



الجمهورية الجزائرية الديمقراطية الشعبية
République Algérienne Démocratique et Populaire
وزارة التعليم العالي والبحث العلمي



Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
جامعة الشاذلي بن جديد - الطارف
Université Chadli Bendjedid – El Tarf
كلية العلوم و التكنولوجيا
Faculté des Sciences et de la Technologie
قسم الرياضيات
Département de Mathématiques

Mémoire de fin d'études

En vue de l'obtention du diplôme de Master

Domaine : Mathématiques et Informatique

Filière : Mathématiques

Spécialité: Analyse fonctionnelle et calcul stochastique

Thème

Etude de la performance du système
M/M/C/K avec rappels

Présenté par:

Siad Roguia

Devant le Jury :

Dr. Grabsia Imen	MCB	Univ Chadli Bendjedid El Tarf	Présidente
Dr. Zidani Nesrine	MCB	Univ Chadli Bendjedid El Tarf	Rapporteur
Dr. Grine Razika	MCB	Univ Chadli Bendjedid El Tarf	Examinatrice

Année Universitaire 2021-2022

Table des matières

Introduction	4
1 Les éléments de la théorie des files d'attente	5
1.1 Introduction	5
1.2 Les systèmes de files d'attente classiques	6
1.2.1 Description du modèle d'attente classique	6
1.2.2 Analyse mathématique d'un système de files d'attente	9
1.2.3 Modèles markoviens	10
1.2.4 Modèles semi-markoviens	10
1.3 Les systèmes de files d'attente avec rappels	11
1.3.1 Modèle général	12
1.3.2 Modèles markoviens	13
1.3.2.1 Modèle M/M/1 avec rappels	13
1.3.2.2 Modèles à multiserveurs	15
2 Approches d'analyse des modèles à multiserveurs avec rappels	20
2.1 Introduction	20
2.2 modèles tronqués	21
2.2.1 Description du modèle	21
2.2.2 Régime stationnaire	22
2.2.3 Condition d'existence d'un régime stationnaire	22
2.2.3.1 Condition suffisante	22
2.2.3.2 Condition nécessaire	24

2.2.4	Un algorithme pour le calcul numérique de la distribution stationnaire	26
2.3	la méthode d'extrapolation de valeur	28
2.3.1	Processus de Markov Décisionnels	28
2.3.2	Ajustement polynomial	30
2.3.3	Fonction Revenu	34
2.3.4	Effet de l'extrapolation de la valeur dans les équations d'Howard . .	34
3	Étude et analyse du modèle d'attente M/M/C/K avec rappels	38
3.1	Introduction	38
3.2	Description du modèle	38
3.2.1	Condition d'existence d'un régime stationnaire	40
3.2.1.1	Conditions suffisante	40
3.2.1.2	Condition nécessaire	41
3.3	Régime stationnaire	43
3.4	Troncature finie	44
3.5	Extrapolation de Valeur	48

Liste des Figures

1.2.1 Schéma de files d'attente classique	6
1.3.1 Schéma général d'un système avec rappels	13
1.3.2 modèle M/M/1 avec rappels	14

Dédicaces

Avec ma gratitude et tout mon amour je dédie ce modeste travail:

À l'être la plus chère de ma vie ma mère, tu trouves toujours les mots qui m'encourageaient et qui me poussaient à aller de l'avant, qui a partagé avec moi chaque pas que j'ai fait dans ma vie avec ses prières. Je prie Dieu de te rendre heureux.

À mon très cher père, pour son soutien, son affection et la confiance qu'il m'a accordé.

À mes chers frères.

À tous les membres de ma famille, et tous mes amis.

Remerciement

*Tous les éloges et les remerciements sont à **ALLAH** de m'avoir donné la force et la puissance de pouvoir mener ce travail à terme.*

*Je remercie vivement mon encadreur **Dr. Zidani Nesrine** pour sa continuelle assistance, sa patience, sa disponibilité et ses conseils qui m'ont été d'un grand apport et encouragement sans oublier ses efforts fournis pendant toute la durée de l'élaboration de ce travail.*

*Je remercie les membres du jury : **Dr. Grine Razika** et **Dr. Grabsia Imen** pour l'attention qu'ils ont accordé à ce travail.*

*Finalement, je remercie ma **famille** pour son soutien et ses encouragements tout au long de mes études.*

Résumé

Dans ce travail, nous nous sommes intéressées à l'analyse stochastique d'un système de files d'attente $M/M/C/K$ avec rappels et abandons exponentiels auquel les clients arrivent selon des processus de Poisson en utilisons les méthodes d'approximation.

Dans un premier temps, nous avons effectué une étude sur la théorie des files d'attente classique et avec rappels. À cet effet, nous avons donné une description du modèle général des deux systèmes classiques et avec rappels. Puis, nous avons passé en revue certains résultats consacrés aux systèmes de files d'attente avec rappels à unique serveur et à multiserveurs.

Dans un deuxième temps, nous avons présenté les approches d'approximations utilisées par les chercheurs pour résoudre les systèmes de files d'attente avec rappels et multiserveurs. En effet, nous avons décrit la Troncature Finie et Extrapolation de Valeur. Et à la fin nous avons donné une description d'un modèle particulier des modèles à multiserveur avec rappels, espace d'attente avec capacité limitée et abandons exponentiels, auquel nous avons trouvé la condition d'ergodicité, et présentons les étapes d'application des méthodes de résolution approximative.

Mots-clés: système de file d'attente avec rappels, Condition d'ergodicité, Troncature, Extrapolation, abandon.

Abstract

In this work, we are interested in the stochastic analysis of M/M/C/K retrial queuing system with exponential abandonments at which customers arrive according to Poisson processes using approximation methods .

First, we carried out a study on the theory of classical and retrial queuing. For this purpose, we have given a description of the general model of the two systems classical and retrial. Then we reviewed some results on single-server and multi-server retrail queuing systems.

Secondly, we presented the approximation approaches used by researchers to solve retrial queuing systems and multiservers. Indeed, we have described Finite Truncation and Value Extrapolation. And at the end we gave a description of a particular model of the retrial multiserver queue, waiting space with limited capacity and exponential abandonment, to which we found the ergodicity condition, and presented the steps for applying the approximate resolution methods.

keyword: Retrial queue, Ergodicity Condition, Truncation, Extrapolation, Abandonment.

المخلص

في هذا العمل ، نحن مهتمون بتحليل العشوائي لنظام الطابور $M / M / C / K$ مع عمليات الاسترجاع والتسرب الأسيية التي يصل إليها العملاء وفقاً لعمليات بواسون باستخدام طرق التقريب.

أولاً ، أجرينا دراسة حول نظرية الاصطفاف الكلاسيكي مع التذكيرات. لهذا الغرض ، قدمنا وصفاً للنموذج العام للنظامين الكلاسيكيين مع التذكيرات. ثم قمنا بمراجعة بعض النتائج على أنظمة قائمة انتظار أحادية الخادم ومتعددة الخوادم مع عمليات الاسترجاعات.

ثانياً ، قدمنا مناهج التقريب التي يستخدمها الباحثون لحل أنظمة الطابور باستخدام الاسترجاعات والخوادم المتعددة. في الواقع ، لقد وصفنا الاقتران المحدود واستقراء القيمة. وفي النهاية قدمنا وصفاً لنموذج معين من نماذج الخوادم المتعددة مع عمليات الاسترجاعات ، ومساحة الانتظار ذات السعة المحدودة والمتسربين الأسيين ، والتي وجدنا فيها حالة العملانية ، وعرض مراحل تطبيق طرق الدقة التقريبية.

الكلمات المفتاحية: نظام الطابور ذات الاسترجاع، شرط الاستقرار، الاقتران، استقراء، فرار.

Introduction générale

La théorie des files d'attente est l'un des outils analytiques les plus puissants pour la modélisation de systèmes de logistique et de communication. Cette théorie a pour objet l'étude de systèmes et réseaux où des entités, appelées clients, cherchent à accéder à des ressources, généralement limitées, afin d'en obtenir un service. La demande concurrente d'une même ressource par plusieurs clients engendre des délais dans la réalisation des services ainsi que des files de clients désireux d'accéder à une ressource indisponible. L'analyse théorique de tels systèmes permet d'établir à l'avance les performances de l'ensemble, d'identifier les éléments critiques ou encore, d'appréhender les effets d'une modification des conditions de fonctionnement.

Ce domaine de recherche est né en 1917, des travaux de l'ingénieur électricien danois Erlang sur la gestion des réseaux téléphoniques de Copenhague entre 1909 et 1920. Par la suite, les files d'attente ont été utilisées dans la modélisation des systèmes de production et des systèmes informatiques. Initialement, leur utilisation était très limitée par le fait qu'elles faisaient vite appel à une puissance de calcul non disponible à cette époque. Ainsi, pendant longtemps, c'est surtout des développements de nature théorique qui ont vu le jour dans le domaine. Le sujet a pourtant inspiré et continue à inspirer de nombreux chercheurs comme en témoignent les nombreuses publications parues à ce jour dans le domaine. Avec l'automatisation des systèmes technologiques, notamment l'avènement des systèmes de production flexibles, les convoyeurs automatiques, etc, les applications nécessitant le traitement par une telle théorie se sont également multipliées.

Ce n'est pourtant que rarement que les files d'attente apparaissent de manière isolée. Le plus souvent ce sont des réseaux entiers avec un nombre élevé de serveurs interconnectés qu'il s'agit de modéliser. La grande percée dans le domaine est venue avec la découverte en 1957, par Jackson, d'une manière particulièrement élégante et efficace de traiter certains réseaux de files d'attente. C'est grâce aux apports des mathématiciens Khintchine, Palm, Kendall, Pollaczek et Kolmogorov que la théorie s'est vraiment développée.

La théorie des files d'attente classique offre deux possibilités pour résoudre le conflit qui apparaît lorsqu'un client arrive dans le système à serveur unique et trouve le serveur occupé: soit il quitte le système sans recevoir le service (modèle d'Erlang à demandes refusées), soit il prend place dans une file d'attente. Une possibilité alternative est que le client qui ne peut être servi, libère l'espace du serveur mais après une durée de temps aléatoire revient au système pour répéter sa demande de service. Entre les appels successifs, le client en question se trouve en "orbite". Un tel système est appelé système de files d'attente avec "rappels" [1]. De nombreuses études sont consacrées aux modèles d'attente avec rappels [20]. Les tentatives de classer les résultats obtenus pour la période allant jusqu'à 2019 sont présentées dans les articles bibliographiques de Artalejo [6], [3], et Shekhar. La nature des résultats obtenus, des méthodes d'analyse [22] et des domaines d'application nous permettent de diviser les files d'attente avec rappels en deux groupes: les systèmes à serveur unique, et à multiserveurs. Les modèles à multiserveurs attirent beaucoup d'attention parce que la conception d'un système de files d'attente avec rappels est déduite de ces importantes applications dans les systèmes de téléphonie [19] [5] [7] [23].

Due à l'absence des formules analytiques explicites pour les caractéristiques probabilistes principales des modèles de files d'attente avec rappels et multiserveurs, de nombreuses tentatives ont été faites pour développer un certain nombre de méthodes approximatives. Cependant, de nombreux chercheurs recourent à des approximations qui sont souvent basées sur des modèles tronqués finis [28], des modèles tronqués généralisés ou sur une homogénéisation de l'espace d'états [13]. Dans les deux premiers cas, le processus

stochastique hétérogène, décrivant l'état du système et possédant un espace d'états infini, est remplacé par un autre homogène et ayant l'espace d'états fini ou infini mais résoluble. Le troisième type d'approximation consiste à analyser le système de files d'attente comme un processus de quasi-naissance et-mort dont les probabilités stationnaires peuvent être obtenus en utilisant une méthode matricielle géométrique[16] . Ces approches fournissent une solution numérique à la distribution stationnaire des chaînes de Markov à temps continu. Une autre approche a été proposée dans[15] appelée Extrapolation de Valeur, consiste à étudier un processus stochastique de Markov en tant qu'un processus de Markov décisionnel c-à-d on ne considère plus la probabilité d'être dans un certain état, mais on s'intéresse à une nouvelle métrique, appelée valeur d'état relative. Le but est de trouver la valeur relative moyenne, représentant une mesure de performance du modèle, en résolvant les équations de Howard.

Dans ce mémoire, nous considérons un système de files d'attente M/M/C/K avec rappels et abandons exponentiels, auxquels les clients arrivent selon des processus de Poisson. Pour le modèle à l'étude, nous trouvons les conditions d'ergodicité et aussi la solution approximative pour les caractéristiques de performance essentielles. A cet fin, nous appliquons deux approches alternatives: Troncature Finie et Extrapolation de Valeur.

Ce mémoire est constitué d'une introduction générale, trois chapitres, d'une conclusion générale, d'une bibliographie.

Dans le premier chapitre, nous présentons les systèmes de files d'attente classiques et avec rappels.

Dans second chapitre, nous décrivons le modèle tronqué et l'approche d'extrapolation de valeur pour les systèmes de files d'attente avec rappels et multiserveurs.

Dans le troisième chapitre, nous réalisons l'analyse stochastique du modèle à l'étude: M/M/C avec rappels, espace d'attente de capacité limitée, abandons exponentiels. À cet effet, nous donnons une description mathématique du modèle, trouvons la condition d'ergodicité, présentons les étapes d'application des méthodes de résolution approximative: Troncature Finie et Extrapolation de Valeur.

Les éléments de la théorie des files d'attente

1.1 Introduction

La théorie des files d'attente constitue une approche pour la modélisation stochastique, l'évaluation des performances et le contrôle des systèmes de production, systèmes informatiques, réseaux de communication..... Elle est la plus apte à fournir une estimation quantitative d'un système.

Toutefois, la théorie des files d'attente classique offre deux possibilités pour résoudre le conflit qui apparaît lorsqu'un client arrive dans le système et trouve l'espace de service occupé: soit le client quitte le système sans recevoir le service, soit il prend place dans une file d'attente. Une possibilité alternative est de permettre au client de répéter sa demande de service après une durée de temps aléatoire. Entre deux tentatives successives (rappels), le client en question est en orbite. Un tel système est appelé système de files d'attente avec rappels.

1.2 Les systèmes de files d'attente classiques

1.2.1 Description du modèle d'attente classique

Les systèmes de files d'attente classiques peuvent être décrits comme suit: des clients arrivent à un certain endroit et réclament un certain service, ils attendent d'être servis, et quittent ce système s'ils ont eu le service. Un système d'attente comprend donc un espace de service avec une ou plusieurs stations de service montées en parallèle et un espace d'attente dans lequel se forme une éventuelle file d'attente[14]. Le terme clients ne signifie pas qu'on parle que des êtres humains, les clients peuvent être par exemple, des camions en attente de chargement ou de déchargement, des machines en attente de réparation, des programmes d'ordinateurs qui attendent d'être exécutés, des avions qui attendent l'autorisation de décoller ou d'atterrir. Un système de base des files d'attente peut schématiquement représenter comme dans la figure 1.2.1.

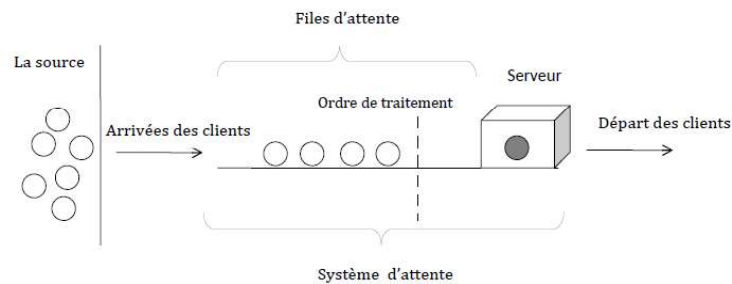


Figure 1.2.1 : Schéma de files d'attente classique

Pour identifier un système de files d'attente, on doit spécifier:

- **Le processus des arrivées des clients:** Les arrivées des clients sont caractérisées par l'ensemble des instants d'arrivées de chaque client ou d'un groupe de clients dans le système. La collection de ces instants forment un processus des arrivées. Souvent, on suppose que les temps entre deux arrivées consécutives sont indépendants et identiquement distribués.
- **La source des clients:** La population source, d'où proviennent les clients, peut être finie ou infinie, unique ou multiples.

- **La capacité du système:** Elle représente le nombre maximal de clients dans le système. Un client arrivant et trouvant ce nombre de clients présents dans le système sera perdu.
- **La durée de service:** Le temps de service d'un client est défini par le temps séparant le début et la fin de son service. En général, on suppose que les durées de service sont indépendantes et identiquement distribuées, et indépendantes des temps d'inter-arrivées.
- **La discipline de service:** Elle spécifie la manière avec laquelle le serveur sélectionne le prochain client à servir. Cependant, plusieurs possibilités existent quant à l'ordre selon lequel les clients seront servis. les principales disciplines du service sont:
 - FIFO (first in, first out) ou FCFS (first come first served): c'est la file standard dans laquelle les clients sont servis dans leur ordre d'arrivée. Notons que les disciplines FIFO et FCFS ne sont pas équivalentes lorsque la file contient plusieurs serveurs. Dans la première, le premier client arrivé sera le premier à quitter la file alors que dans la deuxième, il sera le premier à commencer son service. Rien n'empêche alors qu'un client qui commence son service après lui, dans un autre serveur, termine avant lui.
 - LIFO (last in, first out) ou LCFS (last come, first served): Cela correspond à une pile, dans laquelle le dernier client arrivé (donc posé sur la pile) sera le premier traité (retiré de la pile). À nouveau, les disciplines LIFO et LCFS ne sont équivalentes que pour une file monoserveur.
 - SIRO (Served In Random Order): les clients sont servis au aléatoirement.
 - PNP (Priority service): les clients sont servis à leur priorité . Tous les clients de la plus haute priorité sont servis premier, puis les clients de moindre priorité sont servis, et ainsi de suite.
 - PS (Processor Sharing): les clients sont servis de manière égale. La capacité du système est partagée entre les clients.

- **Le nombre de serveurs:** Il s'agit du nombre de serveurs composant l'espace de service. La plupart du temps, les serveurs sont considérés identiques et indépendants les uns des autres.

Puisque les instants d'arrivée et les durées de service sont généralement des quantités aléatoires, la modélisation décrivant le fonctionnement d'un système de files d'attente est un processus stochastique. Par ailleurs, on suppose généralement que toutes les variables aléatoires introduites pour décrire un système d'attente sont mutuellement indépendantes.

Pour la classification des systèmes de files d'attente, on a recours à une notation symbolique (Notation de Kendall) $A/B/C/m/O/H$ où:

- A : La distribution des temps entre deux arrivées successives. Pour spécifier la distribution A , les symboles suivants sont utilisés:
 - M : distribution exponentielle.
 - H_k : distribution hyperexponentielle d'ordre k .
 - G : distribution générale.
 - E_k : distribution d'Erlang d'ordre k .
- B : la distribution des temps de service d'un client. les symboles sont les mêmes que A .
- C : nombre de serveurs.
- m : capacité de la file (c'est le nombre de places dans le système en d'autre terme c'est le nombre maximal de clients permis dans le système y compris ceux en service).
- O : population des usagers.
- H : discipline de service (c'est la façon dont les clients sont ordonnés pour être servi).

1.2.2 Analyse mathématique d'un système de files d'attente

L'étude mathématique d'un système de files d'attente se fait généralement par l'introduction d'un processus stochastique défini de façon appropriée. En fonction des quantités qui définissent la structure du système, on cherche à calculer les probabilités d'état définissant le régime transitoire du processus en question, puis le régime stationnaire. Les probabilités d'état définissant le régime transitoire dépendent de la distribution initiale du processus étudié.

Ainsi le calcul explicite du régime transitoire s'avère généralement pénible, voire impossible, pour la plupart des modèles donnés. La distribution stationnaire du processus introduit permet d'obtenir les mesures de performance du système :

- Le nombre moyen de client dans le système \bar{N} .
- Le nombre moyen de client dans la file d'attente \bar{N}_f .
- Le temps moyen d'attente d'un client dans la file d'attente \bar{W} .
- Le temps moyen d'attente d'un client dans le système \bar{W}_s .

Ces valeurs ne sont pas indépendantes les unes des autres, mais sont liées par les relations suivantes (*Formules de Little*):

$$\begin{aligned}\bar{N} &= \lambda \bar{W}_s, \\ \bar{N}_f &= \lambda \bar{W}, \\ \bar{W}_s &= \bar{W} + \frac{1}{\mu}, \\ \bar{W} &= \frac{\bar{N}_f}{\lambda}, \\ \bar{N} &= \bar{N}_f + \frac{\lambda}{\mu}.\end{aligned}$$

Où λ est le taux d'entrée des clients dans le système, et $\frac{1}{\mu}$ la durée moyenne du service ($\mu > 0$). Une autre mesure importante d'un système de files d'attente, celle qui mesure le degré de saturation du système, est l'intensité du trafic ρ .

$$\rho = \frac{\text{temps moyen du service}}{\text{temps moyen entre deux arrivées successives}}$$

1.2.3 Modèles markoviens

Les modèles markoviens caractérisent les systèmes dans lesquels tous les événements tels que les arrivées, les fins de services, les changements d'état du système dépendent uniquement de l'état actuel du système et non de son comportement antérieur. L'étude mathématique de tels systèmes se fait par l'introduction d'un processus stochastique approprié. Ce processus est souvent le processus de naissance et de mort $\{N(t); t \geq 0\}$ qui définit le nombre de clients dans le système à l'instant t . L'évolution temporelle du processus markovien $\{N(t); t \geq 0\}$ est complètement définie grâce à la propriété d'absence de mémoire.

1.2.4 Modèles semi-markoviens

En l'absence de l'exponentialité c'est à dire lorsque l'on s'écarte de l'hypothèse d'exponentialité de l'une des deux quantités stochastiques: le temps des interarrivées et la durée de service, où en prenant en compte certaines spécificités des problèmes par l'introduction de paramètres supplémentaires, on aboutit à un modèle semi-markovien. La combinaison de tous ces facteurs rend l'étude mathématique du modèle très délicate, voire impossible. On essaye alors de se ramener à un processus de Markov judicieusement choisi à l'aide de l'une des méthodes d'analyse suivantes:

- **Méthode des étapes d'Erlang:** Son principe est d'approximer toute loi de probabilité ayant une transformée de Laplace rationnelle par une loi de Cox (mélange de lois exponentielles).
- **Méthode des variables supplémentaires:** Elle consiste à compléter l'information sur le processus $\{N(t), t \geq 0\}$ en lui donnant un caractère Markovien, ce qui nous ramène à l'étude du processus $\{N(t), \xi(t), t \geq 0\}$, où $\xi(t)$ sont dites alors variables supplémentaires.
- **Méthode de la chaîne de Markov induite:** Elle consiste à choisir une séquence d'instant $1, 2, 3, \dots, n$ (déterministes ou aléatoires) telle que la chaîne induite $\{q_n, n \geq 0\}$,

où q_n est le nombre de clients dans le système à l'instant $n \geq 0$, soit markovienne et homogène.

- **Simulation:** C'est un procédé d'imitation artificielle d'un processus réel donné sur ordinateur qui permet d'étudier les systèmes les plus complexes, de prévoir leurs comportements et de calculer leurs caractéristiques. Les résultats obtenus ne sont qu'approximatifs, mais peuvent être utilisés avec une bonne précision. Cette technique se base sur la génération de variables aléatoires. D'autres méthodes d'analyse de systèmes semi-markoviens existent, telle que l'approche utilisant les martingales et les méthodes d'approximations.

1.3 Les systèmes de fils d'attente avec rappels

La naissance des modèles de files d'attente avec rappels a été entraînée par des phénomènes d'appels répétés dans des réseaux informatiques et téléphoniques. Elles sont caractérisées par l'hypothèse de base suivante: le client qui arrive et trouve le serveur occupé doit rejoindre une file supplémentaire de client s'appelée "orbite", et réessaye ultérieurement de faire des tentatives de service après une durée de temps aléatoire d'après une politique particulière de rappels. Si par contre à l'arrivée du client il y a un ou plusieurs serveurs libres et en bon état, le client sera servi immédiatement et quitte le système après son service. L'orbite peut être de capacité finie ou infinie, dans le premier cas, si l'orbite est pleine, le client doit quitter le système sans être servi dès qu'il trouve tous les serveurs et les positions d'attente occupés.

Ces modèles d'attente apparaissent dans la modélisation stochastique de plusieurs situations réelles et des réseaux de télécommunications [23]. Par exemple dans la transmission de données, un paquet transmis de la source à la destination peut être retourné, et le processus doit être répété jusqu'à ce que le paquet soit finalement transmis. Les premières tentatives rigoureuses sur les systèmes de files avec rappels remontent aux travaux de Kosten(1947) [24], Wilkinson [29] et cohen [9]. Les progrès dans ce domaine

sont résumés dans les monographies de Falin et templeton (1997)[13] , et une classification bibliographique sur les systèmes avec rappels est donné par Artalejo (1998) [5] et (1999) [6] et (2006) [7] et (2010) [3].

1.3.1 Modèle général

En général, un système de files d'attente avec rappels contient un espace de service composé de $C \geq 1$ dispositifs de service et un espace d'attente de $m - C$ ($m \geq C$) positions d'attente. Les clients arrivent dans le système selon un processus aléatoire avec une loi de probabilité donnée, et forment un flux d'appels primaires. À l'arrivée d'un client, s'il y a un ou plusieurs serveurs libres, le client sera immédiatement pris en charge. Sinon, s'il y a des positions d'attente libres, le client va rejoindre la file d'attente selon la discipline de service adaptée dans le système. D'autre part, si un client arrive et trouve tous les serveurs et positions d'attentes occupés, il quitte le système définitivement avec une probabilité $1 - H_0$ ou fait des tentatives de service après une durée de temps aléatoire avec une probabilité H_0 . Entre les tentatives, le client est en orbite et devient source d'appels secondaires (rappels). La capacité de l'orbite O peut être finie ou infinie. Dans le cas où O est finie et si l'orbite est pleine, le client quitte le système pour toujours. Lorsqu'un client est rappelé de l'orbite, il est traité de la même manière qu'un client primaire avec une probabilité H_k (s'il s'agit de la $k^{\text{ème}}$ tentative échouée).

La notation de Kendall est: $A/B/C/m/O/H$, où A et B décrivent respectivement les codes de la distribution du temps inter-arrivées et la distribution du temps de service, C est le nombre de serveurs identiques et indépendants, $m - C$ est la capacité de la file d'attente, O est la capacité de l'orbite, H est la fonction de persistance $H = \{H_k; k \geq 0\}$. Si m , O et H sont absents dans la notation de Kendall, alors $m = C$, $O = \infty$ et $H_k = 1$ pour tout $k \geq 0$. La distribution du temps inter-rappels n'est pas indiquée. Le schéma général d'un système de files d'attente avec rappels est présenté dans la figure 1.3.1.

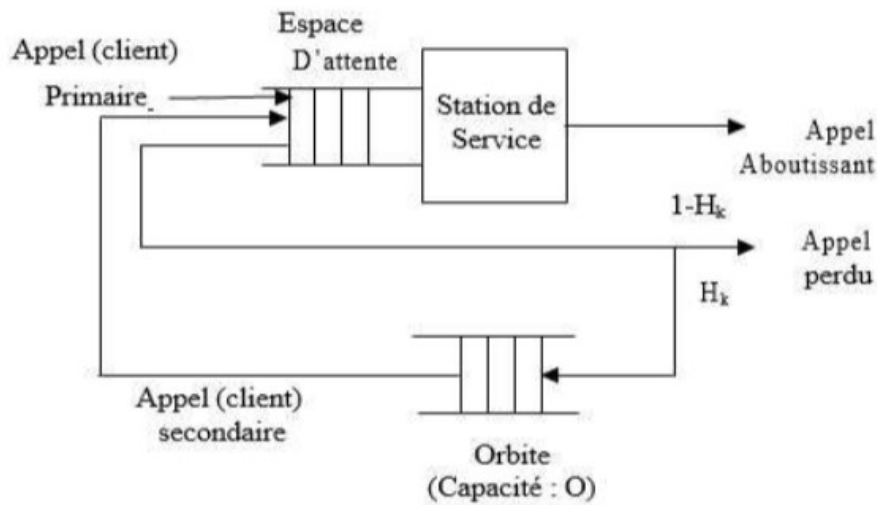


Figure 1.3.1 : Schéma général d'un système avec rappels

1.3.2 Modèles markoviens

1.3.2.1 Modèle M/M/1 avec rappels

On considère un système de files d'attente sans positions d'attente. Le service est assuré par un seul serveur[1]. Les clients primaires arrivent selon un processus de Poisson de taux $\lambda > 0$. Les durées de service suivent une loi exponentielle de fonction de répartition $B(x) = 1 - e^{-\mu x}$, $x \geq 0$ et de moyenne finie $\frac{1}{\mu}$. Les temps entre deux rappels consécutifs sont également exponentiels de paramètre $\theta > 0$ (la fonction de répartition $T(x) = 1 - e^{-\theta x}$, $x \geq 0$). Nous admettons que les durées de service, les durées entre deux rappels consécutifs ainsi que entre deux arrivées primaires successives sont mutuellement indépendantes. L'état du système peut être décrit par le processus

$$\{C(t), N_0(t), t \geq 0\}, \quad (1.3.1)$$

Où $C(t)$ est égale à 0 ou 1 selon le fait que le serveur est libre ou non, $N_0(t)$ est le nombre de clients en orbite l'instant t . Supposons que le régime stationnaire existe ($\rho = \frac{\lambda}{\mu} < 1$). Le processus (1.3.1) est de Markov d'espace d'états $S = \{0,1\} \times N$. Le schéma de modèle M/M/1 avec rappels est donné par la figure 1.3.2.

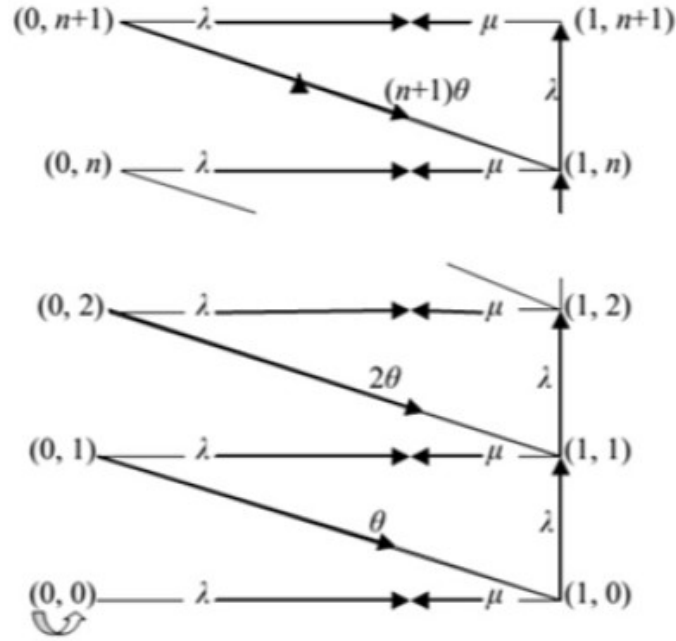


Figure 1.3.2 : modèle M/M/1 avec rappels

Les équations d'équilibre statistique sont:

$$(\lambda + j\theta) p_{0j} = \mu p_{1j}; \quad (1.3.2)$$

$$(\lambda + i\mu) p_{1j} = \lambda p_{0j} + (j + 1)\theta p_{0j+1} + \lambda p_{1j-1}. \quad (1.3.3)$$

Ici, $p_{ij} = \lim_{t \rightarrow \infty} P(C(t) = i, N_0(t) = j)$, $i = 0, 1$ et $j \geq 0$, représentent la distribution stationnaire conjointe de l'état du serveur et du nombre de clients en orbite. Introduisons les fonctions génératrices suivantes:

$$P_0(z) = \sum_{j=0}^{\infty} z^j p_{0j};$$

$$P_1(z) = \sum_{j=0}^{\infty} z^j p_{1j}.$$

À l'aide de ses fonctions et à partir des équations (1.3.2) et (1.3.3), on obtient:

$$P_0(z) = (1 - \rho) \left(\frac{1 - \rho}{1 - z\rho} \right)^{\frac{\lambda}{\theta}}; \quad (1.3.4)$$

$$P_1(z) = \rho \left(\frac{1 - \rho}{1 - z\rho} \right)^{\frac{\lambda}{\theta} + 1}. \quad (1.3.5)$$

Les transformées inverses des (1.3.4) et (1.3.5) nous donnent les formules analytiques explicites [4]:

$$p_{0j} = \frac{\rho}{j!\theta^j} \prod_{k=0}^{j-1} (1 + k\theta)(1 - \rho)^{\frac{\lambda}{\theta}+1};$$

$$p_{1j} = \frac{\rho^{j+1}}{j!\theta^j} \prod_{k=1}^j (\lambda + k\theta)(1 - \rho)^{\frac{\lambda}{\theta}+1};$$

1.3.2.2 Modèles à multiserveurs

Nous Considérons un groupe de C serveurs entièrement disponibles dans lequel un flux des arrivées primaires poissonnien arrive avec un taux $\lambda > 0$. Si un client primaire arrivant trouve un serveur libre, il occupe immédiatement un serveur et quitte le système après le service. Sinon, si tous les serveurs sont occupés, il entre en orbite. Nous supposons que la durée entre deux rappels consécutifs est distribuée exponentiellement avec le paramètre θ , et les temps de service sont distribués exponentiellement avec le paramètre μ .

Le fonctionnement du système peut être décrit par le processus stochastique de Markov $\{C(t), N_0(t); t \geq 0\}$, où $C(t)$ est le nombre de clients dans l'espace de service et $N_0(t)$ est le nombre de clients en orbite à l'instant t [13]. Son espace d'états est donc $S = \{0, 1, \dots, C\} \times N$. Ses taux de transition infinitésimaux $q_{(ij)(nm)}$ sont donnés par:

1. Pour $0 \leq i \leq C - 1$

$$q_{ij}^{(M)}(k, l) = \begin{cases} \lambda & \text{si } (k, l) = (i + 1, j) \\ i\mu & \text{si } (k, l) = (i - 1, j) \\ j\theta & \text{si } (k, l) = (i + 1, j - 1) \quad ; \\ -(\lambda + i\mu + j\theta) & \text{si } (k, l) = (i, j) \\ 0 & \text{sinon} \end{cases}$$

2. Pour $i=C$

$$q_{ij}^{(M)}(k, l) = \begin{cases} \lambda & \text{si } (k, l) = (C, j) + 1 \\ C\mu & \text{si } (k, l) = (C - 1, j) \\ -(\lambda + C\mu) & \text{si } (k, l) = (i, j) \\ 0 & \text{sinon} \end{cases} ;$$

Les caractéristiques importantes de la qualité de service sont:

- Probabilité que tous les serveurs occupés $P_C = \lim_{t \rightarrow \infty} P(C(t) = C)$;
- Nombre moyen de clients en orbite $\bar{N}_0 = \lim_{t \rightarrow \infty} E[N_0(t)]$;
- Nombre moyen de serveurs occupés $\bar{C} = \lim_{t \rightarrow \infty} E[C(t)]$.

La condition suffisante et nécessaire d'existence d'un régime stationnaire du système est $\lambda < C\mu$. La distribution stationnaire $p_{ij} = \lim_{t \rightarrow \infty} p_{ij}(t)$ satisfait le système d'équations de Kolmogorov suivant:

$$(\lambda + i\mu + j\theta)p_{ij} = \lambda p_{i-1, j} + (j+1)\theta p_{i-1, j+1} + (i+1)\mu p_{i+1, j}, \quad (1.3.6)$$

si $0 \leq i \leq C-1$, *et* $j \geq 0$;

$$(\lambda + C\mu)p_{Cj} = \lambda p_{C-1, j} + (j+1)\theta p_{C-1, j+1} + p_{C, j-1}, \quad (1.3.7)$$

si $i = C$ *et* $j \geq 0$.

Pour les fonctions génératrices

$$p_i(z) = \sum_{j=0}^{\infty} z^j p_{ij}, \quad 0 \leq i \leq C$$

Ces équations deviennent

$$(\lambda + i\mu)p_i(z) + \theta z p'_i(z) = \lambda p_{i-1}(z) + \theta p'_{i-1}(z) + (i+1)\mu p_{i+1}(z), \quad (1.3.8)$$

si $0 \leq i \leq C-1$,

$$(\lambda + C\mu)p_C(z) = \lambda p_{C-1}(z) + \theta p'_{C-1}(z) + \lambda z p_C(z), \quad (1.3.9)$$

si $i = C$.

Maintenant, introduisons la fonction génératrice bivariée

$$p(x, z) = \sum_{i=0}^C x^i p_i(z).$$

Nous supposons que le taux de service $\mu = 1$, donc les équations (1.3.8), (1.3.9) deviennent:

$$\begin{aligned} \lambda(1-x)p(x, z) + \theta(z-x)p'_z(x, z) + (x-1)p'_x(x, z) \\ + \lambda x^C(x-z)p_C(z) + \theta x^C(x-z)p'_C(z) = 0. \end{aligned} \quad (1.3.10)$$

En différenciant l'équation (1.3.10) par rapport à z, x, xx, zz au point $x = 1, z = 1$, on obtient les équations suivantes:

$$\begin{aligned} \theta \bar{N}_0 - \lambda P_C - \theta \bar{N}_C &= 0, \\ \lambda + \theta \bar{N}_0 - \bar{C} - \lambda P_C - \theta \bar{N}_C &= 0, \\ \lambda \bar{C} + \theta p''_{xz} - p''_{xx} - \lambda C P_C - \theta C \bar{N}_C &= 0, \\ -\lambda \bar{N}_0 - \theta p''_{zz} + (1 + \theta) p''_{xz} + \lambda - \theta C \bar{N}_C - \lambda C P_C + \theta p''_{Czz} &= 0, \\ \theta p''_{zz} - \lambda \bar{N}_C - \theta p''_{Czz} &= 0, \end{aligned} \quad (1.3.11)$$

Où $\bar{N}_0 = P'_z(1.1)$, $P_C = P_C(1)$, $\bar{C} = P'_x(1.1)$, $\bar{N}_C = \lim_{t \rightarrow \infty} E[N_0(t), C(t) = C] = P'_C(1)$. En éliminant de ces équation les variables $\bar{N}_C, P''_{xz}, P''_{zz}, P''_{Czz}$ et en tenant compte du fait que $P''_{xx}(1.1) = (Var[C(t)] + (E[C(t)]^2) - E[C(t)])$, nous obtenons

$$\bar{C} = \lambda. \quad (1.3.12)$$

$$\bar{N}_0 = \frac{1 + \theta}{\theta} \cdot \frac{\lambda - \lim_{t \rightarrow \infty} Var[C(t)]}{C - \lambda}. \quad (1.3.13)$$

L'équation (1.3.13) peut être réécrite sous une forme équivalente telle que:

$$\bar{N}_0 = \frac{1 + \theta}{\theta} \cdot \frac{\lambda + \lambda^2 - \lim_{t \rightarrow \infty} E[(C(t))^2]}{C - \lambda}. \quad (1.3.14)$$

L'équation (1.3.12) peut être considérée comme une variante de la formule de Little et représente l'égalité du trafic offert et de celui transporté. L'équation (1.3.13) donne une description partielle de la dépendance de la taille moyenne de l'orbite des paramètres du système étudié et réduit le calcul de cette mesure de performance (\bar{N}_0) au calcul des

caractéristiques du nombre de serveurs occupés, ce qui est un problème plus simple.

De même, les moments supérieurs du nombre de clients en orbite peuvent être exprimés en termes de la distribution du nombre de serveurs occupés. Par exemple, le deuxième moment factoriel $\phi_2 \equiv E[N_0(t).(N_0(t) - 1)] = p''_{zz}(1.1)$ est donné par[13] :

$$\begin{aligned} \phi_2 &= \frac{1 + 2\theta}{2\theta^2(C - \lambda)^2} \{ -(2 + \theta)(C - \lambda)p''_{xxx}(1.1) \\ &+ [\lambda(4 + \theta)(C - \lambda) + 2C(1 + \theta)(C - \lambda - 1)]p''_{xx}(1.1) \\ &- 2\lambda^2C(1 + \theta)(C - \lambda - 1) - 2\lambda^3(C - \lambda). \end{aligned}$$

Une autre information intéressante sur la distribution stationnaire du nombre de serveurs occupés $P_i = \lim_{t \rightarrow \infty} P(C(t) = i)$ peut être obtenu comme suit. Mettons dans les équation (1.3.8), (1.3.9) $z = 1$:

$$\begin{aligned} \lambda P_i + \theta \bar{N}_i - (i + 1)P_{i+1} &= \lambda P_{i-1} + \theta \bar{N}_{i-1} - iP_i, \quad 0 \leq i \leq C - 1 \\ \lambda P_{C-1} + \theta \bar{N}_{C-1} - Cp_C &= 0, \end{aligned}$$

Où $\bar{N}_i = \lim_{t \rightarrow \infty} E[N_0(t), C(t) = i]$. Les équations précédentes donnent que

$$\lambda P_i + \theta \bar{N}_i - (i + 1)P_{i+1} = 0, \quad 0 \leq i \leq C - 1 \quad (1.3.15)$$

Définissons le taux du flux des clients secondaires étant donné que le nombre de serveurs occupés est égal à i par $r_i = \frac{\theta \bar{N}_i}{P_i}$. Alors, l'équation (1.3.15) devient:

$$P_{i+1} = \frac{\lambda + r_i}{i + 1} P_i, \quad 0 \leq i \leq C - 1.$$

Enfin

$$P_i = \frac{(\lambda + r_{i-1}) \dots (\lambda + r_0)}{i} P_0, \quad 0 \leq i \leq C, \quad (1.3.16)$$

La probabilité P_0 s'obtient à partir de l'équation de normalisation $\sum_{i=0}^C P_i = 1$. En effet,

$$P_0 = \left[\sum_{i=0}^C \frac{(\lambda + r_{i-1}) \dots (\lambda + r_0)}{i} \right]^{-1}. \quad (1.3.17)$$

Bien que les taux $r_i, 0 \leq i \leq C - 1$, soient inconnus et que les équations (1.3.16)-(1.3.17) ne présentent pas de solution sous forme "close", ces dernières fournissent un aperçu du

problème et seront utilisées plus tard. En outre, l'équation (1.3.12) permet d'avoir une relation, telle que

$$\sum_{i=0}^C i \frac{(\lambda + r_{i-1}) \dots (\lambda + r_0)}{i} = \lambda \sum_{i=0}^C \frac{(\lambda + r_{i-1}) \dots (\lambda + r_0)}{i} \quad (1.3.18)$$

Approches d'analyse des modèles à multiserveurs avec rappels

2.1 Introduction

Due à l'absence des formules analytiques explicites pour les caractéristiques probabilistes principales des modèles de files d'attente avec rappels à multiserveurs, la seule méthode pour obtenir des données numériques précises consiste à résoudre les équations de Kolmogorov numériquement. Mais dès que ce système d'équations est infini, il ne peut pas être résolu directement même sur ordinateur. Les transformations qui réduisent ces équations à une solution d'un petit problème fini dans le cas général ne sont pas disponibles. C'est pourquoi, nous nous adressons aux approches d'approximation.

Dans ce chapitre, nous présentons deux approches d'obtention des solutions, rencontrées dans la littérature, pour les modèles avec rappels et multiserveur.

2.2 modèles tronqués

2.2.1 Description du modèle

Dans ce modèle, par opposition au modèle principal, la taille de l'orbite est limité par un constante M . Si le nombre de sources est égal à M , alors les clients bloqués sont perdus et n'ont aucune influence sur le fonctionnement du système. La dynamique stochastique du système peut être décrite par un processus bidimensionnel[13]

$$\{C^{(M)}(t), N_0^{(M)}(t)\},$$

où $C^{(M)}(t)$ est le nombre de serveurs occupés et $N_0^{(M)}(t)$ est le nombre de clients en orbite à l'instant t . Sous les hypothèses ci-dessus, le processus $(C^{(M)}(t), N_0^{(M)}(t))$ est Markovien avec l'espace d'états $S^{(M)} = \{0, 1, \dots, C\} \times \{0, 1, \dots, M\}$. Les taux de transition $q_{(ij)(nm)}^{(M)}$ sont donnés par:

1. Pour $0 \leq i \leq C - 1, 0 \leq j \leq M$

$$q_{ij}^{(M)}(k, l) = \begin{cases} \lambda & \text{si } (k, l) = (i + 1, j) \\ i\mu & \text{si } (k, l) = (i - 1, j) \\ j\theta & \text{si } (k, l) = (i + 1, j - 1) \\ -(\lambda + i\mu + j\theta) & \text{si } (k, l) = (i, j) \\ 0 & \text{sinon} \end{cases} ;$$

2. Pour $i = C, 0 \leq j \leq M - 1$

$$q_{ij}^{(M)}(k, l) = \begin{cases} \lambda & \text{si } (k, l) = (C, j + 1) \\ C\mu & \text{si } (k, l) = (C - 1, j) \\ -(\lambda + C\mu) & \text{si } (k, l) = (C, j) \\ 0 & \text{sinon} \end{cases} ;$$

3. Pour $i = C, j = M$

$$q_{ij}^{(M)}(k, l) = \begin{cases} C\mu & \text{si } (k, l) = (C - 1, M) \\ -C\mu & \text{si } (k, l) = (C, M) \\ 0 & \text{sinon} \end{cases} ;$$

Ainsi les taux $q_{(ij)(nm)}^{(M)}$ sont les mêmes que ceux du modèle principal [10] sauf pour $i = C, j = M$.

2.2.2 Régime stationnaire

Puisque l'espace d'état du processus $\{C^{(M)}(t), N_0^{(M)}(t)\}$ est fini, le processus est toujours ergodique. La distribution stationnaire $P_{ij}^{(M)} = \lim_{t \rightarrow \infty} P(C^{(M)}(t) = i, N_0^{(M)}(t) = j)$ satisfait le système d'équations suivant:

$$\begin{aligned} (\lambda + i\mu + j\theta)p_{ij}^{(M)} &= \lambda p_{i-1,j}^M + (j+1)\theta p_{i-1,j+1}^{(M)} + (i+1)\mu p_{i+1,j}^{(M)}, \\ 0 \leq i \leq C-1, 0 \leq j \leq M-1; \end{aligned} \quad (2.2.1)$$

$$\begin{aligned} (\lambda + i\mu + M\theta)p_{i,M}^{(M)} &= \lambda p_{i-1,M}^{(M)} + (i+1)\mu p_{i+1,M}^{(M)}, \\ 0 \leq i \leq C-1, j = M; \end{aligned} \quad (2.2.2)$$

$$\begin{aligned} (\lambda + C\mu)p_{i,j}^{(M)} &= \lambda p_{i-1,j}^M + (j+1)\theta p_{i-1,j+1}^{(M)} + \lambda p_{i,j-1}^{(M)}, \\ i = C, 0 \leq j \leq M-1; \end{aligned} \quad (2.2.3)$$

$$\begin{aligned} C\mu p_{C,M}^{(M)} &= \lambda p_{C-1,M}^{(M)} + \lambda p_{C,M-1}^{(M)}, \\ i = C, j = M; \end{aligned} \quad (2.2.4)$$

qui satisfait à la condition de normalisation:

$$\sum_{i=0}^C \sum_{j=0}^M p_{ij}^{(M)} = 1. \quad (2.2.5)$$

2.2.3 Condition d'existence d'un régime stationnaire

2.2.3.1 Condition suffisante

Théorème 2.2.1 [13] *Soit $X(t)$ un processus de Markov avec l'espace des états discret S et les taux de transition q_{sp} , où $s, p \in S, \sum_p q_{sp} = 0$. Supposons qu'il existe:*

1. *une fonction $\varphi(s), s \in S$ (cette fonction est dite une fonction de Lyapunov).*

2. un nombre positif ϵ tel que:

- les variables $y_s = \sum_{p \neq s} q_{sp}(\varphi(p) - \varphi(s)) < \infty$ pour tout $s \in S$;
- $y_s \leq -\epsilon$ pour tous les $s \in S$ sauf un nombre fini d'états.

Alors le processus $X(t)$ est régulier et ergodique.

Preuve. Pour les systèmes de files d'attente avec rappels, une combinaison linéaire des coordonnées du vecteur de Processus de Markov qui décrit le fonctionnement d'un modèle généralement [11] [12] peut être utilisé comme la fonction de Lyapunov $\varphi(s)$.

$$\varphi(s) \equiv \varphi(i, j) = ai + j,$$

où a est un paramètre, qui sera déterminé par la suite. Alors les accroissements moyens $y_s \equiv y_{ij}$ sont donnés par:

$$y_{ij} = \begin{cases} \lambda a + i\mu(-a) + j\theta(a-1), & \text{si } 0 \leq i \leq C-1, \\ \lambda + C\mu(-a), & \text{si } i = C. \end{cases}$$

Puisque pour tout $i = 0, 1, \dots, C$ il existe

$$\lim_{j \rightarrow \infty} y_{ij} = L_i = \begin{cases} (a-1) \cdot \infty, & \text{si } 0 \leq i \leq C-1, \\ \lambda - aC\mu, & \text{si } i = C. \end{cases};$$

les suppositions du théorème restent vrai si toutes les variables L_i sont négatives, c'est-à-dire:

$$\begin{aligned} a - 1 &< 0; \\ \lambda - a[C\mu] &< 0. \end{aligned}$$

Ces conditions [2] représentent un ensemble d'inégalités linéaires pour le paramètre inconnu a . Ces conditions peuvent être écrites sous la forme :

$$\frac{\lambda}{C\mu} < a < 1.$$

■

2.2.3.2 Condition nécessaire

La non ergodicité peut être déterminée aussi à l'aide des fonctions de Lyapunov. Le théorème suivant est le plus convenable.

Théorème 2.2.2 [13] *Soit $\{\zeta_n\}$ une chaîne de Markov avec l'espace des états S et les probabilités de transition r_{sp} . Supposons qu'il existe:*

1. une fonction positive $\varphi(s)$, $s \in S$, telle que pour un certain d on a

$$r_{sp} \neq 0 \implies |\varphi(s) - \varphi(p)| \leq d;$$

2. un nombre positif b tel que:

- L'ensemble $A_b \equiv \{s \in S \mid \varphi(s) > b\} \neq \emptyset$;
- $\inf_{s \in A_b} x_s = \inf_{s \in A_b} E(\varphi(\zeta_{n+1}) - \varphi(s) \mid \zeta_n = s)$ est non négatif (ou positif).

Alors la chaîne $\{\zeta_n\}$ est non ergodique (n'admet pas un régime stationnaire en elle est transitoire).

Preuve. On considère le modèle principale S qui a pour processus de Markov $\{C(t), N_0(t), t \geq 0\}$.

Les probabilités de transition de la chaîne de Markov $\{\zeta_n\}$ sont données par:

1. Pour $0 \leq i \leq C - 1$

$$r_{ij}(k, l) = \begin{cases} \frac{\lambda}{\lambda + i\mu + j\theta} & \text{si } (k, l) = (i + 1, j) \\ \frac{i\mu}{\lambda + i\mu + j\theta} & \text{si } (k, l) = (i - 1, j) \\ \frac{j\theta}{\lambda + i\mu + j\theta} & \text{si } (k, l) = (i + 1, j - 1) \\ 0 & \text{sinon} \end{cases} ;$$

2. Pour $i = C$

$$r_{ij}(k, l) = \begin{cases} \frac{\lambda}{\lambda + C\mu} & \text{si } (k, l) = (C, j + 1) \\ \frac{C\mu}{\lambda + C\mu} & \text{si } (k, l) = (C - 1, j) \\ 0 & \text{sinon} \end{cases} ;$$

Soit $\lambda \geq C\mu$. Considérons la fonction de Lyapunov $\varphi(i, j) = i + j$. Alors:

1. Si $r_{(ij)(kl)} \neq 0$ alors $|i - k| \leq 1, |j - l| \leq 1$ et par conséquent

$$|\varphi(i, j) - \varphi(k, l)| = |i - k + j - l| \leq |i - k| + |j - l| \leq 2,$$

c'est-à-dire $d = 2$.

2. Puisque la variable b peut prendre un nombre positif quelconque. Alors:

(a) $A_b = (i, j) | i = 0, 1, \dots, C; j > \max(b - i, 0) \neq \emptyset$;

(b)

$$x_{ij} = \begin{cases} \frac{\lambda - i\mu}{\lambda + i\mu + j\theta} & \text{si } 0 \leq i \leq C - 1 \\ \frac{\lambda - C\mu}{\lambda + C\mu} & \text{si } i = C \end{cases}.$$

Par conséquent, pour $\lambda \geq C\mu$ la variable x_{ij} est toujours non négative, de plus $x_{ij} \geq 0$ pour $(i, j) \in A_b$. Alors, pour $\lambda \geq C\mu$ le processus $\{C(t), N_o(t); t \geq 0\}$ est non ergodique.

A présent, soit $\lambda > C\mu$. On considère la fonction de Lyapunov $\varphi(i, j) = ai + j$ où $a \in (1, \lambda/C\mu)$. Alors:

1. Si $r_{ij}(n, m) \neq 0$, alors $|i - n| \leq 1, |j - m| \leq 1$ et par conséquent

$$|\varphi(i, j) - \varphi(n, m)| = |a(i - n) + j - m| \leq a|i - n| + |j - m| \leq a + 1,$$

c'est à dire $d = a + 1$.

2. L'ensemble A_b est non vide quelque soit $b > 0$, alors on doit choisir b de telle façon que $\inf_{(i,j) \in A_b} x_{ij} > 0$. Mais pour $i = C$

$$x_{Cj} = \frac{\lambda - aC\mu}{\lambda + C\mu} > 0 \text{ pour tout } j,$$

et pour $0 \leq i \leq C - 1$

$$\lim_{j \rightarrow \infty} x_{ij} = a - 1 > 0.$$

Alors, pour $i = 0, 1, \dots, C$ il existe N_i tel que

$$x_{ij} > \frac{a - 1}{2} \text{ pour tout } N_i.$$

Si

$$\epsilon = \min \left\{ \frac{\lambda - aC\mu}{\lambda + C\mu}, \frac{a-1}{2} \right\},$$

alors $x_{ij} \geq \epsilon$ pour tout $(i, j) \in S$ sauf que pour (i, j) vérifiant $i = 0, 1, \dots, C-1$ et $0 \leq j \leq N_i - 1$. Soit b suffisamment grand de telle façon que l'ensemble A_b ne contient pas ces états. Alors

$$x_{ij} \geq \epsilon, \text{ pour tout } (i, j) \in A_b,$$

c'est à dire

$$\inf_{(i,j) \in A_b} x_{ij} \geq \epsilon > 0.$$

Alors, pour $\lambda > C\mu$ la chaîne $\{C(t), N_o(t), t \geq 0\}$ est transitoire. Par conséquent, la condition $\rho = \frac{\lambda}{C\mu} < 1$ implique l'ergodicité et $\rho = \frac{\lambda}{C\mu} > 1$ implique que la chaîne $\{C(t), N_o(t), t \geq 0\}$ est transitoire.

2.2.4 Un algorithme pour le calcul numérique de la distribution stationnaire

Pour obtenir une précision correcte de la solution dans le cas d'un trafic fort ou internité de rappels faible, la valeur M doit être vraiment large (environ les centaines). Si le nombre de serveurs est aussi large, alors le nombre d'inconnues peut atteindre quelques centaines de milliers et, ainsi l'application de la méthode de troncation rencontre des difficultés considérables. Donc il est important pour les modèles essentiels de bien simplifier les calculs. Comment ceci peut être réalisé? une idée est due à Wilkinsen [29]:

On introduit les nouvelles variables $r_{ij}^{(M)}$, $0 \leq i \leq C$, $0 \leq j \leq M$, comme suit:

$$r_{ij}^{(M)} = \frac{P_{ij}^{(M)}}{P_{0M}^{(M)}}.$$

Si on peut trouver les variables $r_{ij}^{(M)}$, alors on peut calculer les probabilités $P_{ij}^{(M)}$:

$$P_{ij}^{(M)} = \frac{r_{ij}^{(M)}}{\sum_{i=0}^C \sum_{j=0}^M r_{ij}^{(M)}}.$$

Et on peut calculer les principales caractéristiques probabilistes du système tronqué $S^{(M)}$ [13]:

- La probabilité de borage:

$$P_C^{(M)} = \sum_{j=0}^M r_{Cj}^{(M)} \cdot P_{0M}^{(M)};$$

- Le nombre moyen de serveurs occupés:

$$\bar{C}^M = \sum_{i=0}^C \sum_{j=0}^M i r_{ij}^{(M)} \cdot P_{0M}^{(M)};$$

- Le nombre moyen de clients en orbite:

$$\bar{N}_0^{(M)} = \sum_{i=0}^C \sum_{j=0}^M j r_{ij}^{(M)} \cdot P_{0M}^{(M)}.$$

Les variables $r_{ij}^{(M)}$ satisfont l'ensemble d'équations suivant, obtenu à partir des équations (2.2.1) – (2.2.5) pour les probabilités $P_{ij}^{(M)}$:

$$r_{0M}^{(M)} = 1, \tag{2.2.6}$$

$$(\lambda + i\mu + j\theta)r_{ij}^{(M)} = \lambda r_{i-1,j}^{(M)} + (j+1)\theta r_{i-1,j+1}^{(M)} + (i+1)\mu r_{i+1,j}^{(M)}, \tag{2.2.7}$$

$$si \ 0 \leq i \leq C-1, \ 0 \leq j \leq M-1,$$

$$(\lambda + i\mu + M\theta)r_{iM}^{(M)} = \lambda r_{i-1,M}^{(M)} + (i+1)\mu r_{i+1,M}^{(M)}, \tag{2.2.8}$$

$$si \ 0 \leq i \leq C-1, \ j = M,$$

$$(\lambda + C\mu)r_{Cj}^{(M)} = \lambda r_{C-1,j}^{(M)} + (j+1)\theta r_{C-1,j+1}^{(M)} + \lambda r_{C,j-1}^{(M)}, \tag{2.2.9}$$

$$si \ i = C, \ 0 \leq j \leq M-1,$$

$$C\mu r_{CM}^{(M)} = \lambda r_{C-1,M}^{(M)} + \lambda r_{C,M-1}^{(M)}. \tag{2.2.10}$$

On commence à partir de $r_{0M}^{(M)} = 1$. Puis, les variables $r_{ij}^{(M)}$ peuvent être calculées récursivement. L'algorithme a été publié par Johnin et Sedol .

2.3 la méthode d'extrapolation de valeur

L'approche présentée ci-dessus fournit une solution numérique [17] à la distribution stationnaire P_{ij} des chaînes de Markov à temps continu. Elles utilisent les équations de Kolmogorov pour calculer les caractéristiques de performance souhaitées. Le problème à résoudre est de trouver la solution du système d'équation suivant:

$$P_{ij} \sum_{s' \neq s} q_{s's} = \sum_{s' \neq s} q_{s's} P_{ij}, \forall s \in S;$$

$$\sum_s P_s = 1,$$

où $q_{s's}$ représente le taux de transition de l'état $s = (i, j)$ à $s' = (k, l)$.

Dans ce qui suit, nous considérons une approche alternative, appelée Extrapolation de valeur [15]. L'approche en question est basée sur les outils de la théorie des Processus de Markov Décisionnels et également sur le principe de troncation. L'idée est de trouver la valeur relative moyenne en résolvant les équations de Howard, écrites pour un espace d'états tronqué. Sa particularité consiste dans le fait qu'au lieu d'une simple troncation, les valeurs d'état relatives à l'extérieur de l'espace d'états tronqué sont estimées par l'extrapolation polynomiale des valeurs d'état relatives à l'intérieur de l'espace tronqué en question[18]. De cette manière, nous obtenons un système fermé ainsi que les résultats précis avec de faibles niveaux de troncation.

2.3.1 Processus de Markov Décisionnels

Un processus de Markov Décisionnel peut être défini comme $\{S, A, P, R\}$, où S est un ensemble d'états, A est un ensemble d'actions, P est une fonction de transition d'état, et R est une fonction de revenu. L'état du système peut être contrôlé, en choisissant les actions "a" parmi "A", influençant ainsi les transitions d'état. La fonction de transition $P : S \times S \times A \rightarrow R_+$ spécifie le taux de transition entre une

paire d'états lorsqu'une certaine action est prise à l'état d'origine[27].

La première caractéristique de la technique d'extrapolation de valeur est la nécessité de la définition d'une fonction de revenu qui doit être une fonction de l'état du système[26], qui est $r(s)$. À la suite de la définition de la fonction de revenu pour chaque état, en régime stationnaire, un taux de revenu moyen du processus peut être introduit comme suit $\bar{r} = \sum_{(i,j) \in S} P_{ij}r(i, j)$. Dans la technique d'extrapolation de valeur, la fonction de revenu R doit être définie de sorte que le revenu moyen résultant \bar{r} coïncide avec la mesure de performance souhaitée. Une fois que nous avons défini le cadre PMD ainsi que la fonction de revenu, nous sommes en position pour définir les valeurs d'état relatives. Il est évident qu'après avoir exécuté une action dans un état $s \in S$ le système collectera un revenu pour cette action ($r(s)$), mais comme le nombre de transitions augmente, le revenu moyen collecté converge vers \bar{r} . La valeur d'état relative $v(s)$ montre la différence entre le total des revenus cumulé lorsque le processus débute à l'état s et le revenu total cumulé par le processus, le taux moyen de revenu étant \bar{r} :

$$v(s) = E\left[\int_0^{\infty} (r(S(t)) - \bar{r})dt / S(0) = s\right].$$

Les équations de Howard [21] relient les revenus, les valeurs d'état relatives et les taux de transition de la manière suivante:

$$r(s) - \bar{r} + \sum q_{ss'}(v(s') - v(s)) = 0; \forall s \in S.$$

Les équations de Howard qui correspondent au système M/M/C avec rappels sont:

$$\begin{aligned} r(i, j) - \bar{r} + \lambda(v(i+1, j) - v(i, j)) + i\mu(v(i-1, j) - v(i, j)) \\ + j\theta(v(i+1, j-1) - v(i, j)) = 0, \\ 0 \leq i \leq C-1; \end{aligned}$$

$$\begin{aligned} r(i, j) - \bar{r} + \lambda(v(C, j+1) - v(C, j)) + C\mu(v(C-1, j) - v(C, j)) = 0, \\ i = C. \end{aligned}$$

Comme nous pouvons observer le nombre d'états est infini car j peut prendre n'importe quelle valeur dans Z_+ , nous devons donc tronquer l'espace d'état S à $S^{(M)}$:

$$S^{(M)} = \{s = (i, j); 0 \leq i \leq C; 0 \leq j \leq M\}.$$

En effet, il y aura autant d'équations de Howard que de nombre d'états, $|S^{(M)}|$. Le nombre d'inconnus sera les $|S^{(M)}|$ valeurs d'état relatives plus le revenu attendu \bar{r} , c'est-à-dire $|S^{(M)}| + 1$ inconnues. Cependant, comme seules les différences dans les valeurs relatives apparaissent dans les équations de Howard, nous pouvons définir $v(0) = 0$. Par conséquent on aura un système linéaire d'équations résoluble ayant le même nombre d'équations que le nombre d'inconnues.

2.3.2 Ajustement polynomial

L'Extrapolation de Valeur considère les valeurs d'état relatives en dehors $S^{(M)}$ qui apparaissent dans les équations de Howard comme une extrapolation de certaines valeurs d'état relatives correspondant aux états se trouvant à l'intérieur $S^{(M)}$. En résumé, l'objectif de l'extrapolation de valeur est de trouver une fonction d'extrapolation qui correspond à certains points de $S^{(M)}$ de sorte qu'il se rapproche également des points extérieurs de $S^{(M)}$. Il est important de choisir une fonction d'ajustement qui fait que les équations de Howard forment un système fermé d'équations linéaires. Les fonctions d'ajustement les plus courantes qui remplissent cette condition sont les polynômes. Nous pouvons utiliser tous les états de $S^{(M)}$ dans la procédure d'ajustement global ou, ce qui est le plus souvent utilisé, seulement un sous-ensemble S_f de leur localisation. Par souci de simplicité, dans la description suivante[25], nous supposons qu'il existe une application W à partir de l'ensemble bidimensionnel d'états vers un ensemble à une dimension, par exemple, les nombres réels: $W : \bar{S}_f \rightarrow R$.

Par conséquent, l'application W traite des états comme s'il s'agissait de valeurs réelles $w = W(s)$. Le choix de W dépendra fortement des états dont lesquels nous voulons extrapoler sa valeur d'état relative. Notons également que la fonction $f(w)$ et l'ensemble $S_f^{(M)}$ doivent être choisis de sorte que les paramètres de $f(w)$ des valeurs non ambigus, c'est-à-dire dans le cas du choix d'un polynôme comme fonction d'interpolation, le nombre de points différents dans $S_f^{(M)}$ doit être égal ou supérieur au nombre de coefficients dans le polynôme. En général, la procédure du calcul des coefficients du polynôme consiste à minimiser l'erreur quadratique minimale:

$$E = \sum_{w \in W} (f(w) - v(w))^2. \quad (2.3.1)$$

Alors les valeurs optimales pour les coefficients du polynôme a_i peuvent être calculées par en résolvant les équations:

$$\frac{\partial E}{\partial a_i} = 0, \quad \forall i. \quad (2.3.2)$$

Dans notre cas, nous utilisons autant de points que le nombre de paramètres du polynôme d'interpolation, de sorte que la procédure d'ajustement est une interpolation polynomiale ordinaire et $E = 0$, c'est-à-dire que tous les points considérés se trouveront dans la courbe du polynôme. Dans ce cas, le problème peut être formulé comme suit: Étant donné un ensemble de $n = (W(S_f^{(M)})) = |S_f^{(M)}|$ points $(w_0, v(w_0), \dots, (w_{n-1}, v(w_{n-1})))$, on peut déterminer un polynôme de degré $(n - 1)$ de sorte que $f(w_i) = v(w_i)$, pour $i = 0, \dots, n - 1$, où

$$f(w) = a_0 + a_1w + a_2w^2 + \dots + a_{n-1}w^{n-1}.$$

Le polynôme d'interpolation satisfait aux n équations linéaires suivantes:

$$f(w_i) = a_0 + a_1w_i + a_2w_i^2 + \dots + a_{n-1}w_i^{n-1} = v(w_i), \quad i = 0, \dots, n - 1.$$

Ces dernières, sous forme matricielle, se présentent de la manière suivante:

$$Aa = \begin{bmatrix} 1 & w_0 & \cdots & w_0^{n-1} \\ 1 & w_1 & \cdots & w_1^{n-1} \\ \vdots & \vdots & \cdots & \vdots \\ 1 & w_{n-1} & \cdots & w_{n-1}^{n-1} \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_{n-1} \end{bmatrix} = \begin{bmatrix} v(w_0) \\ v(w_1) \\ \vdots \\ v(w_{n-1}) \end{bmatrix} = b.$$

La matrice de coefficients de ce système A est une matrice de Vandermonde, dont le déterminant est non nul et donc A est inversible. Ainsi, il existe toujours une unique solution au système linéaire d'équations considéré ou, de manière équivalente, il existe un unique polynôme qui passe par tous les n points. Cependant, les matrices de Vandermonde sont souvent mal conditionnées, surtout si certains w_i sont très proches, donc la procédure pour calculer le polynôme est également mal conditionnée. Il est important de noter que l'unicité du polynôme d'interpolation ne signifie pas qu'il ne peut pas être écrit dans une base différente de la base standard. Plus concrètement dans ce travail, nous avons utilisé la base de Lagrange. Pour le problème d'interpolation considéré, le polynôme est une combinaison linéaire

$$L(w) = \sum_{j=0}^{n-1} v(w_j) l_j(w).$$

où

$$l_j(w) = \prod_{i=0, i \neq j}^{n-1} \frac{w - w_i}{w_j - w_i} = \frac{w - w_0}{w_j - w_0} \cdots \frac{w - w_{j-1}}{w_j - w_{j-1}} \frac{w - w_{j+1}}{w_j - w_{j+1}} \cdots \frac{w - w_{n-1}}{w_j - w_{n-1}};$$

Pour le problème étudié, nous aurons une équation de Howard dans laquelle apparaît $v(C, M + 1)$, c'est-à-dire une valeur relative d'un état qui n'appartient pas à $S^{(M)}$. Nous devons donc approximer la valeur relative $v(C, M + 1)$ en utilisant certaines valeurs d'état relatives des états appartenant à $S^{(M)}$. Il est important de souligner que pour l'extrapolation de $v(C, M + 1)$, nous n'utilisons que des états de la forme $s = (C, j)$, avec j variables. De ce fait, on définit l'application w comme $w((C, j)) = j$. De plus, on utilise un polynôme de degré $(n - 1)$ qui interpole les n points de $S_f = \{s_i = (C, M - i) / i = 0, \dots, n - 1\}$ et alors $w(S_f) = \{w_i = M - i / i = 0, \dots, n - 1\}$.

En effet,

$$\begin{aligned}
 w_0 = M &\rightarrow v(w_0) = v(C, M); \\
 w_1 = M - 1 &\rightarrow v(w_1) = v(C, M - 1); \\
 &\vdots \\
 w_j = M - j &\rightarrow v(w_j) = v(C, M - j); \\
 w_{n-1} = M - (n - 1) &\rightarrow v(w_{n-1}) = v(C, M - (n - 1)).
 \end{aligned}$$

De cette façon, la forme générale de l'état d'extrapolation lors de l'utilisation d'un polynôme de degré $(n - 1)$ est:

$$v^{(n)}(C, M + 1) = L^{(n)}(M + 1) = \sum_{j=0}^{n-1} v(C, M - j)l_j(M + 1).$$

Par exemple, dans le cas d'une extrapolation linéaire $n = 2$, on utilise $(M, v(C, M))$ et $(M - 1, v(C, M - 1))$:

$$\begin{aligned}
 v^{(2)}(C, M + 1) &= L^{(2)}(M + 1) = v(C, M)l_0(M + 1) + v(C, M - 1)l_1(M + 1) \\
 &= v(C, M) \frac{(M + 1) - (M - 1)}{M - (M - 1)} + v(C, M - 1) \frac{(M + 1) - M}{(M - 1) - M} \\
 &= 2v(C, M) - v(C, M - 1).
 \end{aligned}$$

suivant la procédure similaire, nous obtenons les relations suivantes pour $n = 3$ et $n = 4$:

$$\begin{aligned}
 v^{(3)}(C, M + 1) &= 3v(C, M) - 3v(C, M - 1) + v(C, M - 2); \\
 v^{(4)}(C, M + 1) &= 4v(C, M) - 6v(C, M - 1) + 4v(C, M - 2) - v(C, M - 3).
 \end{aligned}$$

Finalement

$$v^{(n)}(C, M + 1) = \sum_{k=0}^{n-1} (-1)^k \binom{n}{k+1} v(C, M - k).$$

où n est le nombre de coefficients pris pour les polynômes de Lagrange.

2.3.3 Fonction Revenu

Par définition, $r(s)$ est le taux de revenu obtenu lorsque le système est dans l'état s . Par conséquent, nous devons définir le revenu comme la mesure de performance que nous voulons calculer. De plus, les entrées $r(s)$ dans les équations de Howard doivent être correctement définies. Le tableau 1 donne plusieurs exemples sur la façon dont $r(s)$ peuvent être déterminés pour obtenir certains mesure de performance. Par exemple, nous choisissons le nombre moyen de client en orbite et nous définissons la fonction de revenu comme étant j pour tout les états de $S^{(M)}$, C'est-à-dire $r(i, j) = j$, pour $0 \leq i \leq C$ et $0 \leq j \leq M$, et $r(i, j) = 0$ dans les états qui se trouve en dehors de $S^{(M)}$.

Table 2.1: Définition de la fonction revenu.

Probabilité de blocage	P_C	$r(i, j) = 1$ pour $i = C$ et $0 \leq j \leq M$ $r(i, j) = 0$ sinon
Nombre moyen de clients en orbite	\bar{N}_o	$r(i, j) = j$, pour $0 \leq i \leq C$ et $0 \leq j \leq M$ $r(i, j) = 0$ sinon
Nombre moyen de serveurs occupés	\bar{C}	$r(i, j) = i$, pour $0 \leq i \leq C$ et $0 \leq j \leq M$ $r(i, j) = 0$ sinon

2.3.4 Effet de l'extrapolation de la valeur dans les équations d'Howard

Dans notre problème, et comme mentionné ci-dessus, nous n'aurons qu'à remplacer $v(C, M + 1)$ par sa valeur approximative dans l'équation de Howard qui correspond à l'état $v(C, M + 1)$ Par exemple, si nous utilisons l'extrapolation linéaire $n = 2$, cette équation devient

$$\begin{aligned} & r(C, M) - r + v(C, M)(-\lambda - C\mu) + \lambda v(C, M + 1) + C\mu v(C - 1, M) \\ & = r(C, M) - r + v(C, M)(\lambda - C\mu) + C\mu v(C - 1, M) + v(C, M - 1) = 0. \end{aligned}$$

Comme $v(C, M + 1)$ n'apparaît plus dans les équations de Howard, nous avons un système linéaire de $(C + 1) \times (M + 1)$ équations ayant le même nombre d'inconnues. Ce système peut être exprimé sous forme matricielle pour des raisons de simplicité. Par conséquent, le système peut être réécrit comme $xT = b$, où x est un vecteur avec les inconnues $(C + 1) \times (M + 1)$ (\bar{r} et les valeurs d'état relatives $v(s)$) et b comporte les taux de revenu négatifs pour les différents états [15]:

$$x = [r, v(0, 1), \dots, v(0, Q), v(1, 0), \dots, v(C, M)];$$

$$b = [-r(0, 0), -r(0, 1), \dots, -r(C, 0), -r(C, M)].$$

La matrice T représente la matrice des coefficients et peut être construite en mettant tous les éléments de la première rangée de la matrice T_0 égale à -1. La matrice T_0 est donnée par:

$$T_0 = \begin{bmatrix} A_1^0 & A_0^0 & O & \cdots & O & O & O \\ A_2^1 & A_1^1 & A_0^1 & \cdots & O & O & O \\ O & A_2^2 & \ddots & \ddots & O & O & O \\ \vdots & \vdots & \ddots & A_1^C & A_0^C & \vdots & \vdots \\ O & O & O & \cdots & \ddots & A_1^{C-1} & A_0^{C-1} \\ O & O & O & \cdots & O & A_2^C & A_1^C \end{bmatrix},$$

Où les sous-matrices sont définies comme

$$A_0^i = (i + 1)\mu I, \quad 0 \leq i \leq C - 1;$$

$$A_2^i = \begin{bmatrix} \lambda & \theta & 0 & \cdots & 0 & 0 \\ 0 & \lambda & 2\theta & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda & M\theta \\ 0 & 0 & 0 & \cdots & 0 & \lambda \end{bmatrix}, \quad 1 \leq i \leq C;$$

$$A_1^i = \begin{bmatrix} \alpha & 0 & 0 & \cdots & 0 \\ 0 & \alpha - \theta & 0 & \cdots & 0 \\ 0 & 0 & \alpha - 2\theta & \ddots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \alpha - M\theta \end{bmatrix}, \text{ pour } \alpha = -\lambda - i\mu, 0 \leq i \leq C - 1$$

Lorsque $i = C$, en utilisant l'extrapolation linéaire $n = 2$ et quadratique $n = 3$, nous obtenons respectivement:

$$A_1^C = \begin{bmatrix} \beta & 0 & \cdots & 0 & 0 \\ \lambda & \beta & \cdots & 0 & 0 \\ 0 & \lambda & \cdots & 0 & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & \cdots & \beta & -\lambda \\ 0 & 0 & \cdots & \lambda & \lambda - C\mu \end{bmatrix},$$

$$A_1^C = \begin{bmatrix} \beta & 0 & \cdots & 0 & 0 \\ \lambda & \beta & \cdots & 0 & 0 \\ 0 & \lambda & \cdots & 0 & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & \lambda \\ 0 & 0 & \cdots & \beta & -3\lambda \\ 0 & 0 & \cdots & \lambda & 2\lambda - C\mu \end{bmatrix},$$

Où $\beta = -\lambda - C\mu$. En général, si l'extrapolation est faite avec $n \leq M + 1$ points, la matrice A_1^C est donnée par:

$$A_1^C = \begin{bmatrix} \beta & 0 & \cdots & 0 & \lambda c_M^{(n)} \\ \lambda & \beta & \cdots & 0 & \lambda c_{M-1}^{(n)} \\ 0 & \lambda & \cdots & 0 & \lambda c_{M-2}^{(n)} \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & \lambda c_2^{(n)} \\ 0 & 0 & \cdots & \beta & \lambda c_1^{(n)} \\ 0 & 0 & \cdots & \lambda & -\lambda - C\mu + \lambda c_0^{(n)} \end{bmatrix}$$

Où

$$c_l^{(n)} = \begin{cases} (-1)^l C_n^{l+1}, & \text{si } l < n \\ 0, & \text{si } l \geq n \end{cases} .$$

Nous soulignons que la taille de la matrice T ne dépend pas du degré de polynôme utilisé pour effectuer l'extrapolation, seule la dernière colonne de la matrice A dépend de l'ajustement polynômial. Cette caractéristique à l'avantage suivant: il n'y aura pas de différence dans le coût de calcul lorsqu'on utilise un degré d'extrapolation plus élevé.

L'inconvénient principal de la technique d'extrapolation de la valeur est que cette technique est seulement capable de calculer une mesure de performance chaque fois que nous résolvons le système. Néanmoins, nous pouvons surmonter cet inconvénient de la suivante. D'une manière générale, la solution du système $xT = b$ peut être obtenue en utilisant la matrice inverse de T en faisant $x = bT^{-1}$. Notons également que le choix d'une autre mesure de performance n'affectera que les valeurs dans b . Par conséquent, le calcul d'une seconde mesure de performance augmente seulement les dépenses de calcul par le coût du produit bT^{-1} , car le reste du processus de calcul de la matrice inverse T^{-1} n'est résolu qu'une seule fois.

Étude et analyse du modèle d'attente

M/M/C/K avec rappels

3.1 Introduction

Dans ce chapitre, nous réalisons l'analyse stochastique du modèle à l'étude : M/M/C avec rappels, espace d'attente de capacité limitée, abandons exponentiels. À cet effet, nous donnons une description mathématique du modèle, trouvons les conditions d'ergodicité, présentons les étapes d'application des méthodes de résolution approximative : Troncature Finie et Extrapolation de Valeur.

3.2 Description du modèle

Nous considérons un système de files d'attente avec C serveurs et $K - C$ positions d'attente vers lequel les clients primaires arrivent selon des processus de Poisson avec un taux $\lambda > 0$. Un client primaire reçoit un service immédiat s'il y a un serveur inactif; sinon nous avons deux possibilités selon le fait que le nombre de clients dans l'espace d'attente est égal à $K - C$ ou inférieur à $K - C$. Dans le premier cas, il

quitte temporairement la zone de service pour rejoindre le groupe de clients bloqués (orbite); dans le second cas, il est tenu d'attendre dans l'espace d'attente.

Tout client en orbite va répéter à plusieurs reprises ses tentatives jusqu'au moment où il trouve un serveur inactif ou une position d'attente libre. Les temps de rappels sont répartis de façon exponentielle avec la fonction de répartition $T(x) = 1 - e^{-\theta x}$, $x \geq 0$ et le taux $\theta > 0$. Il s'agit d'une politique des rappels classique: les tentatives répétées sont faites individuellement par chaque client bloqué et le taux total des rappels est $j\theta$ lorsque la taille de l'orbite est $j \geq 0$. Les clients en attente (les clients dans l'espace d'attente) abandonneront le système si leur seuil de patience est dépassé. Nous supposons que les clients abandonnent après avoir attendu un temps exponentiellement distribué avec la fonction de répartition $W(x) = 1 - e^{-\gamma x}$, $x \geq 0$ et le taux $\gamma > 0$. Les temps de service suivent une loi exponentielle de fonction de répartition $B(x) = 1 - e^{-\mu x}$, $x \geq 0$ et de moyenne finie $\frac{1}{\mu}$. Enfin, nous admettons l'hypothèse de l'indépendance mutuelle entre toutes les variables aléatoires définies ci-dessus.

L'état du système à l'instant t peut être décrit au moyen du processus $\{C(t), N_0(t), t \geq 0\}$, où $C(t)$ est le nombre de clients dans les zones de service et d'attente, $N_0(t)$ représente le nombre de clients en orbite à l'instant t . Il s'agit d'une chaîne de Markov continue avec un espace d'états $S = \{0, 1, \dots, C, C + 1, \dots, K\} \times N$. Ses taux de transition $q_{ij}(k, l)$ sont donnés par:

1. Pour $0 \leq i \leq C - 1$

$$q_{ij}(k, l) = \begin{cases} \lambda & \text{si } (k, l) = (i + 1, j) \\ i\mu & \text{si } (k, l) = (i - 1, j) \\ j\theta & \text{si } (k, l) = (i + 1, j - 1) \quad ; \\ -(\lambda + i\mu + j\theta) & \text{si } (k, l) = (i, j) \\ 0 & \text{sinon} \end{cases}$$

2. Pour $C \leq i \leq K - 1$

$$q_{ij}(k, l) = \begin{cases} \lambda & \text{si } (k, l) = (i + 1, j) \\ C\mu + (i - C)\gamma & \text{si } (k, l) = (i - 1, j) \\ j\theta & \text{si } (k, l) = (i + 1, j - 1) \\ -(\lambda + C\mu + (i - C)\gamma + j\theta) & \text{si } (k, l) = (i, j) \\ 0 & \text{sinon} \end{cases} ;$$

3. Pour $i = K$

$$q_{ij}(k, l) = \begin{cases} \lambda & \text{si } (k, l) = (K, j + 1) \\ C\mu + (K - C)\gamma & \text{si } (k, l) = (K - 1, j) \\ -(\lambda + C\mu + (K - C)\gamma) & \text{si } (k, l) = (i, j) \\ 0 & \text{sinon} \end{cases} ;$$

Puisque le modèle à étudier a une structure complexe incluant les rappels et les abandons, son analyse mathématique présente un intérêt particulier ainsi qu'une difficulté analytique particulière.

3.2.1 Condition d'existence d'un régime stationnaire

La question suivante à étudier est l'ergodicité du processus de Markov $\{C(t), N_0(t), t \geq 0\}$.

3.2.1.1 Conditions suffisante

Théorème 3.2.1 *Soit $X(t)$ un processus de Markov avec l'espace d'état discret S et les taux de transition q_{sp} , où $s, p \in S$, $\sum_p q_{sp} = 0$. Supposons qu'il existe:*

- (a) *une fonction $\varphi(s)$, $s \in S$ (cette fonction est dite une fonction de Lyapunov).*
- (b) *un nombre positif ϵ tel que:*

- *les variables $y_s = \sum_{p \neq s} q_{sp}(\varphi(p) - \varphi(s)) < \infty$ pour tout $s \in S$;*
- *$y_s \leq -\epsilon$ pour tous les $s \in S$ sauf un nombre fini d'états.*

Alors le processus $X(t)$ est régulier et ergodique.

Preuve. Nous considérons la fonction de test suivante $\varphi(s) \equiv \varphi(i, j) = ai + j$, où a est un paramètre inconnu et $s \in S$. Les accroissements moyens $y_s \equiv y_{ij}$ sont données par:

$$y_{ij} = \begin{cases} \lambda a + i\mu(-a) + j\theta(a-1), & \text{si } 0 \leq i \leq C-1, \\ \lambda a + (C\mu + (i-C)\gamma)(-a) + j\theta(a-1), & \text{si } C \leq i \leq K-1, \\ \lambda + (C\mu + (K-C)\gamma)(-a), & \text{si } i = K, \end{cases} .$$

Puisque pour tout $i = 0, 1, \dots, K$ il existe

$$\lim_{j \rightarrow \infty} y_{ij} = L_i = \begin{cases} (a-1) \cdot \infty, & \text{si } 0 \leq i \leq C-1, \\ (a-1) \cdot \infty, & \text{si } C \leq i \leq K-1, \\ \lambda - a[C\mu + (K-C)\gamma], & \text{si } i = K. \end{cases} ;$$

les suppositions du théorème restent vraies si et seulement si toutes les variables L_i sont négatives, Ainsi:

$$\begin{aligned} a-1 &< 0; \\ a-1 &< 0; \\ \lambda - a[C\mu + (K-C)\gamma] &< 0. \end{aligned}$$

Ces conditions représentent un ensemble d'inégalités linéaires pour le paramètre inconnu a . Ces conditions peuvent être écrites sous la forme:

$$\frac{\lambda}{[C\mu + (K-C)\gamma]} < a < 1.$$

■

3.2.1.2 Condition nécessaire

Pour prouver que la condition ci-dessus est aussi une condition nécessaire, on applique le théorème 2.2.2.

3.2. Description du modèle

Considérons $\{\zeta_n\}$ une chaîne de Markov induite pour le processus de Markov défini ci-dessus $\{C(t), N_o(t), t \geq 0\}$. Ses probabilités de transition sont données par

1. Pour $0 \leq i \leq C - 1$

$$r_{ij}(k, l) = \begin{cases} \frac{\lambda}{\lambda+i\mu+j\theta} & \text{si } (k, l) = (i+1, j) \\ \frac{i\mu}{\lambda+i\mu+j\theta} & \text{si } (k, l) = (i-1, j) \\ \frac{j\theta}{\lambda+i\mu+j\theta} & \text{si } (k, l) = (i+1, j-1) \\ 0 & \text{sinon} \end{cases} ;$$

2. Pour $C \leq i \leq K - 1$

$$r_{ij}(k, l) = \begin{cases} \frac{\lambda}{\lambda+C\mu+(i-C)\gamma+j\theta} & \text{si } (k, l) = (i+1, j) \\ \frac{C\mu+(i-C)\gamma}{\lambda+C\mu+(i-C)\gamma+j\theta} & \text{si } (k, l) = (i-1, j) \\ \frac{j\theta}{\lambda+C\mu+(i-C)\gamma+j\theta} & \text{si } (k, l) = (i+1, j-1) \\ 0 & \text{sinon} \end{cases} ;$$

3. Pour $i = K$

$$r_{Kj}(k, l) = \begin{cases} \frac{\lambda}{\lambda+C\mu+(K-C)\gamma} & \text{si } (k, l) = (K, j+1) \\ \frac{C\mu+(K-C)\gamma}{\lambda+C\mu+(K-C)\gamma} & \text{si } (k, l) = (K-1, j) \\ 0 & \text{sinon} \end{cases} .$$

Supposons que $\lambda \geq C\mu + (K - C)\gamma$ et considérons la fonction de test $\varphi(i, j) = i + j$.

Alors,

- Si $r_{ij}(n, m) \neq 0$, alors $|i - n| \leq 1$, $|j - m| \leq 1$. Par conséquent:
 $|\varphi(i, j) - \varphi(n, m)| = |i - n + j - m| \leq |i - n| + |j - m| \leq 2$.
- Puisque la variable b peut prendre un nombre positif quelconque. Alors,
 - i. $A_b = \{(i, j)/i = 0, 1, \dots, K; j > \max(b - i, 0)\} \neq \emptyset$;

ii.

$$x_{ij} = \begin{cases} \frac{\lambda-i}{\lambda+i\mu+j\theta} & \text{si } 0 \leq i \leq C - 1 \\ \frac{\lambda-i}{\lambda+C\mu+(i-C)\gamma+j\theta} & \text{si } C \leq i \leq K - 1 \\ \frac{\lambda-K}{\lambda+C\mu+(K-C)\gamma} & \text{si } i = K \end{cases} .$$

Ainsi, la variable x_{ij} est toujours non négative et de plus $x_{ij} \geq 0$ pour $(i, j) \in A_b$. Par conséquent, le processus de Markov $\{C(t), N_o(t); t \geq 0\}$ est non-ergodique.

A présent, considérons le cas où $\lambda > C\mu + (K - C)\gamma$ et la fonction test est $\varphi(i, j) = ai + j$ avec $a \in (1, (\lambda/(C\mu + (K - C)\gamma)))$. Alors:

- Si $r_{ij}(n, m) \neq 0$ alors $|i - n| \leq 1, |j - m| \leq 1$.
Par conséquent, $|\varphi(i, j) - \varphi(n, m)| = |a(i - n) + j - m| \leq a|i - n| + |j - m| \leq a + 1$, c'est à dire $d = a + 1$.
- L'ensemble A_b est non vide quelque soit $b > 0$. Alors, on doit choisir b de telle sorte que $\inf_{(i,j) \in A_b} x_{ij} > 0$. Mais pour $i = K$, $x_{Kj} = \frac{\lambda - aK}{\lambda + C\mu + (K - C)\gamma} > 0$ (pour tous j); et pour $0 \leq i \leq K - 1$, $\lim_{j \rightarrow \infty} x_{ij} = a - 1 > 0$.

Par conséquent pour tout $i = 0, 1, \dots, K$, il existe N_i tel que $x_{ij} > \frac{a-1}{2}$ pour tout N_i . Ici, N_i est le nombre de i -clients en orbite. Si $\epsilon = \min \left\{ \frac{\lambda - aK}{\lambda + C\mu + (K - C)\gamma}, \frac{a-1}{2} \right\}$, alors $x_{ij} \geq \epsilon$ pour tout $(i, j) \in S$ sauf pour (i, j) tels que $i = 0, 1, \dots, K - 1$ et $0 \leq j \leq N_i - 1$. Supposons que b est assez grand pour que l'ensemble A_b ne contienne pas ces états. Alors $x_{ij} \geq \epsilon$, pour tout $(i, j) \in A_b$, ou $\inf_{(i,j) \in A_b} x_{ij} \geq \epsilon > 0$. En conséquence, la chaîne de Markov à temps continu $\{C(t), N_o(t), t \geq 0\}$ est non-ergodique.

Enfin, le processus de Markov considéré $\{C(t), N_o(t), t \geq 0\}$ est ergodique si et seulement si $\lambda < C\mu + (K - C)\gamma$. ■

3.3 Régime stationnaire

Supposons que la condition d'ergodicité $\lambda < C\mu + (K - C)\gamma$ est satisfait. Soit

$$P_{ij} = \lim_{t \rightarrow \infty} P(C(t) = i, N_o(t) = j), \quad (i, j) \in S,$$

la distribution conjointe stationnaire du nombre de clients dans les zones de service et d'attente et du nombre de clients en orbite, pour laquelle les équations de Kolmogorov

sont données par:

$$(\lambda + i\mu + j\theta)p_{ij} = \lambda p_{i-1,j} + (j+1)\theta p_{i-1,j+1} + (i+1)\mu p_{i+1,j},$$

$$0 \leq i \leq C-1 \text{ et } j \geq 0;$$

$$(\lambda + C\mu + (i-C)\gamma + j\theta)p_{ij} = \lambda p_{i-1,j} + (j+1)\theta p_{i-1,j+1}$$

$$+ (C\mu + (i+1-C)\gamma)p_{i+1,j},$$

$$C \leq i \leq K-1 \text{ et } j \geq 0;$$

$$(\lambda + C\mu + (K-C)\gamma)p_{K,j} = \lambda p_{K-1,j} + (j+1)\theta p_{K-1,j+1} + \lambda p_{K,j-1},$$

$$i = K \text{ et } j \geq 0;$$

$$\sum_{i=0}^K \sum_{j=0}^{\infty} p_{ij} = 1.$$

La distribution stationnaire p_{ij} contribue à l'obtention des mesures de performance, telles que:

- probabilité de blocage $P_K = \lim_{t \rightarrow \infty} P(C(t) = K)$;
- nombre moyen de clients en orbite $\bar{N}_o = \lim_{t \rightarrow \infty} E[N_o(t)]$;
- nombre moyen de clients dans les zones de service et d'attente $\bar{C} = \lim_{t \rightarrow \infty} E[C(t)]$.

Les caractéristiques essentielles de l'espace d'états du modèle considéré sont: sa dimension infinie (due à l'orbite illimitée) et la non-homogénéité le long de celui-ci (produite par la caractéristique des rappels). On ne peut donc pas obtenir la solution analytique exacte pour la distribution stationnaire P_{ij} . Dans ces circonstances, nous nous intéressons à deux approches approximatives: la troncature finie et l'extrapolation de valeur.

3.4 Troncature finie

La manière intuitive d'exécuter la procédure de troncature consiste à inclure dans la description du modèle la condition selon laquelle le nombre maximal de clients en orbite ne

dépasse pas une valeur M (choisie suffisamment grande). Cependant, la mise en oeuvre de ce schéma de troncature peut conduire à prendre en compte un grand nombre d'états avec des probabilités négligeables. Vu la méthodologie, qui est proposée dans [30] et décrite dans le paragraphe 2.2.2, cet inconvénient peut être atténué en explorant toutes les directions dans l'espace d'états considéré où les probabilités d'état diminuent, puis en déterminant des frontières de troncature. Nous avons donc la possibilité de fournir l'erreur relative donnée de l'estimation des mesures de performance.

Supposons que la taille de l'orbite de notre modèle soit bornée par une constante donnée M , choisie de manière appropriée. Dans le cadre de l'hypothèse ci-dessus, le nouveau processus $\{C^{(M)}(t), N_o^{(M)}(t), t \geq 0\}$ est un Markovien avec l'espace d'états $S^{(M)} = \{0, \dots, C, C+1, \dots, K-1, K\} \times \{0, 1, \dots, M\}$. Nous avons les taux de transition infinitésimaux suivants $q_{ij}^{(M)}(k, l)$:

1. Pour $0 \leq i \leq C-1, 0 \leq j \leq M-1$

$$q_{ij}^{(M)}(k, l) = \begin{cases} \lambda & \text{si } (k, l) = (i+1, j) \\ i\mu & \text{si } (k, l) = (i-1, j) \\ j\theta & \text{si } (k, l) = (i+1, j-1) \\ -(\lambda + i\mu + j\theta) & \text{si } (k, l) = (i, j) \\ 0 & \text{sinon} \end{cases} ;$$

2. Pour $0 \leq i \leq C-1, j = M$

$$q_{iM}^{(M)}(k, l) = \begin{cases} \lambda & \text{si } (k, l) = (i+1, M) \\ i\mu & \text{si } (k, l) = (i-1, M) \\ M\theta & \text{si } (k, l) = (i+1, M-1) \\ -(\lambda + i\mu + M\theta) & \text{si } (k, l) = (i, M) \\ 0 & \text{sinon} \end{cases} ;$$

3. Pour $C \leq i \leq K - 1, 0 \leq j \leq M - 1$

$$q_{ij}^{(M)}(k, l) = \begin{cases} \lambda & \text{si } (k, l) = (i + 1, j) \\ C\mu + (i - C)\gamma & \text{si } (k, l) = (i - 1, j) \\ j\theta & \text{si } (k, l) = (i + 1, j - 1) \text{ ;} \\ -(\lambda + C\mu + (i - C)\gamma + j\theta) & \text{si } (k, l) = (i, j) \\ 0 & \text{sinon} \end{cases}$$

4. Pour $C \leq i \leq K - 1, j = M$

$$q_{iM}^{(M)}(k, l) = \begin{cases} \lambda & \text{si } (k, l) = (i + 1, M) \\ C\mu + (i - C)\gamma & \text{si } (k, l) = (i - 1, M) \\ M\theta & \text{si } (k, l) = (i + 1, M - 1) \text{ ;} \\ -(\lambda + C\mu + (i - C)\gamma + M\theta) & \text{si } (k, l) = (i, M) \\ 0 & \text{sinon} \end{cases}$$

5. Pour $i = K, 0 \leq j \leq M - 1$

$$q_{Kj}^{(M)}(k, l) = \begin{cases} \lambda & \text{si } (k, l) = (K, j + 1) \\ C\mu + (K - C)\gamma & \text{si } (k, l) = (K - 1, j) \\ -(\lambda + C\mu + (K - C)\gamma) & \text{si } (k, l) = (K, j) \text{ ;} \\ 0 & \text{sinon} \end{cases}$$

6. Pour $i = K, j = M$

$$q_{KM}^{(M)}(k, l) = \begin{cases} C\mu + (K - C)\gamma & \text{si } (k, l) = (K - 1, M) \\ -(C\mu + (K - C)\gamma) & \text{si } (k, l) = (K, M) \\ 0 & \text{sinon} \end{cases} .$$

Comme l'espace d'états $S^{(M)}$ est fini, le processus $\{C^{(M)}(t), N_o^{(M)}(t), t \geq 0\}$ est toujours ergodique. Sa distribution en régime stationnaire

$P_{ij}^{(M)} = \lim_{t \rightarrow \infty} P(C^{(M)}(t) = i, N_o^{(M)}(t) = j)$ satisfait aux équations suivantes:

$$\begin{aligned} (\lambda + i\mu + j\theta)p_{ij}^{(M)} &= \lambda p_{i-1,j}^{(M)} + (i + 1)\mu p_{i+1,j}^{(M)} + (j + 1)\theta p_{i-1,j+1}^{(M)}, \\ 0 \leq i \leq C - 1, 0 \leq j \leq M - 1; \end{aligned} \quad (3.4.1)$$

$$(\lambda + i\mu + M\theta)p_{iM}^{(M)} = \lambda p_{i-1,M}^{(M)} + (i+1)\mu p_{i+1,M}^{(M)}, \quad (3.4.2)$$

$$0 \leq i \leq C-1, \quad j = M;$$

$$(\lambda + [C\mu + (i-C)\gamma] + j\theta)p_{i,j}^{(M)} = \lambda p_{i-1,j}^{(M)} + (j+1)\theta p_{i-1,j+1}^{(M)} \quad (3.4.3)$$

$$[C\mu + (i+1-C)\gamma]p_{i+1,j}^{(M)},$$

$$C \leq i \leq K-1, \quad 0 \leq j \leq M-1;$$

$$(\lambda + [C\mu + (i-C)\gamma] + j\theta)p_{i,j}^{(M)} = \lambda p_{i-1,j}^{(M)} + [C\mu + (i+1-C)\gamma]p_{i+1,j}^{(M)}, \quad (3.4.4)$$

$$C \leq i \leq K-1, \quad j = M;$$

$$(\lambda + [C\mu + (K-C)\gamma])p_{K,j}^{(M)} = \lambda p_{K-1,j}^{(M)} + \lambda p_{K,j-1}^{(M)} + (j+1)\theta p_{K-1,j+1}^{(M)}, \quad (3.4.5)$$

$$i = K, \quad 0 \leq j \leq M-1;$$

$$(C\mu + (K-C)\gamma)p_{K,M}^{(M)} = \lambda p_{K-1,M}^{(M)} + \lambda p_{K,M-1}^{(M)}, \quad (3.4.6)$$

$$i = K, \quad j = M$$

Cet ensemble d'équations peut être résolu par un algorithme récursif, décrit dans le paragraphe 2.2.3, et nous permet de trouver certaines variables $r_{ij}^{(M)} = \frac{P_{ij}^{(M)}}{P_{0M}^{(M)}}$, $0 \leq i \leq K$ et $0 \leq j \leq M$. Ensuite, les probabilités $P_{ij}^{(M)} = \frac{r_{ij}^{(M)}}{\sum_{i=0}^K \sum_{j=0}^M r_{ij}^{(M)}}$. En prenant en considération les spécificités de notre modèle, nous avons adapté cet algorithme pour calculer les mesures de performance introduites ci-dessus pour le système tronqué:

$$P_K^{(M)} = \sum_{j=0}^M r_{Kj}^{(M)} P_{0M}^{(M)};$$

$$\bar{N}_o^{(M)} = \sum_{i=0}^K \sum_{j=0}^M j r_{ij}^{(M)} P_{0M}^{(M)};$$

$$\bar{C}^{(M)} = \sum_{i=0}^K \sum_{j=0}^M i r_{ij}^{(M)} P_{0M}^{(M)}.$$

3.5 Extrapolation de Valeur

L'idée de la méthode Extrapolation de Valeur est de considérer le système dans le contexte des Processus de Markov Décisionnels et d'obtenir la valeur prévue à partir des équations de Howard écrites pour un espace d'état tronqué.

Considérons notre système. En utilisant les taux de transition donnés dans le paragraphe 3.2 et en suivant la méthodologie d'application de cette approche, décrite dans le paragraphe 2.3, nous pouvons écrire les équations de Howard suivantes:

$$\begin{aligned} r(i, j) - \bar{r} + \lambda(v(i+1, j) - v(i, j)) + i\mu(v(i-1, j) - v(i, j)) & \quad (3.5.1) \\ + j\theta(v(i+1, j-1) - v(i, j)) & = 0, \\ 0 \leq i \leq C-1; & \end{aligned}$$

$$\begin{aligned} r(i, j) - \bar{r} + \lambda(v(i+1, j) - v(i, j)) + j\theta(v(i+1, j-1) - v(i, j)) & \quad (3.5.2) \\ + (C\mu + (i-C)\gamma)(v(i-1, j) - v(i, j)) & = 0, \\ C \leq i \leq K-1; & \end{aligned}$$

$$\begin{aligned} r(K, j) - \bar{r} + (C\mu + (K-C)\gamma)(v(K-1, j) - v(K, j)) & \quad (3.5.3) \\ + \lambda(v(K, j+1) - v(K, j)) & = 0, \\ i = K. & \end{aligned}$$

Ici, \bar{r} représente soit la probabilité de blocage P_K , ou le nombre moyen de clients en orbite \bar{N}_o , ou le nombre moyen de clients dans les zones de service et d'attente \bar{C} .

Comme l'espace d'état S du processus $\{C(t), N_o(t), t \geq 0\}$ est infini, nous effectuerons une procédure de troncature pour obtenir $\hat{S} := \{s = (i, j) : 0 \leq i \leq K, 0 \leq j \leq M\}$. Le nombre d'équations de Howard correspond au nombre d'états, $|\hat{S}|$, et le nombre d'inconnues est défini par $|\hat{S}|$ valeurs d'état relatives plus revenu prévu \bar{r} . Cependant, comme seules les différences dans les valeurs relatives apparaissent dans les équations de Howard, on

peut fixer $v(0) = 0$. Nous avons donc un système linéaire d'équations avec le même nombre d'équations que d'inconnues. Maintenant nous pouvons fixer les taux de revenus afin d'obtenir les mesures de performance (voir Table 3.1):

Table 3.1: Définition de la fonction revenu.

Probabilité de blocage	P_K	$r(i, j) = 1$ pour $i = K$ et $0 \leq j \leq M$ $r(i, j) = 0$ sinon
Nombre moyen de clients en orbite	\bar{N}_o	$r(i, j) = j$, pour $0 \leq i \leq K$ et $0 \leq j \leq M$ $r(i, j) = 0$ sinon
Nombre moyen de serveurs occupés	\bar{C}	$r(i, j) = i$, pour $0 \leq i \leq K$ et $0 \leq j \leq M$ $r(i, j) = 0$ sinon

Le point suivant à étudier est la détermination des valeurs d'état relatives extrapolées, $v(s)$. Pour cela, nous devons choisir une fonction d'extrapolation $f(s)$ qui interpole quelques points $(s, v(s))$ pour $s \in \hat{S}$ de sorte qu'il s'approche également de $(s, v(s))$ pour $s \notin \hat{S}$. Comme c'était mentionné au paragraphe 2.3, nous avons deux types de procédures d'extrapolation selon le fait que nous utilisons toutes les paires $(s, v(s))$ de l'espace d'états ou un sous-ensemble de celles-ci, S_f . Le choix d'une fonction d'extrapolation et d'un sous-ensemble S_f est régi par le fait que le nombre de différentes paires $(s, v(s))$ dans le sous-ensemble en question doit être égal ou supérieur au nombre de coefficients dans le polynôme. De plus, le choix de S_f dépend aussi de la valeur d'état relative que nous voulons extrapoler. Ainsi, nous avons un problème d'interpolation. Le polynôme, sous forme de Lagrange, est $L(w) = \prod_{j=0}^{n-1} v(w_j)l_j(w)$, où

$$l_j(w) = \prod_{\substack{i=0 \\ j \neq i}}^{n-1} \frac{w-w_i}{w_j-w_i} = \frac{w-w_0}{w_j-w_0} \cdots \frac{w-w_{j-1}}{w_j-w_{j-1}} \frac{w-w_{j+1}}{w_j-w_{j+1}} \cdots \frac{w-w_{n-1}}{w_j-w_{n-1}}.$$

Pour le problème de troncature dans l'intérêt [31], nous aurons une équation de Howard [voir l'équation (3.6.3) pour $j = M + 1$] dans laquelle apparaît la valeur d'état $v(K, M + 1)$ de l'état $(K, M + 1)$ qui n'appartient pas à \hat{S} . Par conséquent, nous devons approximer

$v(K, M + 1)$ en utilisant certaines valeurs d'état relatives des états appartenant à \widehat{S} . Il est important de souligner que l'extrapolation de $v(K, M + 1)$ sera réalisée en utilisant les états de forme $s = (K, j)$ (avec des valeurs différentes de j). Avec ce choix, on peut définir la cartographie w comme $w = w((K, j)) = j$. De plus, nous utilisons un polynôme de degré $(n - 1)$ pour interpoler les n points dans $S_f := \{s_i = (K, M - i) | i = 0, \dots, n - 1\}$. Par conséquent, $w(S_f) = \{w_i = M - i | i = 0, \dots, n - 1\}$:

$$w_0=M \longrightarrow v(w_0) = v(K, M);$$

$$w_1=M - 1 \longrightarrow v(w_1) = v(K, M - 1);$$

$$w_{n-1}=M - (n - 1) \longrightarrow v(w_{n-1}) = v(K, M - (n - 1)).$$

De cette manière, l'expression générale de la valeur d'état relative extrapolée est

$$v^{(n)}(K, M + 1)=L^{(n)}(M + 1) = \sum_{j=0}^{n-1} v(K, M - j)l_j(M + 1).$$

Par exemple, dans le cas de l'extrapolation linéaire avec $n = 2$, nous utilisons $(M, v(K, M))$ et $(M - 1, v(K, M - 1))$. Alors

$$\begin{aligned} v^{(2)}(K, M + 1)=L^{(2)}(M + 1) &= v(K, M)l_0(M + 1) + v(K, M - 1)l_1(M + 1) \\ &= v(K, M)\frac{(M+1)-(M-1)}{(M-(M-1))} + v(K, M - 1)\frac{((M+1)-M)}{((M-1)-M)} \\ &= 2v(K, M) - v(K, M - 1). \end{aligned}$$

Suivant une procédure similaire, nous obtenons les relations suivantes pour $n = 3$ et $n = 4$, respectivement:

$$v^{(3)}(K, M + 1) = 3v(K, M) - 3v(K, M - 1) + v(K, M - 2);$$

$$v^{(4)}(K, M + 1) = 4v(K, M) - 6v(K, M - 1) + 4v(K, M - 2) - v(K, M - 3).$$

Enfin, pour le polynôme de degré $(n - 1)$, en utilisant la base de Lagrange pour réduire la complexité de la procédure, on obtient l'expression de forme fermée suivante pour valeur d'état relative extrapolée:

$$v^{(n)}(K, M + 1) = \sum_{k=0}^{n-1} (-1)^k C_n^{k+1} v(K, M - j),$$

où n est le nombre de coefficients pris pour le polynôme de Lagrange.

Pour le modèle étudié, il suffit de remplacer $v(K, M + 1)$ par sa valeur approximative dans l'équation de Howard qui correspond à l'état (K, M) . Par exemple, si on utilise l'extrapolation linéaire ($n = 2$), cette équation (voir (3.5.3)) devient:

$$r(K, M) - \bar{r} + v(K, M)(-\lambda - C\mu - (K - C)\gamma) + (C\mu + (K - C)\gamma)v(K - 1, M) + \lambda v(K, M + 1) = 0;$$

$$r(K, M) - \bar{r} + v(K, M)(\lambda - C\mu - (K - C)\gamma) + (C\mu + (K - C)\gamma)v(K - 1, M) - \lambda v(K, M - 1) = 0.$$

Comme $v(K, M + 1)$ n'apparaît plus dans les équations de Howard, nous avons un système linéaire de $(K + 1) \times (M + 1)$ équations avec le même nombre d'inconnues. Les équations de Howard (3.5.1)-(3.5.3) peuvent être exprimées sous une forme matricielle $xT = b$, où x est un vecteur ayant $(K + 1) \times (M + 1)$ inconnues et b contient les taux de revenus négatifs:

$$x = [\bar{r}, v(0, 1), \dots, v(0, M), v(1, 0), \dots, v(C, 0), \dots, v(C, M), \dots, v(K, 0), \dots, v(K, M)];$$

$$b = [-r(0, 0), -r(0, 1), \dots, -r(C, 0), -r(C, 1), \dots, -r(K, M)].$$

La matrice T est composée des coefficients et peut être construite en faisant tous les éléments de la première ligne de T' égaux à -1, c'est-à-dire tous les éléments de la première

rangée de toutes les sous-matrices qui sont en première ligne de la matrice T' sera remplacée par -1. Initialement, la matrice T' est donnée par

$$T' = \begin{bmatrix} A_1^0 & A_0^0 & O & \cdots & O & O & O \\ A_2^1 & A_1^1 & A_0^1 & \cdots & O & O & O \\ O & A_2^2 & \ddots & \ddots & O & O & O \\ \vdots & \vdots & \ddots & A_1^C & A_0^C & \vdots & \vdots \\ O & O & O & A_2^C & \ddots & \ddots & O \\ O & O & O & \cdots & \ddots & A_1^{K-1} & A_0^{K-1} \\ O & O & O & \cdots & O & A_2^K & A_1^K \end{bmatrix},$$

où chaque sous-matrice a la dimension $(M + 1) \times (M + 1)$ et O est la matrice nulle.

$$A_0^i = ((i + 1)\mu)I, \quad 0 \leq i \leq C - 1;$$

$$A_0^i = (C\mu + (i + 1 - C)\gamma)I, \quad C \leq i \leq K - 1;$$

$$A_2^i = \begin{bmatrix} \lambda & \theta & 0 & \cdots & 0 & 0 \\ 0 & \lambda & 2\theta & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda & M\theta \\ 0 & 0 & 0 & \cdots & 0 & \lambda \end{bmatrix},$$

$1 \leq i \leq K;$

$$A_1^i = \begin{bmatrix} \alpha & 0 & 0 & \cdots & 0 \\ 0 & \alpha - \theta & 0 & \cdots & 0 \\ 0 & 0 & \alpha - 2\theta & \ddots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \alpha - M\theta \end{bmatrix},$$

où $\alpha = -\lambda - i\mu$ pour $0 \leq i \leq C - 1$ et $\alpha = -\lambda - C\mu - (i - C)\gamma$ pour $C \leq i \leq K - 1$.

Lorsque $i = K$, en utilisant les extrapolations linéaire ($n = 2$) et quadratique ($n = 3$), nous obtenons respectivement,

$$A_1^K = \begin{bmatrix} \beta & 0 & \cdots & 0 & 0 \\ \lambda & \beta & \cdots & 0 & 0 \\ 0 & \lambda & \cdots & 0 & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & \cdots & \beta & -\lambda \\ 0 & 0 & \cdots & \lambda & \lambda - C\mu - (K - C)\gamma \end{bmatrix}, \quad A_1^K = \begin{bmatrix} \beta & 0 & \cdots & 0 & 0 \\ \lambda & \beta & \cdots & 0 & 0 \\ 0 & \lambda & \cdots & 0 & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & \lambda \\ 0 & 0 & \cdots & \beta & -3\lambda \\ 0 & 0 & \cdots & \lambda & 2\lambda - C\mu - (K - C)\gamma \end{bmatrix},$$

où $\beta = -\lambda - C\mu - (K - C)\gamma$. Ainsi, si l'extrapolation est faite avec $n \leq M + 1$ points, la matrice A_1^K sera définie comme

$$A_1^K = \begin{bmatrix} \beta & 0 & \cdots & 0 & \lambda c_M^{(n)} \\ \lambda & \beta & \cdots & 0 & \lambda c_{M-1}^{(n)} \\ 0 & \lambda & \cdots & 0 & \lambda c_{M-2}^{(n)} \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & \lambda c_2^{(n)} \\ 0 & 0 & \cdots & \beta & \lambda c_1^{(n)} \\ 0 & 0 & \cdots & \lambda & -\lambda - C\mu - (K - C)\gamma + \lambda c_0^{(n)} \end{bmatrix},$$

où

$$c_l^{(n)} = \begin{cases} (-1)^l C_n^{l+1}, & \text{si } l < n \\ 0, & \text{si } l \geq n \end{cases}.$$

Conclusion générale

Dans ce travail, nous avons analysé un système de files d'attente $M/M/C/K$ avec rappels, espace d'attente, et possibilité d'abandonner le système dans le cas où la durée d'attente dans l'espace d'attente dépasse un certain seuil. Dans un premier temps, nous avons effectué une étude sur les systèmes de files d'attente classiques et avec rappels. Puis nous avons passé en revue certains résultats consacrés aux systèmes de files d'attente avec rappels ainsi que à deux approches utilisées par les chercheurs dans leurs investigations. Dans la première approche permet d'obtenir une solution numérique approchée à la distribution stationnaire de l'état du système, puis, à l'aide des outils fondamentaux de la théorie des probabilités, calculer les mesures de performance, tandis que la seconde approche utilise les moyens fournis par la théorie des Processus de Markov Décisionnels et permet d'exprimer et par la suite de déterminer les indices de performance en question en termes d'une certaine métrique, appelée valeur d'état relative.

Notre apport est théorique. Dans un premier temps, nous avons décrit le modèle général d'un système de files d'attente classiques et avec rappels ainsi que d'un modèle markovien à unique serveur et à multiserveurs. Nous avons également présenté deux approches les plus intéressantes et les plus performantes pouvant être appliquée lors de l'analyse des modèles avec rappels et multiserveurs.

Bibliographie

- [1] A. Aissani. A Survey on Retrial Queueing Models. *Actes des Journées Statistiques Appliquées*. U.S.T.H.B., Alger, 1-11,1994.
- [2] A. Anisimov and M. Kurtulush. Some Markovian queueing retrial system under light-traffic conditions, *Cybernetics and Systems Analysis*, 37: 876- 887.2001.
- [3] J. R. Artalejo. Accessible bibliography on retrial queues : Progress in 2000-2009. *Mathematical and computer modelling* 51 : 1071-1081,2010.
- [4] J. R. Artalejo and A. Gmez-Corral. Advances in Retrial Queues. *European Journal of Operation Research*, 2008.
- [5] J.R. Artalejo. Retrial queues with a finite number of sources. *Journal of the Korean Mathematical Society*, 35(3) :503-525,1998.
- [6] J.R. Artalejo. Retrial Queueing systems, *Mathematical and Computer Modelling* ,30 : 3-4, 1-228,1999.
- [7] J. R. Artalejo. Algorithmic Methods in Retrial Queues. *Annals of Operation Research*, 141 : 1-301,2006.
- [8] J.R. Artalejo and A.Gomez-Corral.Retrial Queueing Systems :A *Computational Approach*. Springer, 2008.
- [9] J.W. Cohen. Basic problems of the telephone traffic theory and the influence of repeated calls. *Philips Telecom*, 18(2) : 49-100,1957.
- [10] V. Fabrice. Les files d'attentes : Modélisation et évaluation de perforamances de réseaux, Technical report, *Université Lyon, France*, 2003.

- [11] G.I. Falin. A Single-line System with Secondary Orders. *Engineering Cybernetics Review*, 17(2) : 76-83,1979.
- [12] G.I. Falin. Single-line Repeated Orders Queueing Systems. *Mathematische Operationsforschung und Statistik, Optimization*, 5 : 649-667,1986.
- [13] G.I Falin and J.G.C. Templeton. Retrial queues. *New Jersey : Chapman and Hall*, 1997.
- [14] L. Ferhat. Approximation dans le système de files d'attente avec rappels, à serveur non fiable et client impatients, Thèse de doctorat L.M.D,Tizi-ouzou, 2019.
- [15] J.M. Gimenez-Guzman, M.J. Domenech-Belloch, V. Pla, V. Casares-Giner and J. Martinez-Bauset. Value extrapolation technique to solve retrial queues : a comparative perspective. *ETRI journal*, 30 :492-494, 2008.
- [16] J.M. Gimenez-Guzman, M.J. Domenech-Belloch, V. Pla, J. Martinez-Bauset and V. Casares-Giner. Efficient method to approximately solve retrial systems with impatience. *Journal of Applied Mathematics*, 2012 : 1-18, 2012.
- [17] P. Goatin. Analyse Numérique. Cours de 1^{ère} année- Var ISITV, *du université du Sud Toulon*.
- [18] S.K. Godunov and V.S. Ryabenkii. Difference Schemes. *Northe Holland*, 47-50, 1987.
- [19] A. Gomez-Corral and M.F. Ramalhoto. On the waiting time distribution and the busy period of a retrial queue with constant retrial rate. *Stochastic Modelling and Applications*, 3 : 37-47,2000.
- [20] O. Hashida and K. Kawashima. Buffer Behavior with Repeated Calls. *Electronics and Communication in Japan*, 62-B. 27,1979.
- [21] R.A. Howard. Dynamic Programming and Markov Processus, The Technology Press of MIT, Cambridge, Mass, *USA*, 1960.
- [22] F. Jędrzejewski, Introduction aux méthodes numériques. *Springer*, 2005.
- [23] E.V. Koba and S.V. Pustovaya. Call center as retrial queueing system. *Journal of Automation and Information Sciences*, 39 : 37-47,2007.

- [24] L. Kosten. On the influence of repeated calls in the theory of probabilities of blocking. *De Ingenieur*, 59 : 1-125, 1947.
- [25] V.G. Kulkarni and H.M. Liang. Retrial Queues Revisited. *Frontiers in Queueing (J.H. Dshalov, ed.) CRC Press Boca Raton*, pp 19-34, 1997.
- [26] J. Leino and J. Virtamo. An approximate method for calculating performance measures of Markov processes. *Proceedings of VALUETOOLS*, 2006.
- [27] M.L. Puterman. Markov Decision Processes : Discrete Stochastic Dynamic Programming. *Wiley series in Probability and Statistics, Wiley*, 2008.
- [28] J.G.C. Templeton. Retrial Queues. *Top*, vol. 7, p : 351-353, 1999.
- [29] R.I. Wilkinson. Theories for toll traffic engineering in the USA. *Bell systems technical journal*, 35 : 421-507, 1956.
- [30] N. Zidani et N. Djellab. Calcul numérique des caractéristiques stationnaires du modèle tronqué, Journées Nationales sur les Mathématiques Appliquées "JNMA'14", Université 20 août 1955 Skikda, 26-27 Novembre 2014.
- [31] N. Zidani et N. Djellab. L'approximation de la distribution stationnaires de l'état des systems des files d'attente avec rappels et multiserveurs. Journées Jeunes chercheurs, Université Badji Mokhtar- Annaba, 30 septembre - 1 Octobre 2014.