une distribution des temps de rappels générale étudiée à travers les PDMPs.

Nous intéressons dans ce qui suit à l'approximation par les processus de diffusion dans des systèmes de file d'attente avec une distribution des temps de rappels exponentielle.

4.3.1 File M/M/c dans le cas de taux de rappels faible

Le comportement asymptotique d'une file d'attente avec rappels est souvent influencé par le taux de rappels. Pour plusieurs raisons, dans les applications, le cas le plus important est le franchissement de la limite $\nu \rightarrow 0$. Pour le modèle de base à plusieurs serveurs décrit le chapitre 2 de ce travail, ce problème a été considéré en premier lieu par Cohen [11] qui a étudié le comportement asymptotique du nombre moyen de clients dans l'orbite. Les investigations dans le régime permanent ($\lambda < c$) ont prouvé que la limite $\lim_{\mu \rightarrow 0} E(N) = a$ existe et elle est unique, où a est la racine positive de l'équation algébrique

$$\lambda \frac{(\lambda + a)^c}{c!} = a \sum_{i=0}^{c-1} \frac{(\lambda + a)^i}{i!}; \quad (4.11)$$

avec $a < \frac{\lambda}{(c-\lambda)}$. Ensuite en 1985, Falin [21] a montré que la distribution du nombre de clients dans l'orbite en régime stationnaire est asymptotiquement gaussienne de moyenne $\frac{a}{\nu}$ et de variance $\frac{1}{\nu} \times \left\{ a + \frac{a^3}{\lambda - a(c-\lambda)} \frac{c!}{(a+\lambda)^{c+1}} \sum_{i=0}^{c-1} \frac{i!}{(a+\lambda)^i} \left(\sum_{j=0}^i \frac{(\lambda+a)^j}{j!} \right)^2 \right\}$.

Dans ce qui suit, nous présentons l'étude approximative du nombre de clients dans l'orbite établie par Falin [23]. Pour ce faire, l'auteur a considéré le processus centré réduit $(C(\frac{t}{\nu}), X(\frac{t}{\nu})) = (C(\frac{t}{\nu}), \sqrt{\nu}N(\frac{t}{\nu}) - \frac{a}{\sqrt{\nu}})$ à l'équilibre ($\lambda < c$) et lorsque $\nu \rightarrow 0$. Le défi principal rencontré lors de cette approximation est le fait que seule la deuxième coordonnée du processus, qui correspond au nombre de clients dans l'orbite, converge vers un processus de diffusion. Pour remédier à ce problème, l'auteur a utilisé la méthode de démonstration des théorèmes de limite fonctionnelle en théorie des files d'attente (Voir Burman [7]).

Le générateur infinitésimal du processus $\left(C\left(\frac{t}{\nu}\right), \sqrt{\nu}N\left(\frac{t}{\nu}\right) - \frac{a}{\sqrt{\nu}}\right)$ défini sur l'espace des états $\mathbb{E} = \{0, 1, \dots, c\} \times]0, +\infty[$ est déterminé comme suit. En partant d'un état initial $(i, x) \in \mathbb{E}$ avec $i < c$ et après un temps infinitésimal h , les transitions possibles du processus $\left(C\left(\frac{t}{\nu}\right), X\left(\frac{t}{\nu}\right)\right)$ sont :

- $(i, x) \longrightarrow (i + 1, x)$ avec la probabilité $\lambda \frac{t}{\nu}$;
- $(i, x) \longrightarrow (i + 1, x - \sqrt{\nu})$ avec la probabilité $\left(\frac{x}{\sqrt{\nu}} + \frac{a}{\nu}\right) t$;
- $(i, x) \longrightarrow (i - 1, x)$ avec la probabilité $i \frac{t}{\nu}$.

Les probabilités de transition ci-dessus sont obtenues en considérant le changement de variables suivant. Le temps passe de t à $\frac{t}{\nu}$ et le nombre de clients dans l'orbite passe de n à $\frac{x}{\sqrt{\nu}} + \frac{a}{\nu}$ pour chaque réalisation du processus $\left(C\left(\frac{t}{\nu}\right), X\left(\frac{t}{\nu}\right)\right)$.

Ainsi, le générateur infinitésimal est donné par

$$\mathcal{G}_\nu f(i, x) = \frac{\lambda}{\nu} [f(i+1, x) - f(i, x)] + \left(\frac{x}{\sqrt{\nu}} + \frac{a}{\nu}\right) [f(i+1, x - \sqrt{\nu}) - f(i, x)] + \frac{i}{\nu} [f(i-1, x) - f(i, x)],$$

pour $i < c$ et par

$$\mathcal{G}_\nu f(c, x) = \frac{\lambda}{\nu} [f(c, x + \sqrt{\nu}) - f(c, x)] + \frac{c}{\nu} [f(c-1, x) - f(c, x)],$$

pour l'état initial (c, x) .

Le domaine $\mathfrak{D}(\mathcal{G}_\nu)$ du générateur infinitésimal \mathcal{G}_ν se constitue des fonctions f qui sont différentiables par rapport à x pour tout $0 \leq i \leq c$.

Définissons maintenant la fonction

$$f_\nu(i, x) = f(x) + \sqrt{\nu} f'(x) g_i + \nu f''(x) h_i,$$

où les constantes g_i et h_i seront déterminées ultérieurement.

Pour ces fonctions, le g n rateur infinitesimal s' crit alors

$$\begin{aligned}
\mathcal{G}_\nu f_\nu(i, x) &= \frac{1}{\sqrt{\nu}} f'(x) [(a + \lambda)(g_{i+1} - g_i) - a + i(g_{i-1} - g_i)] \\
&\quad + x f'(x)(g_{i+1} - g_i - 1) \\
&\quad + f''(x) \left[(a + \lambda)(h_{i+1} - h_i) + \frac{a}{2} - a g_{i+1} + i(h_{i-1} - h_i) \right] \\
&\quad + o(1), \quad \text{pour } 0 \leq i \leq c - 1; \\
\mathcal{G}_\nu f_\nu(c, x) &= \frac{1}{\sqrt{\nu}} f'(x) [\lambda + c(g_{c-1} - g_c)] \\
&\quad + f''(x) \left[\frac{\lambda}{2} + \lambda g_c + c(h_{c-1} - h_c) \right] \\
&\quad + o(1).
\end{aligned}$$

Pour $\nu \rightarrow 0$, le g n rateur $\mathcal{G}_\nu f_\nu(i, x)$ converge vers une limite qui ne d pend pas de i si les fonctions g_i et h_i sont choisies de telle sorte que, pour une certaine fonction $B(x)$, nous avons

$$(a + \lambda)(g_{i+1} - g_i) - a + i(g_{i-1} - g_i) = 0 \quad \text{pour } 0 \leq i \leq c - 1; \quad (4.12)$$

$$\lambda + c(g_{c-1} - g_c) = 0; \quad (4.13)$$

$$\begin{aligned}
x f'(x)(g_{i+1} - g_i - 1) + f''(x) \left\{ (a + \lambda)(h_{i+1} - h_i) \right. \\
\left. + \frac{a}{2} - a g_{i+1} + i(h_{i-1} - h_i) \right\} = B(x) \quad 0 \leq i \leq c - 1;
\end{aligned} \quad (4.14)$$

$$f''(x) \left\{ \frac{\lambda}{2} + \lambda g_c + c(h_{c-1} - h_c) \right\} = B(x). \quad (4.15)$$

Le syst me Eq.(4.12)-Eq.(4.13) admet une solution si et seulement si le param tre a satisfait Eq.(4.11).

Ainsi,

$$g_i = g_0 + \frac{a}{\lambda + a} \sum_{j=0}^{i-1} \frac{j!}{(\lambda + a)^j} \sum_{k=0}^j \frac{(\lambda + a)^k}{k!} \quad 0 \leq i \leq c - 1.$$

Le syst me Eq.(4.14)-Eq.(4.15) admet une solution si et seulement si

$$\begin{aligned}
B(x) &= \left\{ \left(\frac{\lambda}{2} + g_c \right) f''(x) \frac{(\lambda + a)^c}{c!} + \sum_{j=0}^{c-1} \frac{(a + \lambda)^j}{j!} \right. \\
&\quad \left. \left[x f'(x)(g_{j+1} - g_j - 1) + f''(x) \left(\frac{a}{2} - a g_{j+1} \right) \right] \right\} \times \left(\sum_{j=0}^c \frac{(a + \lambda)^j}{j!} \right)^{-1}.
\end{aligned} \quad (4.16)$$

Pour les fonctions g_i et h_i ainsi choisies, nous avons $\lim_{\nu \rightarrow 0} \mathcal{G}_\nu f_\nu(i, x) = B(x)$.
 En utilisant Eq.(4.11), l'expression de Eq.(4.16) s'écrit donc

$$\begin{aligned}
 B(x) &= -\frac{\lambda - a(c - \lambda)}{\lambda + a} x f'(x) \\
 &+ f''(x) \left\{ \frac{a}{\lambda + a} (\lambda - a(c - \lambda)) + \frac{a^3 c!}{(\lambda + a)^{c+2}} \sum_{j=0}^{c-1} \frac{j!}{(\lambda + a)^j} \left(\sum_{k=0}^j \frac{(\lambda + a)^k}{k!} \right)^2 \right\} \\
 &= -\alpha x f'(x) + \frac{\sigma^2}{2} f''(x) \\
 &= \mathcal{G}_0 f_0(x).
 \end{aligned}$$

$\mathcal{G}_0 f_0(x)$ est le générateur correspondant à un processus de diffusion avec une moyenne infinitésimale $-\alpha x$ et une variance infinitésimale σ^2 avec

$$\alpha = \frac{\lambda - a(c - \lambda)}{\lambda + a};$$

$$\sigma^2 = 2 \times \left\{ \frac{a}{\lambda + a} (\lambda - a(c - \lambda)) + \frac{a^3 c!}{(\lambda + a)^{c+2}} \sum_{j=0}^{c-1} \frac{j!}{(\lambda + a)^j} \left(\sum_{k=0}^j \frac{(\lambda + a)^k}{k!} \right)^2 \right\}.$$

4.3.2 File M/M/1 avec une politique de rappels classique et une source finie de clients

Considérons le système de file d'attente de type M/M/1 avec une source de clients finie de taille K décrit dans [76] comme suit. Si le serveur est inactif, alors le client entrant reçoit son service immédiatement et quitte le système après l'achèvement du service. Sinon, si le serveur est occupé, le client arrivé rejoint l'orbite. Les arrivées se présentent dans le système selon un processus de poisson avec un taux λ/K . Les temps de service sont indépendants et exponentiellement distribués avec une moyenne de $1/\mu$. Après un temps aléatoire dans l'orbite, les clients bloqués répètent leurs tentatives d'entrée en service indépendamment les uns des autres et les rappels sont générés selon un processus de Poisson avec un taux ν/K . Les temps d'inter-arrivée, les périodes de service et les temps de rappel sont supposés être mutuellement indépendants.

Soit $(C(t), N(t))$ le processus de Markov décrivant l'état du système à l'instant t . La composante $C(t)$ représente l'état du serveur ($C(t)$ est égal à 0 ou 1 selon que le serveur est libre ou occupé) et $N(t)$ est le nombre de clients bloqués où $N(t) \in \{0, 1, \dots, K - 1\}$. Le générateur infinitésimal du processus $(C(t), N(t))$ est donné par

$$\mathcal{G}f(0, n) = \frac{\lambda}{K}(K - n)[f(1, n) - f(0, n)] + \frac{\nu}{K}n[f(1, n - 1) - f(0, n)]; \quad (4.17)$$

$$\mathcal{G}f(1, n) = \frac{\lambda}{K}(K - n - 1)[f(1, n + 1) - f(1, n)] + \mu[f(0, n) - f(1, n)]. \quad (4.18)$$

Nous allons présenter, dans cette section, la démarche de Meziani et Kernane [63] suivie pour montrer la convergence du générateur donné par Eq.(4.17) et Eq.(4.18) vers un générateur infinitésimal qui a la forme suivante

$$\mathcal{G}f(x) = a(x)f'(x) + \frac{b(x)}{2}f''(x). \quad (4.19)$$

Approximation par un processus de diffusion

Définissons le processus réduit suivant $X(t) = \frac{N(Kt)}{\sqrt{K}}$. Ce processus est défini sur l'espace des états continu $[0, \dots, \frac{K-1}{\sqrt{K}}]$. Ainsi, il est possible de l'approximer par un processus de diffusion lorsque $K \rightarrow \infty$ comme le nombre de clients bloqués augmentent avec la croissance de la taille de la source des clients.

Le générateur infinitésimal du processus $(C(Kt), X(t))$ a la forme

$$\begin{aligned} \mathcal{G}_K f(0, x) &= \lambda K(1 - x)[f(1, x) - f(0, x)] + \nu x K \left[f \left(1, x - \frac{1}{\sqrt{K}} \right) - f(0, x) \right]; \\ \mathcal{G}_K f(1, x) &= \lambda K \left(1 - x - \frac{1}{K} \right) \left[f \left(1, x + \frac{1}{\sqrt{K}} \right) - f(1, x) \right] + \mu K [f(0, x) - f(1, x)]. \end{aligned}$$

De même qu'en Falin [25] et Falin et Templeton [26], pour une fonction arbitraire deux fois différentiable f_K , on définit la fonction

$$f_K(i, x) = f(x) + \frac{1}{\sqrt{K}}f'(x)c_i + \frac{1}{K}f''(x)d_i, \quad (4.20)$$

Où c_i et d_i pour $i \in \{0, 1\}$ sont des constantes.

Pour ces fonctions $f_K(i, x)$, lorsque $K \rightarrow \infty$, nous avons

$$\begin{aligned} \mathcal{G}_K f_K(0, x) &= \sqrt{K}f'(x)[\lambda(c_1 - c_0)] + f''(x)[\lambda(d_1 - d_0)] + o(1); \\ \mathcal{G}_K f_K(1, x) &= \sqrt{K}f'(x)[\lambda + \mu(c_1 - c_0)] + f''(x)[\lambda(c_1 - c_0) + \mu(d_1 - d_0)] + o(1). \end{aligned}$$

D'après ce qui a précédé, $\mathcal{G}_K f_K$ peut converger pour $K \rightarrow \infty$ vers une limite qui ne dépend pas de i si c_i et d_i satisfont aux conditions suivantes

$$\lambda(c_1 - c_0) = \frac{a(x)}{\sqrt{K}}. \quad (4.21)$$

$$\lambda(d_1 - d_0) = \frac{b(x)}{2}. \quad (4.22)$$

$$[\lambda + \mu(c_1 - c_0)] = \frac{a(x)}{\sqrt{K}}. \quad (4.23)$$

$$\lambda(c_1 - c_0) + \mu(d_1 - d_0) = \frac{b(x)}{2}. \quad (4.24)$$

La résolution du système des équations Eq.(4.21)-Eq.(4.24) nous permet de trouver que

$$a(x) = -\sqrt{K} \frac{\lambda\rho}{1-\rho}. \quad (4.25)$$

$$b(x) = 2\lambda \left(\frac{\rho}{1-\rho} \right)^2. \quad (4.26)$$

Où $\rho = \frac{\lambda}{\mu}$.

En utilisant Eq.(4.25) et Eq.(4.26), nous obtenons

$$\lim_{K \rightarrow \infty} \mathcal{G}_K f_K(i, x) = -\sqrt{K} \frac{\lambda\rho}{1-\rho} f'(x) + \lambda \left(\frac{\rho}{1-\rho} \right)^2 f''(x) = \mathcal{G}_0 f(x).$$

Le générateur \mathcal{G}_0 correspond à un mouvement brownien de moyenne infinitésimale $-\sqrt{K} \frac{\lambda\rho}{1-\rho}$ et de variance infinitésimale $2\lambda \left(\frac{\rho}{1-\rho} \right)^2$.

Distribution stationnaire

Soit $p(x)$ la distribution stationnaire associée au processus de diffusion $\{X(t)\}$. Comme l'a montré Newell [67], la fonction $p(x)$ est de la forme

$$p(x) = \frac{cte}{b(x)} \exp(\gamma(x)); \quad 0 < x < x_{K-1}. \quad (4.27)$$

Où $\gamma(x) = 2 \int_0^x \frac{a(u)}{b(u)} du$.

Comme $x_{K-1} = \frac{K-1}{\sqrt{K}} \rightarrow \infty$ lorsque $K \rightarrow \infty$, alors en substituant Eq.(4.25) et Eq.(4.26) dans Eq.(4.27), nous obtenons

$$p(x) = \frac{cte}{2\lambda} \left(\frac{1-\rho}{\rho} \right)^2 \exp \left(-\sqrt{K} \left(\frac{1-\rho}{\rho} \right) x \right); \quad x > 0.$$

La constante est déterminée en utilisant la condition de normalisation $\int_0^{\infty} p(x)dx = 1$.

Ainsi,

$$p(x) = \sqrt{K} \left(\frac{1-\rho}{\rho} \right) \exp \left(-\sqrt{K} \left(\frac{1-\rho}{\rho} \right) x \right); \quad x > 0. \quad (4.28)$$

Donc, pour $K \rightarrow \infty$, la distribution stationnaire du mouvement brownien approximant le nombre de clients bloqués dans l'orbite correspond à une distribution exponentielle de paramètre $\sqrt{K} \left(\frac{1-\rho}{\rho} \right)$.

4.3.3 File M/M/1 avec une politique de rappels constante

Le modèle considéré est une file d'attente infinie et d'un serveur unique. Les clients arrivent selon un processus de Poisson de taux λ . Les durées de service sont i.i.d d'une même loi exponentielle de paramètre μ . Si le serveur est libre au moment d'une arrivée alors le client est servi immédiatement et quitte le système à la fin du son service. Sinon, si le serveur est occupé, le client arrivant se dirige vers l'orbite. Le client en tête de l'orbite rappelle pour le service et les temps de rappels se produisent selon un processus de Poisson de taux ν . Les inter-arrivées, les temps de service et les temps de rappels sont mutuellement indépendants. D'après Fayolle [28], la charge du système est donnée par $\rho = \frac{\lambda(\lambda+\nu)}{\mu\nu}$ et le système est stable si $\rho < 1$.

Approximation par un processus de diffusion

Le nombre de clients dans l'orbite est décrit par le processus $\{N(t), t \geq 0\}$ défini sur l'espace des états $\mathbb{E}_d = \mathbb{N}$. La probabilité que le serveur soit occupé est ρ et la probabilité que le serveur soit libre est $1 - \rho$. Par conséquent, $\lambda\rho$ est le flux d'entrée dans l'orbite et $\nu(1 - \rho)$ est le flux de sortir de l'orbite (voir figure 4.1).

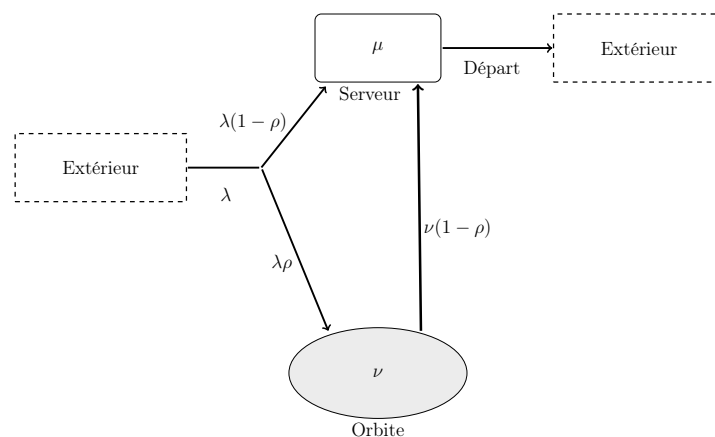


FIGURE 4.1 – File M/M/1 avec politique de rappels constante.

Le processus $\{N(t), t \geq 0\}$ peut être vu comme un processus de naissance et de mort avec un taux de naissance $\lambda\rho$ et un taux de mortalité $\nu(1 - \rho)$. En supposant que $E(t)$ est le nombre de clients entrant dans l'orbite dans l'intervalle du temps $]0, t]$ et $D(t)$ le nombre de clients sortant de l'orbite dans $]0, t]$, alors :

$$N(t) = N(0) + E(t) - D(t), \quad t \geq 0. \quad (4.29)$$

Détermination de l'intervalle \mathbb{E}_c

Soit I_n un intervalle de \mathbb{E}_c correspondant à l'évènement $P(N(t) = n), n \in \mathbb{E}_d = \mathbb{N}$ tel que

$$I_0 = \{0\} \quad \text{et} \quad I_n =]n - 1, n];$$

Ainsi,

$$E_c = \bigcup_{n \in \mathbb{N}_d} I_n = [0, \infty[.$$

Calcul des paramètres de diffusion

D'après Heyman [40], Kimura [52] et Pujolle et Wu [71], les paramètres de diffusion peuvent être calculés à partir de Eq.(1.8), Eq.(1.9) et Eq.(4.29) en se basant sur le comportement asymptotique des processus $\{E(t)\}$ et $\{D(t)\}$ et en utilisant le théorème de renouvellement. Ainsi, $\{N(t), t \geq 0\}$ peut être approximé par un processus de diffusion dont les paramètres infinitésimaux sont

$$a(x, t) = \lambda\rho - \nu(1 - \rho); \quad (4.30)$$

$$b(x, t) = \lambda\rho + \nu(1 - \rho);$$

et la densité de probabilité, considérée sous la condition limite $\lim_{t \rightarrow 0} p(x, t|x_0, t_0) = 0$ pour $x \neq x_0$, est

$$\frac{\partial}{\partial t} p(x, t|x_0, t_0) = \frac{\lambda\rho + \nu(1 - \rho)}{2} \frac{\partial^2}{\partial x^2} p(x, t|x_0, t_0) - (\lambda\rho - \nu(1 - \rho)) \frac{\partial}{\partial x} p(x, t|x_0, t_0); \quad x > 0. \quad (4.31)$$

Il a été montré dans Prabhu [70] et ensuite dans Gross et al. [33] que la solution de (4.31) est donnée par :

$$p(x, t|x_0, t_0) = \frac{e^{-[x-x_0-(\lambda\rho-\nu(1-\rho))t]^2/2(\lambda\rho+\nu(1-\rho))t}}{\sqrt{2\pi(\lambda\rho+\nu(1-\rho))t}}. \quad (4.32)$$

Mais ce résultat n'est pas très significatif à cause de la négligence de la barrière lorsque $x = 0$ puisque le système peut se vider à n'importe quel instant durant le fonctionnement de la file d'attente. En tenant compte de cette possibilité, nous rajoutons une condition

limite supplémentaire :

$$\lim_{x \rightarrow 0} \frac{\partial}{\partial t} p(x, t | x_0, t_0) = 0, \text{ pour tout } t \geq 0.$$

Cette condition est utilisée aussi car le processus ne peut aller au-delà de zéro vers les valeurs négatives.

La nouvelle solution est donc donnée par :

$$\begin{aligned} p(x, t | x_0, t_0) = & \frac{1}{\sqrt{2\pi(\lambda\rho + \nu(1-\rho))t}} \left[e^{-[x-x_0 - (\lambda\rho - \nu(1-\rho))t]^2 / 2(\lambda\rho + \nu(1-\rho))t} \right. \\ & + e^{2\frac{\lambda\rho - \nu(1-\rho)}{\lambda\rho + \nu(1-\rho)}} \left(e^{-[x+x_0 - (\lambda\rho - \nu(1-\rho))t]^2 / 2(\lambda\rho + \nu(1-\rho))t} \right. \\ & \left. \left. - \frac{2(\lambda\rho - \nu(1-\rho))}{\lambda\rho + \nu(1-\rho)} \int_x^\infty e^{-[y+x_0 - (\lambda\rho - \nu(1-\rho))t]^2 / 2(\lambda\rho + \nu(1-\rho))t} dy \right) \right]. \end{aligned} \quad (4.33)$$

Si $a(x, t) < 0$, alors la solution stationnaire obtenue lorsque $t \rightarrow \infty$ est donnée par :

$$p(x) = \frac{2(\nu(1-\rho) - \lambda\rho)}{\nu(1-\rho) + \lambda\rho} e^{-2(\nu(1-\rho) - \lambda\rho)x / (\nu(1-\rho) + \lambda\rho)} ; x > 0. \quad (4.34)$$

C'est une distribution exponentielle indépendante de x_0 .

Simulation et étude numériques

Soit $\{X(t), t \geq 0\}$ le processus de diffusion approximant le nombre de clients bloqués $\{N(t), t \geq 0\}$. Alors, en utilisant les paramètres donnés par Eq.(4.30), $\{X(t)\}$ vérifie l'EDS suivante :

$$dX(t) = (\lambda\rho - \nu(1-\rho))dt + \sqrt{(\lambda\rho + \nu(1-\rho))}dW(t) \text{ avec } X(0) = x_0. \quad (4.35)$$

Pour évaluer l'efficacité de la méthodologie d'approximation décrite précédemment, une étude de simulation a été réalisée. Les algorithmes ont été codés en R. Le nombre moyen de clients dans l'orbite a été simulé en utilisant Eq. (4.35) en se basant sur la méthode d'Euler pour $\lambda \in \{1, 3, 5, 6, 7\}$, $\nu = 10$, $\mu = 12$ et $x_0 \in \{0, 10, 40, 70, 100\}$. Le temps de simulation considéré est $T = 100$ pour un pas de $dt = 1/1000$. Les valeurs exactes du nombre de clients dans l'orbite ont été calculées à l'aide de la formule $E(N) = \lim_{t \rightarrow \infty} E(N(t)) = \frac{\lambda\rho}{1-\rho} \left[\frac{1}{\lambda + \nu} + \frac{1}{\mu} \right]$ déterminée à partir des probabilités Eq.(2.4) et Eq.(2.5) comme suit. Les fonctions génératrices correspondantes sont $P_0(z) = (1-\rho) \left[1 - \frac{\lambda}{\lambda + \nu} + \frac{\lambda}{(\lambda + \nu)(1-\rho z)} \right]$ et $P_1(z) = (1-\rho) \frac{\lambda}{\mu(1-\rho z)}$, respectivement. La fonction gé-

nératrice du nombre de clients dans l'orbite est donnée par

$$P(z) = P_0(z) + P_1(z) = (1 - \rho) \left[1 - \frac{\lambda}{\lambda + \nu} + \left(\frac{\lambda}{\lambda + \nu} + \frac{\lambda}{\mu} \frac{1}{1 - \rho z} \right) \right]. \quad (4.36)$$

Le nombre moyen de clients dans l'orbite est obtenu en posant $z = 1$ dans $P'(z)$.

Les expériences ont été répliquées 1000 fois pour estimer les moyennes. À chaque étape nous avons calculé l'écart relatif en utilisant la formule $|(E(N) - \bar{X})/\bar{X}|$ qui a été exprimé en pourcentage dans le tableau 4.1.

ρ	$E(N)$	\bar{X}	ER $\times 100\%$
0.0917	0.0176	0.4557	96.13
0.325	0.2316	0.5606	58.45
0.625	1.25	1.0028	26.15
0.8	3.5	1.9413	11.86
0.9917	118.4167	117.6811	5.44

TABLE 4.1 – Comparaison entre les résultats exacts et approximatifs du nombre moyen de clients dans l'orbite.

Les valeurs de $\rho \in \{0.0917, 0.325, 0.625, 0.8, 0.9917\}$ correspondent aux valeurs de $\lambda \in \{1, 3, 5, 6, 7\}$, respectivement. Le choix des valeurs $\nu = 10$ et $\mu = 12$ a été fait de sorte que la condition de stabilité du système $\rho = \frac{\lambda(\lambda + \nu)}{\mu\nu} < 1$ soit toujours vérifiée. Les valeurs du tableau 4.1 indiquent que le nombre moyen approximé de clients dans l'orbite se rapproche du nombre moyen exacte au fur et à mesure que les valeurs de ρ tendent vers 1. Cette constatation est confirmée par la diminution de l'écart relatif lorsque les valeurs de ρ rapprochent intensivement de 1. Ce résultat rejoint ceux de Heyman [40], Kimura [52] et Pujolle et Wu [71] qui affirment qu'un processus de naissance et mort peut être approximé par un processus de diffusion dans un trafic intense c'est-à-dire $\rho \simeq 1$. Ainsi, l'étude à titre illustratif de la file considérée, pour laquelle nous avons traité plusieurs échantillons, a montré l'efficacité de la méthode d'approximation proposée.

Conclusion

Cette thèse vise à apporter une contribution à l'étude des files d'attente avec rappels. Tout d'abord, dans le premier chapitre, nous avons présenté les processus stochastiques qui ont été utilisés au cours de notre étude, notamment les PDMPs qui ont été utilisés afin de décrire un système de files d'attente avec une distribution générale des temps de rappels ainsi que les processus de diffusion qui ont été considérés comme des limites des processus de Markov décrivant des systèmes d'attente avec une distribution exponentielle des temps de rappels. Puis, dans le deuxième chapitre, nous nous sommes orientés vers l'étude des phénomènes d'attente dans un cadre mathématique. Nous nous sommes essentiellement concentrés sur la file $M/G/1$ avec rappels pour laquelle nous avons rappelé certains résultats existant dans la littérature en considérant les différentes politiques de rappels dans le cas d'une distribution exponentielle des temps de rappels. Nous avons également présenté les conditions de stabilité de la file $M/G/1$ avec une distribution générale des temps de rappels pour les diverses politiques de rappels. Le troisième chapitre a été consacré à la modélisation des systèmes d'attente par les PDMPs. Comme une première contribution, nous nous sommes intéressés à l'étude de la file $M/G/1$ avec une distribution générale des temps de rappels dans le cadre des PDMPs. Cette nouvelle démarche nous a permis de mieux comprendre la dynamique d'un tel modèle et de déterminer ensuite le nombre moyen de clients bloqués dans l'orbite en régime transitoire. Enfin, l'approximation par les processus de diffusion dans les systèmes d'attente a été établie dans le quatrième chapitre. L'approximation par les processus de diffusion dans les files d'attente avec une distribution exponentielle des temps de rappels a fait l'objet de notre deuxième contribution consistant en deux approximations. La première approximation a été réalisée en utilisant la méthode de convergence du générateur infinitésimal et le deuxième modèle a été étudié sous certaines propriétés asymptotiques du processus décrivant le nombre de clients dans orbite. Une analyse numérique des résultats obtenus a également été effectuée, pour la deuxième approximation, afin de montrer l'efficacité de la méthode d'approximation employée.

En effet, le calcul du générateur infinitésimal ainsi que les martingales associées au PDMP modélisant la file $M/G/1$ avec une distribution générale des temps de rappels est la principale contribution de ce travail. De plus, ces résultats nous ont permis de déterminer le nombre moyen de clients dans l'orbite en régime transitoire. Cependant, à l'avenir, nous estimons qu'il serait intéressant d'exploiter les résultats théoriques sur la

distribution stationnaire d'un PDMP pour étudier le régime permanent dans le modèle de file d'attente considéré. Il est également important d'envisager des approximations par des processus de diffusion dans des modèles de files d'attente avec une distribution générale des temps de rappels modélisés par les PDMPs.

Bibliographie

- [1] ARTALEJO, J., AND FALIN, J. Stochastic decomposition for retrial queues. *Top 2* (1994), 329–342.
- [2] ARTALEJO, J., AND GÓMEZ-CORRAL, A. *Retrial queueing systems : a computational approach*. Berlin, Springer Berlin Heidelberg, ISBN-13, 2008.
- [3] ARTALEJO, J. R. Analysis of an M/G/1 queue with constant repeated attempts and server vacations. *Computers & operations research 24*, 6 (1997), 493–504.
- [4] ARTALEJO, J. R., AND GÓMEZ-CORRAL, A. Steady state solution of a single-server queue with linear repeated requests. *Journal of Applied Probability 34*, 1 (1997), 223–233.
- [5] ARTALEJO, J. R., GÓMEZ-CORRAL, A., AND NEUTS, M. F. Analysis of multiserver queues with constant retrial rate. *European Journal of Operational Research 135*, 3 (2001), 569–581.
- [6] BREUER, L. Continuity of the M/G/c queue. *Queueing Systems 58*, 4 (2008), 321–331.
- [7] BURMAN, D. *An analytic approach to diffusion approximation in Queueing Ph. D.* PhD thesis, thesis, Department of Mathematics, New York University, 1979.
- [8] CHAKRAVARTHY, S. R. Analysis of MAP/PH/c retrial queue with phase type retrials–simulation approach. In *Belarusian Workshop on Queueing Theory* (2013), Springer, pp. 37–49.
- [9] CHOUDHURY, G. A two phase batch arrival retrial queueing system with bernoulli vacation schedule. *Applied Mathematics and Computation 188*, 2 (2007), 1455–1466.
- [10] CLANCY, D. SIR epidemic models with general infectious period distribution. *Statistics & Probability Letters 85* (2014), 1–5.
- [11] COHEN, J. Basic problems of telephone traffic theory and the influence of repeated calls, Philips Telecom. *Review*, 18 (1957).
- [12] COSTA, O. L. V. Stationary distributions for piecewise-deterministic Markov processes. *Journal of Applied Probability 27*, 1 (1990), 60–73.

- [13] COSTA, O. L. V., AND DUFOUR, F. Stability and ergodicity of piecewise deterministic Markov processes. *SIAM Journal on Control and Optimization* 47, 2 (2008), 1053–1077.
- [14] COX, D. R., AND ISHAM, V. *Point processes*, vol. 12. CRC Press, 1980.
- [15] DASSIOS, A., AND ZHAO, H. A dynamic contagion process. *Advances in applied probability* 43, 3 (2011), 814–846.
- [16] DAVIS, M. H. Piecewise-deterministic Markov processes : A general class of non-diffusion stochastic models. *Journal of the Royal Statistical Society : Series B (Methodological)* 46, 3 (1984), 353–376.
- [17] DAVIS, M. H. *Markov models & optimization*. Chapman & Hall, 1993.
- [18] DUFOUR, F., AND COSTA, O. L. Stability of piecewise-deterministic Markov processes. *SIAM Journal on Control and Optimization* 37, 5 (1999), 1483–1502.
- [19] ERLANG, A. K. The theory of probabilities and telephone conversations. *Nyt. Tidsskr. Mat. Ser. B* 20 (1909), 33–39.
- [20] ERLANG, A. K. Solution of some problems in the theory of probabilities of significance in automatic telephone exchanges. *Post Office Electrical Engineer's Journal* 10 (1917), 189–197.
- [21] FALIN, G. Limit theorems for queueing systems with retrials. *Abstracts of IV Vilnius Conference on Probability Theory and Mathematical Statistics*, 3 (1985), 235–237.
- [22] FALIN, G. Periodic queues in heavy traffic. *Advances in Applied Probability* 21, 2 (1989), 485–487.
- [23] FALIN, G. A diffusion approximation for retrial queueing systems. *Theory of Probability & Its Applications* 36, 1 (1992), 149–152.
- [24] FALIN, G. I. Single-line repeated orders queueing systems. *Optimization* 17, 5 (1986), 649–667.
- [25] FALIN, G. I. A survey of retrial queues. *Queueing systems* 7 (1990), 127–168.
- [26] FALIN, G. I., AND J. G. C. TEMPLETON. *Retrial Queues*. Chapman & Hall, 1997.
- [27] FARAHMAND, K. Single line queue with repeated demands. *Queueing Systems* 6, 1 (1990), 223–228.
- [28] FAYOLLE, G. A simple telephone exchange with delayed feedback, in : Teletraffic analysis and computer performance evaluation, OJ Boxma, JW Cohen and HC Tijms. *Elsevier Science* 46 (1986), 561–581.
- [29] FUCHS, C. *Inference for diffusion processes : with applications in life sciences*. Springer Science & Business Media, 2013.
- [30] GARRIDO, J., AND LU, Y. On double periodic non-homogeneous poisson processes. *Business Economics Series*, 2002.

- [31] GÓMEZ-CORRAL, A. Stochastic analysis of a single server retrial queue with general retrial times. *Naval Research Logistics (NRL)* 46, 5 (1999), 561–581.
- [32] GÓMEZ-CORRAL, A., AND LÓPEZ-GARCÍA, M. On SIR epidemic models with generally distributed infectious periods : Number of secondary cases and probability of infection. *International Journal of Biomathematics* 10, 02 (2017).
- [33] GROSS, D. *Fundamentals of queueing theory*. John Wiley & Sons, 2008.
- [34] GUESS, F., AND PROSCHAN, F. 12 mean residual life : theory and applications. *Handbook of statistics* 7 (1988), 215–224.
- [35] GUPTA, P., AND KUMAR, N. Cost optimization of single server retrial queueing model with bernoulli schedule working vacation, vacation interruption and balking. *J. Math. Comput. Sci.* 11, 3 (2021), 2508–2523.
- [36] GUPTA, P., KUMAR, N., AND GUPTA, R. (r1971) analysis of feedback queueing model with differentiated vacations under classical retrial policy. *Applications and Applied Mathematics : An International Journal (AAM)* 17, 2 (2022), 2.
- [37] HARRISON, J. M. Brownian motion and stochastic flow systems. *Wiley New York*, 1985.
- [38] HARRISON, J. M., AND LEMOINE, A. J. Limit theorems for periodic queues. *Journal of Applied Probability* 14, 3 (1977), 566–576.
- [39] HASOFER, A. On the single-server queue with non-homogeneous poisson input and general service time. *Journal of Applied Probability* 1, 2 (1964), 369–384.
- [40] HEYMAN, D. P. A diffusion model approximation for the GI/G/1 queue in heavy traffic. *Bell System Technical Journal* 54, 9 (1975), 1637–1646.
- [41] IACUS, S. M. *Simulation and inference for stochastic differential equations : with R examples*, vol. 486. Springer, 2008.
- [42] IGLEHART, D. L., AND WHITT, W. Multiple channel queues in heavy traffic. ii : Sequences, networks, and batches. *Advances in Applied Probability* 2, 2 (1970), 355–369.
- [43] JACOB, V. Analysis of customer induced interruption in a retrial queueing system with classical retrial policy. *International Journal of Applied Engineering Research* 15, 5 (2020), 445–451.
- [44] JAIN, M., AND SANGA, S. S. Admission control for finite capacity queueing model with general retrial times and state-dependent rates. *Journal of Industrial and Management Optimization* 16, 6 (2020), 2625–2649.
- [45] KAPYRIN, V. Stationary characteristics of a queueing system with repeated calls. *Cybernetics* 13, 4 (1977), 584–590.

- [46] KENDALL, D. G. Stochastic processes occurring in the theory of queues and their analysis by the method of the imbedded markov chain. *The Annals of Mathematical Statistics* (1953), 338–354.
- [47] KERNANE, T. *Stabilité de modèles de files d’attente*. PhD thesis, Université des Sciences et de la Technologie USTHB, 2007.
- [48] KERNANE, T. Conditions for stability and instability of retrial queueing systems with general retrial times. *Statistics & probability letters* 78, 18 (2008), 3244–3248.
- [49] KERNANE, T., AND AÏSSANI, A. Stability of retrial queues with versatile retrial policy. *Journal of Applied Mathematics and Stochastic Analysis* 2006 (2006).
- [50] KHINTCHINE, A. Y. Mathematical theory of stationary queues. *Matem. Sbornik* 39 (1932), 73–84.
- [51] KIM, C., KLIMENOK, V., AND DUDIN, A. A G/M/1 retrial queue with constant retrial rate. *Top* 22, 2 (2014), 509–529.
- [52] KIMURA, T. Diffusion approximations for queues with markovian bases. *Annals of Operations Research* 113, 1 (2002), 27–40.
- [53] KIMURA, T., AND MASUYAMA, H. A heavy-traffic-limit formula for the moments of the stationary distribution in GI/G/1-type markov chains. *Operations Research Letters* 49, 6 (2021), 862–867.
- [54] KINGMAN, J. F. On queues in heavy traffic. *Journal of the Royal Statistical Society : Series B (Methodological)* 24, 2 (1962), 383–392.
- [55] KOLMOGOROFF, A. Über die analytischen methoden in der wahrscheinlichkeitsrechnung. *Mathematische Annalen* 104 (1931), 415–458.
- [56] KOROLIUK, V. S., AND LIMNIOS, N. *Stochastic systems in merging phase space*. World Scientific, 2005.
- [57] KUMAR, B. K., AND RAJA, J. On multiserver feedback retrial queues with balking and control retrial rate. *Annals of Operations Research* 141 (2006), 211–232.
- [58] LI, K., AND WANG, J. Equilibrium balking strategies in the single-server retrial queue with constant retrial rate and catastrophes. *Quality Technology & Quantitative Management* 18, 2 (2021), 156–178.
- [59] LILLO, R. A G/M/1-queue with exponential retrial. *Top* 4, 1 (1996), 99–120.
- [60] LIU, X., GONG, Q., AND KULKARNI, V. G. Diffusion models for double-ended queues with renewal arrival processes. *Stochastic Systems* 5, 1 (2015), 1–61.
- [61] MEZIANI, S., AND KERNANE, T. Etude de la file M/G/1 avec rappel général et politique de rappel constante dans le cadre d’un PDMP. In *Colloque International MOAD’2022 (Méthodes et Outils d’Aide à la Décision)* (15-17 Novembre 2022), Béjaia, Algérie.

- [62] MEZIANI, S., AND KERNANE, T. Extended generator and associated martingales for M/G/1 retrial queue with classical retrial policy and general retrial times. *Probability in the Engineering and Informational Sciences* (2022), 1–8.
- [63] MEZIANI, S., AND KERNANE, T. Diffusion Approximation of a Finite-Source M/M/1 Retrial Queueing System. In *1st International Conference On Pure and Applied Mathematics (IC-PAM'21)* (May 26-27, 2021), Ouargla, Algeria (Virtual conference).
- [64] MOROZOV, E., AND PHUNG-DUC, T. Stability analysis of a multiclass retrial system with classical retrial policy. *Performance Evaluation* 112 (2017), 15–26.
- [65] MOROZOV, E., AND ZHUKOVA, K. An upper bound of the large deviation probability in multi-server constant retrial rate system. In *Distributed Computer and Communication Networks : 22nd International Conference, DCCN 2019, Moscow, Russia, September 23–27, 2019, Revised Selected Papers 22* (2019), Springer, pp. 325–337.
- [66] NEKRASOVA, R. Stability analysis of a multi-class retrial queue with general retrials and classical retrial policy. In *2021 28th Conference of Open Innovations Association (FRUCT)* (2021), IEEE, pp. 328–333.
- [67] NEWELL, G. F. *Application of Queueing Theory*. Chapman & Hall, 1971.
- [68] PHUNG-DUC, T. Retrial queueing models : A survey on theory and applications. *arXiv preprint arXiv :1906.09560* (2019).
- [69] POLLACZEK, F. Über eine aufgabe der wahrrscheinlichkeitstheorie. i : Mitteilung aus dem telegraphentechnischen reichsamt. *Mathematische Zeitschrift* 32, 1 (1930), 64–100.
- [70] PRABHU, N. U. *Stochastic Processes*. Wiley, 1965b.
- [71] PUJOLLE, G., AND WU, A. Approximation par un processus de diffusion de files GI/GI/1 avec dépendance de l'état. *RAIRO-Operations Research* 19, 2 (1985), 117–131.
- [72] REICH, E. On the integrodifferential equation of Takács. i. *The Annals of Mathematical Statistics* 29, 2 (1958), 563–570.
- [73] REICH, E. On the integrodifferential equation of Takács. ii. *The Annals of Mathematical Statistics* 30, 1 (1959), 143–148.
- [74] SHIN, Y. W., AND MOON, D. H. Approximation of M/M/c retrial queue with ph-retrial times. *European journal of operational research* 213, 1 (2011), 205–209.
- [75] SHIN, Y. W., AND MOON, D. H. Approximation of PH/PH/c retrial queue with ph-retrial time. *Asia-Pacific Journal of Operational Research* 31, 02 (2014), 1440010.
- [76] SUDYKO, E., NAZAROV, A. A., AND SZTRIK, J. Asymptotic waiting time analysis of a finite-source M/M/1 retrial queueing system. *Probability in the Engineering and Informational Sciences* 33, 3 (2019), 387–403.

- [77] TAKÁCS, L. The limiting distribution of the virtual waiting time and the queue size for a single-server queue with recurrent input and general service times. *Sankhyā : The Indian Journal of Statistics, Series A* (1963), 91–100.
- [78] YANG, T., POSNER, M., TEMPLETON, J. G., AND LI, H. An approximation method for the M/G/1 retrial queue with general retrial times. *European Journal of Operational Research* 76, 3 (1994), 552–562.
- [79] ZHOU, M., LIU, L., CHAI, X., AND WANG, Z. Equilibrium strategies in a constant retrial queue with setup time and the n-policy. *Communications in Statistics-Theory and Methods* 49, 7 (2020), 1695–1711.
- [80] ZIREM, D., BOUALEM, M., ADEL-AISSANOU, K., AND AÏSSANI, D. Analysis of a single server batch arrival unreliable queue with balking and general retrial time. *Quality Technology & Quantitative Management* 16, 6 (2019), 672–695.