

République Algérienne Démocratique et Populaire
Ministère de l'enseignement supérieur et de la recherche scientifique
Université de Saïda - Dr Moulay Tahar.



Faculté des Mathématiques,
d'Informatiques et
Télécommunications.
Département de Mathématiques.



Mémoire présenté en vue de l'obtention du diplôme de

Master Académique

Filière : **MATHEMATIQUES**

Spécialité : **Analyse Stochastique, Statistique des Processus et
Applications**

par

KADDOUR CHERIF Mokhtar¹

Sous la direction de

Dr. YAHIAOUI Lahcene

Thème :

**Analyse de performance des systèmes d'attente
à temps de service général**

Soutenu le 14/06/2026 devant le jury composé de

Dr. HAMLAT Z.	Université de Saïda Dr. Moulay Tahar	Président
Dr. YAHIAOUI L.	Université de Saïda Dr. Moulay Tahar	Encadrent
Dr. KADI M.	Université de Saïda Dr. Moulay Tahar	Examinateur

Année univ.: 2025/2026

1. e-mail : kaddourcherifmokhtar@gmail.com

Dédicace

Je dédie ce travail :

- † *À mes chers parents, source de force, de tendresse et de dignité.*
- † *À ma grande famille et ma petite famille, pour leur amour et leur soutien.*
- † *À tous mes enseignants, et particulièrement ceux qui m'ont formé et dirigé.*
- † *À mes amis et camarades, pour leur amitié sincère.*
- † *À tous ceux qui ont contribué, directement ou indirectement, à ma réussite.*

Merci du fond du cœur.

Remerciements

« *Et le dernier des devoirs de celui qui a reçu du bien,
est d'en témoigner sa reconnaissance.* »

— *Proverbe arabe*

Je tiens tout d'abord à exprimer ma profonde gratitude et mon infini respect à **Dieu le Tout-Puissant**, qui m'a donné la force, le courage et la patience pour mener à bien ce travail.

Encadrement Je remercie chaleureusement mon encadrant, **Monsieur YAHIAOUI L.**, pour sa disponibilité, ses précieux conseils, sa patience et sa rigueur scientifique. Il a su m'orienter avec bienveillance tout au long de ce projet. Qu'il trouve ici l'expression de ma sincère reconnaissance.

Corps enseignant Je tiens également à remercier l'ensemble des enseignants qui ont contribué à ma formation, depuis mes premières années d'études jusqu'à aujourd'hui. Leurs cours, leur passion et leur dévouement m'ont inspiré et préparé aux défis de la recherche.

Jury de soutenance Mes remerciements s'adressent également aux membres du jury qui ont accepté de lire et d'évaluer ce travail. Leurs remarques et suggestions seront d'une grande richesse pour mon perfectionnement.

Famille Un grand merci à ma famille, en particulier à **mon cher père**, à **ma mère bien-aimée**, à mes frères et sœurs, ainsi qu'à toute ma famille grande et petite. Leur soutien moral, leurs encouragements constants et leur amour inconditionnel ont été mon moteur tout au long de ce parcours. Aucun mot ne saurait exprimer ma gratitude infinie.

Amis et camarades Je n'oublie pas mes amis et camarades de promotion, en particulier **[Hadjadj Houari]**, pour leur entraide, leur bonne humeur et les moments inoubliables que nous avons partagés. Leur amitié a rendu ce chemin plus doux et moins solitaire.

Remerciements généraux Enfin, je remercie toutes les personnes qui ont contribué de près ou de loin à la réalisation de ce mémoire, qu'elles trouvent ici l'assurance de ma profonde reconnaissance.

Merci à tous du fond du cœur.

Table des matières

Dédicace	2
Remerciements	3
Introduction générale	6
1 Processus stochastiques pour les files d'attente	8
1.1 Processus de comptage	8
1.2 Processus de renouvellement	9
1.3 Processus de Poisson	10
1.4 Processus de Markov	12
1.4.1 Chaînes de Markov à temps discret (CMTD)	13
1.4.2 Classification des états	14
1.4.3 Comportement stationnaire et asymptotique	15
1.4.4 Processus de Markov à temps continu et générateur infinitésimal	16
1.5 Processus de naissance et de mort	18
1.6 Définition et composants d'un système de file d'attente	19
1.7 Notation de Kendall pour les files d'attente	20
1.8 Application aux modèles de files d'attente	20
1.8.1 Modèle M/M/1	21
1.8.2 Modèle M/M/1/K (capacité finie)	22
1.8.3 Modèle M/M/c (serveurs multiples)	22
1.8.4 Modèle M/M/c/c (système avec pertes)	23
1.8.5 Modèle M/M/∞ (nombre infini de serveurs)	24
1.9 Loi de Little	24
1.10 Résumé des modèles et leurs caractéristiques	26
2 Analyse exacte du système M/G/1 : chaînes de Markov incluses	27
2.1 Limites des processus markoviens à temps continu	27
2.2 Chaînes de Markov incluses (Imbedded Markov Chains)	28
2.2.1 Points de régénération	28
2.2.2 Relation de récurrence fondamentale	29
2.2.3 Matrice de transition	29
2.2.4 Distribution stationnaire de la chaîne incluse	30
2.3 Fonction génératrice et transformée	30
2.3.1 Définition des fonctions génératrices	30
2.3.2 Dérivation de la formule de Pollaczek-Khinchine	31
2.3.3 La formule de Pollaczek-Khinchine pour le nombre moyen de clients	31
2.3.4 La formule de Pollaczek-Khinchine pour le temps d'attente moyen	32

2.4	Analyse de la période d'occupation (Busy Period)	33
2.4.1	Définition et propriétés fondamentales	33
2.4.2	Transformée de Laplace de la période d'occupation	33
2.4.3	Moments de la période d'occupation	34
2.5	Comparaison mathématique selon le type de service	35
2.5.1	Cas particuliers	35
2.5.2	Comparaison des formules de Pollaczek-Khinchine	35
2.5.3	Impact de la variabilité sur les performances	35
2.5.4	Tableau récapitulatif	36
2.6	Résumé du chapitre	36
3	File d'attente M/G/1 non fiable avec temps de rappel général, vacances actives et temps de configuration	38
3.1	Description du modèle	39
3.2	Analyse à l'état stationnaire	40
3.2.1	Description des états du système	40
3.2.2	Condition de stabilité et d'ergodicité	41
3.2.3	Équations régissant le système	41
3.3	Résultats en régime permanent	46
3.4	Mesures de performance du système	47
3.4.1	Probabilités des états du système	47
3.4.2	Taille moyenne du système et de l'orbite	47
3.5	Mesures de fiabilité	48
3.6	Périodes d'occupation et de cycle	49
3.6.1	Application pratique : Centre de tri de déchets recyclables	49
3.7	Analyse de sensibilité et exemples numériques	52
3.8	Configuration des Paramètres et Discussion des Courbes	53
3.8.1	Choix des paramètres numériques	53
3.8.2	Impact du taux de rappel (ξ) sur la taille de l'orbite	53
3.8.3	Effet du taux de service régulier (μ_n) sur la taille de l'orbite	53
3.8.4	Variation de la probabilité de vacances en fonction de son taux (ω)	54
3.8.5	Sensibilité de la probabilité de préparation (<i>Setup</i>) par rapport au taux de service	55
3.8.6	Impact du taux de réparation (η) sur la probabilité de réparation	55
3.8.7	Évolution de la probabilité d'oisiveté (P_{00}) en fonction du taux de service	56
3.9	Conclusion	57
	Conclusion générale	58

Introduction générale

Les systèmes de files d'attente constituent un outil fondamental de modélisation probabiliste pour l'analyse des performances de nombreux systèmes réels, tels que les réseaux de télécommunications, les systèmes informatiques, les centres d'appels, les chaînes de production et les infrastructures logistiques [7, 22, 20]. La théorie des files d'attente permet d'étudier le comportement de ces systèmes soumis à des arrivées aléatoires de clients et à des temps de service variables, dans le but d'optimiser des mesures de performance telles que le temps d'attente moyen, la longueur de la file, le taux d'occupation du serveur ou encore la disponibilité du service.

Au cours des dernières décennies, les modèles classiques de files d'attente ont été enrichis afin de mieux représenter les phénomènes rencontrés dans les systèmes modernes. Parmi ces extensions figurent notamment les files avec rappels (*retrial queues*), les vacances du serveur, les vacances actives (*working vacations*), les interruptions de vacances, les temps de configuration (*setup time*) ainsi que la non-fiabilité du serveur. Les files de rappels ont fait l'objet de nombreux travaux fondamentaux, notamment ceux de Falin et Templeton [7], Gomez-Corral [8], ainsi que les contributions plus récentes de Arrar et al. [2, 3, 4]. Ces modèles sont particulièrement adaptés aux systèmes dans lesquels les clients qui trouvent le serveur indisponible quittent temporairement le système avant de réessayer après un délai aléatoire.

Par ailleurs, le concept de vacances du serveur a été introduit afin de modéliser les périodes d'inactivité ou de maintenance des systèmes de service [12, 14, 22]. Une généralisation importante de ce mécanisme est celle des vacances actives, introduite dans les travaux de Servi et Finn [21], où le serveur continue de fournir un service à un rythme réduit pendant les périodes de vacances. Ce concept a suscité un intérêt croissant dans la littérature scientifique, donnant lieu à plusieurs études analytiques et applications [1, 5, 6, 15, 23, 24]. De plus, plusieurs chercheurs ont étudié l'effet combiné des vacances actives avec d'autres phénomènes tels que les interruptions de vacances, l'impatience des clients, les pannes du serveur et les temps de configuration [9, 10, 16, 17, 18, 19, 13].

Ce mémoire s'inscrit dans cette thématique et propose une analyse approfondie d'un modèle avancé de file d'attente de type M/G/1 intégrant simultanément un serveur non fiable, un temps de rappel général, des vacances actives et un temps de configuration. Ce modèle permet de représenter des situations réalistes où le serveur peut tomber en panne au cours du service, où les clients bloqués réessaient après un temps aléatoire général, où le serveur reste actif à un taux de service réduit durant les périodes de vacances, et où un temps de préparation est nécessaire avant la reprise normale du service. Un tel modèle présente un intérêt particulier dans plusieurs domaines industriels et technologiques, notamment les systèmes informatiques, les réseaux de communication et les systèmes automatisés de production.

L'objectif principal de ce travail est de développer une méthodologie rigoureuse pour l'analyse exacte de ce système, en déterminant ses mesures de performance à l'état station-

naire, sa condition de stabilité ainsi que certains indicateurs de fiabilité. Une application concrète au domaine du tri des déchets recyclables est également présentée afin d'illustrer l'intérêt pratique du modèle étudié. Ce travail s'appuie également sur des contributions récentes consacrées aux systèmes $M/G/1$ avec rappels, vacances actives et temps de configuration, notamment l'étude proposée dans [11].

Le mémoire est structuré en trois chapitres. Le premier chapitre présente les rappels mathématiques nécessaires à la compréhension de la suite : processus de comptage, processus de renouvellement, processus de Poisson, processus de Markov et processus de naissance et de mort. Le deuxième chapitre est consacré à l'analyse exacte du système $M/G/1$ par la méthode des chaînes de Markov incluses, menant à la célèbre formule de Pollaczek-Khinchine et à l'étude de la période d'occupation. Enfin, le troisième chapitre développe le modèle original proposé, incluant les vacances actives, le temps de configuration et la non-fiabilité du serveur, avec une analyse détaillée des équations de Kolmogorov, la dérivation des fonctions génératrices et le calcul des mesures de performance. Une étude de sensibilité et des exemples numériques viennent compléter cette analyse.

Chapitre 1

Processus stochastiques pour les files d'attente

L'étude des systèmes de files d'attente repose sur la modélisation probabiliste de l'évolution du nombre de clients au cours du temps. Cette évolution est décrite par un processus stochastique. Dans cette section, nous présentons les principaux types de processus utilisés dans ce mémoire, en commençant par les processus de comptage et de renouvellement, puis en introduisant le processus de Poisson avec ses propriétés fondamentales, pour aboutir enfin aux processus de Markov et plus particulièrement aux processus de naissance et de mort, qui constituent la base des modèles classiques comme M/M/1, M/M/c, etc.

1.1 Processus de comptage

Définition 1.1.1 (Processus de comptage). *Un processus de comptage est un processus stochastique à temps continu $\{N(t), t \geq 0\}$ où $N(t)$ représente le nombre d'événements survenus jusqu'à l'instant t . Il satisfait les propriétés suivantes [18] :*

1. $N(0) = 0$.
2. $N(t)$ est à valeurs entières non négatives.
3. $N(t)$ est croissant (au sens large) : si $s < t$, alors $N(s) \leq N(t)$.
4. Pour $s < t$, la différence $N(t) - N(s)$ est le nombre d'événements dans l'intervalle $(s, t]$.

Définition 1.1.2 (Temps d'arrivée et temps inter-arrivées). *Soit $\{N(t), t \geq 0\}$ un processus de comptage. On définit [20] :*

- Les temps d'arrivée (ou instants d'occurrence) S_1, S_2, \dots où S_n est l'instant du n -ième événement.
- Les temps inter-arrivées T_1, T_2, \dots où $T_1 = S_1$ et $T_n = S_n - S_{n-1}$ pour $n \geq 2$.

On a la relation :

$$S_n = \sum_{i=1}^n T_i$$

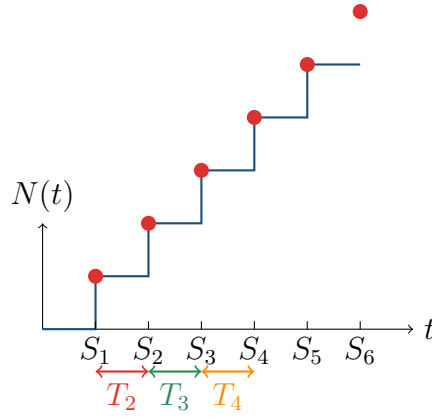


Figure : Processus de comptage : arrivées et temps inter-arrivées

1.2 Processus de renouvellement

Définition 1.2.1 (Processus de renouvellement). *Un processus de renouvellement est un processus de comptage $\{N(t), t \geq 0\}$ pour lequel les temps inter-arrivées T_1, T_2, \dots sont indépendants et identiquement distribués (i.i.d.) selon une loi positive [5, 2].*

Proposition 1.2.1 (Propriété de renouvellement). *Dans un processus de renouvellement, le processus "recommence à nouveau" à chaque événement. Plus précisément, si on pose S_n comme le n -ième instant de renouvellement, alors le processus $\{N(t + S_n) - N(S_n), t \geq 0\}$ a la même distribution que le processus original $\{N(t), t \geq 0\}$ et est indépendant du passé avant S_n [4].*

Théorème 1.2.1 (Théorème élémentaire du renouvellement). *Pour un processus de renouvellement avec des temps inter-arrivées de moyenne $\mathbb{E}[T] = 1/\lambda$, on a [7] :*

$$\lim_{t \rightarrow \infty} \frac{\mathbb{E}[N(t)]}{t} = \lambda$$

et presque sûrement :

$$\lim_{t \rightarrow \infty} \frac{N(t)}{t} = \lambda.$$

Démonstration. Soit S_n le temps du n -ième renouvellement. On a $S_n = \sum_{i=1}^n T_i$. Par la loi forte des grands nombres, $\frac{S_n}{n} \rightarrow \mathbb{E}[T] = 1/\lambda$ presque sûrement. De plus, $N(t)$ est l'inverse généralisé de S_n au sens où $S_{N(t)} \leq t < S_{N(t)+1}$. En divisant par $N(t)$, on obtient :

$$\frac{S_{N(t)}}{N(t)} \leq \frac{t}{N(t)} < \frac{S_{N(t)+1}}{N(t)}.$$

En passant à la limite $t \rightarrow \infty$, on a $N(t) \rightarrow \infty$ presque sûrement, et par la loi des grands nombres, $\frac{S_{N(t)}}{N(t)} \rightarrow \mathbb{E}[T]$ et $\frac{S_{N(t)+1}}{N(t)} \rightarrow \mathbb{E}[T]$. Par conséquent, $\frac{t}{N(t)} \rightarrow \mathbb{E}[T]$ presque sûrement, d'où $\frac{N(t)}{t} \rightarrow \frac{1}{\mathbb{E}[T]} = \lambda$ presque sûrement.

Pour l'espérance, on utilise le théorème de convergence dominée ou des arguments de renouvellement plus fins (théorème de Wald) pour établir que $\mathbb{E}[N(t)]/t \rightarrow \lambda$. \square

1.3 Processus de Poisson

Le processus de Poisson est le plus important des processus de renouvellement. Il constitue la base de nombreux modèles en files d'attente.

Définition 1.3.1 (Processus de Poisson). *Un processus de Poisson de taux $\lambda > 0$ est un processus de comptage $\{N(t), t \geq 0\}$ qui satisfait les conditions suivantes [10, 15] :*

1. $N(0) = 0$.
2. *Le processus a des accroissements indépendants : pour tous $0 \leq t_1 < t_2 < \dots < t_n$, les variables aléatoires $N(t_2) - N(t_1)$, $N(t_3) - N(t_2)$, ..., $N(t_n) - N(t_{n-1})$ sont indépendantes.*
3. *Le processus a des accroissements stationnaires : la distribution de $N(t+s) - N(s)$ ne dépend que de t , pas de s .*
4. *Pour tout intervalle de temps de longueur t , ($t > 0$), le nombre d'événements suit une loi de Poisson de moyenne λt :*

$$\mathbb{P}(N(t) = k) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}, \quad k = 0, 1, 2, \dots$$

Trajectoire typique d'un processus de Poisson ($\lambda = 1.5$)

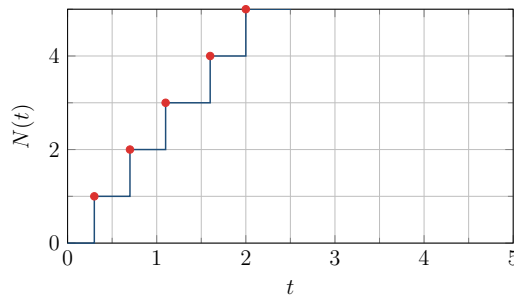


Figure : Trajectoire typique d'un processus de Poisson

Proposition 1.3.1 (Caractérisation équivalente). *Un processus de Poisson peut également être caractérisé par les propriétés suivantes [4] :*

- *Les temps inter-arrivées sont i.i.d. selon une loi exponentielle de paramètre λ :*

$$\mathbb{P}(T_i > t) = e^{-\lambda t}, \quad t \geq 0.$$

- *Pour t petit, la probabilité d'une arrivée est approximativement λdt :*

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

- *La probabilité de plus d'une arrivée dans un intervalle infinitésimal est négligeable :*

$$\mathbb{P}(N(t+dt) - N(t) \geq 2) = o(dt).$$

Démonstration. Montrons l'équivalence entre ces caractérisations.

(\Rightarrow) Supposons que $\{N(t)\}$ est un processus de Poisson selon la définition 4. Alors pour t petit, $\mathbb{P}(N(t) = 1) = \lambda t e^{-\lambda t} = \lambda t + o(t)$ et $\mathbb{P}(N(t) \geq 2) = 1 - e^{-\lambda t} - \lambda t e^{-\lambda t} = o(t)$. De plus, les temps inter-arrivées sont exponentiels car :

$$\mathbb{P}(T_1 > t) = \mathbb{P}(N(t) = 0) = e^{-\lambda t}.$$

Par la propriété d'indépendance des accroissements, les T_i sont i.i.d. exponentielles.

(\Leftarrow) Réciproquement, si les temps inter-arrivées sont i.i.d. exponentielles de taux λ , alors le processus de comptage associé a des accroissements indépendants et stationnaires, et $N(t) \sim \text{Poisson}(\lambda t)$. En effet, l'événement $\{N(t) = k\}$ équivaut à $\{S_k \leq t < S_{k+1}\}$ où $S_k = \sum_{i=1}^k T_i$. La somme de k exponentielles i.i.d. suit une loi Gamma(k, λ), donc :

$$\mathbb{P}(N(t) = k) = \mathbb{P}(S_k \leq t) - \mathbb{P}(S_{k+1} \leq t) = e^{-\lambda t} \frac{(\lambda t)^k}{k!}.$$

□

Proposition 1.3.2 (Propriété d'absence de mémoire). *La distribution exponentielle des temps inter-arrivées confère au processus de Poisson la propriété d'absence de mémoire (ou propriété de Markov) [18] :*

$$\mathbb{P}(T > t + s \mid T > s) = \mathbb{P}(T > t), \quad \forall s, t \geq 0.$$

Cette propriété est fondamentale pour l'analyse des systèmes markoviens.

Démonstration. Pour une variable exponentielle $T \sim \mathcal{E}(\lambda)$:

$$\mathbb{P}(T > t + s \mid T > s) = \frac{\mathbb{P}(T > t + s)}{\mathbb{P}(T > s)} = \frac{e^{-\lambda(t+s)}}{e^{-\lambda s}} = e^{-\lambda t} = \mathbb{P}(T > t).$$

Réciproquement, toute variable positive satisfaisant cette propriété est nécessairement exponentielle. □

Théorème 1.3.1 (Somme de processus de Poisson indépendants). *Soient $\{N_1(t), t \geq 0\}$ et $\{N_2(t), t \geq 0\}$ deux processus de Poisson indépendants de taux λ_1 et λ_2 . Alors leur somme $\{N(t) = N_1(t) + N_2(t), t \geq 0\}$ est un processus de Poisson de taux $\lambda = \lambda_1 + \lambda_2$ [10].*

Démonstration. Vérifions les propriétés caractéristiques. Les accroissements de N sont indépendants car ceux de N_1 et N_2 le sont. La stationnarité découle de celle de N_1 et N_2 . Enfin, pour tout t :

$$\mathbb{P}(N(t) = k) = \sum_{i=0}^k \mathbb{P}(N_1(t) = i, N_2(t) = k-i) = \sum_{i=0}^k e^{-\lambda_1 t} \frac{(\lambda_1 t)^i}{i!} e^{-\lambda_2 t} \frac{(\lambda_2 t)^{k-i}}{(k-i)!} = e^{-(\lambda_1 + \lambda_2)t} \frac{(\lambda_1 + \lambda_2)^k t^k}{k!}$$

C'est bien une loi de Poisson de paramètre $(\lambda_1 + \lambda_2)t$. □

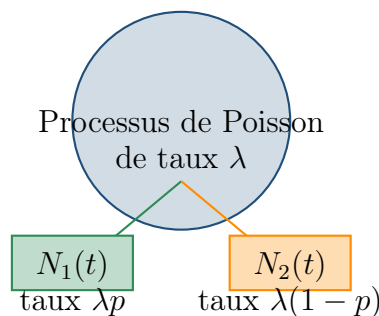


Figure : Décomposition d'un processus de Poisson

Théorème 1.3.2 (Décomposition (échantillonnage) d'un processus de Poisson). *Soit $\{N(t), t \geq 0\}$ un processus de Poisson de taux λ . Si chaque événement est, indépendamment des autres, classé comme étant de type 1 avec probabilité p et de type 2 avec probabilité $1 - p$, alors les processus $\{N_1(t), t \geq 0\}$ et $\{N_2(t), t \geq 0\}$ des événements de type 1 et de type 2 sont des processus de Poisson indépendants de taux respectifs λp et $\lambda(1 - p)$ [18].*

Démonstration. Conditionnellement à $N(t) = n$, le nombre d'événements de type 1 suit une loi binomiale $B(n, p)$. Donc :

$$\mathbb{P}(N_1(t) = k, N_2(t) = \ell) = \mathbb{P}(N_1(t) = k, N(t) = k + \ell) = \mathbb{P}(N(t) = k + \ell) \binom{k + \ell}{k} p^k (1 - p)^\ell.$$

En substituant $\mathbb{P}(N(t) = k + \ell) = e^{-\lambda t} \frac{(\lambda t)^{k + \ell}}{(k + \ell)!}$, on obtient :

$$\mathbb{P}(N_1(t) = k, N_2(t) = \ell) = e^{-\lambda p t} \frac{(\lambda p t)^k}{k!} \cdot e^{-\lambda(1 - p)t} \frac{(\lambda(1 - p)t)^\ell}{\ell!}.$$

Ce qui montre que $N_1(t)$ et $N_2(t)$ sont indépendants et suivent des lois de Poisson de paramètres respectifs $\lambda p t$ et $\lambda(1 - p)t$. Les propriétés d'accroissements indépendants et stationnaires se transmettent également. \square

1.4 Processus de Markov

Les processus de Markov occupent une place centrale en théorie des files d'attente en raison de leur propriété fondamentale d'absence de mémoire. Nous présentons ici une formalisation mathématique rigoureuse, accompagnée de démonstrations clés.

Définition 1.4.1 (Propriété de Markov). *Un processus stochastique $\{X(t), t \geq 0\}$ à valeurs dans un espace d'états E possède la propriété de Markov si, pour tout $t \geq 0$ et tout $s \geq 0$, la distribution future du processus ne dépend que de son état présent, et non de son passé. Formellement [13, 14] :*

$$\mathbb{P}(X(t + s) \leq x \mid X(u), 0 \leq u \leq t) = \mathbb{P}(X(t + s) \leq x \mid X(t))$$

pour tous $t, s \geq 0$ et tout $x \in E$.

Définition 1.4.2 (Processus de Markov). *Un processus stochastique qui satisfait la propriété de Markov est appelé processus de Markov. On distingue [4] :*

- Les chaînes de Markov à temps discret (CMTD) : X_n évolue par sauts discrets.
- Les processus de Markov à temps continu (PMTC) : $X(t)$ évolue continûment.

Remarque 1.4.1 (Lien avec la loi exponentielle). *En temps continu, un processus de Markov à espace d'états discret a des temps de séjour dans chaque état qui suivent une loi exponentielle [11]. En effet, la propriété de Markov implique que la distribution du temps restant dans un état ne dépend pas du temps déjà passé dans cet état, ce qui est une caractérisation de la loi exponentielle.*

Démonstration.[Absence de mémoire et loi exponentielle] Soit T le temps de séjour dans un état donné. La propriété de Markov implique :

$$\mathbb{P}(T > t + s \mid T > s) = \mathbb{P}(T > t), \quad \forall s, t \geq 0.$$

Cette équation fonctionnelle, avec $T > 0$ presque sûrement, n'admet que des solutions de la forme :

$$\mathbb{P}(T > t) = e^{-\lambda t}, \quad \lambda > 0,$$

ce qui est la fonction de survie d'une loi exponentielle de paramètre λ . On a donc bien $T \sim \mathcal{E}(\lambda)$.

Pour le prouver, posons $g(t) = \ln \mathbb{P}(T > t)$. La relation devient $g(t + s) = g(t) + g(s)$ pour tous $s, t \geq 0$, avec g décroissante et $g(0) = 0$. La seule solution continue (mesurable) est $g(t) = -\lambda t$, d'où $\mathbb{P}(T > t) = e^{-\lambda t}$. \square

1.4.1 Chaînes de Markov à temps discret (CMTD)

Avant d'aborder les processus à temps continu, il est essentiel de comprendre les fondements des chaînes de Markov à temps discret. Une chaîne de Markov à temps discret (CMTD) est une séquence de variables aléatoires qui évolue par sauts discrets et qui possède la propriété de Markov. L'étude des CMTD fournit une base théorique solide pour appréhender les processus de naissance et de mort, qui peuvent être vus comme une généralisation en temps continu.

Définition 1.4.1.1 (Chaîne de Markov à temps discret). *Une chaîne de Markov à temps discret est une suite de variables aléatoires $\{X_n, n \in \mathbb{N}\}$ à valeurs dans un espace d'états discret E (fini ou dénombrable). Elle satisfait la propriété de Markov, qui stipule que l'état futur ne dépend que de l'état présent, et non des états passés. Formellement, pour tout $n \geq 0$ et pour toute suite d'états $i_0, i_1, \dots, i_{n-1}, i, j \in E$:*

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

Définition 1.4.1.2 (Homogénéité). *Une CMTD est dite homogène si les probabilités de transition ne dépendent pas de l'instant n . Dans ce cas, on définit la probabilité de transition en une étape de l'état i vers l'état j par :*

$$p_{i,j} = \mathbb{P}(X_{n+1} = j \mid X_n = i), \quad \forall n \in \mathbb{N}.$$

Dans toute la suite, nous considérerons uniquement des chaînes homogènes.

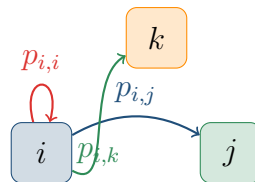


Figure : Transition d'état dans une chaîne de Markov

Matrice de transition. La matrice de transition P est la matrice (éventuellement infinie) dont l'entrée (i, j) est $p_{i,j}$. Cette matrice est dite stochastique, ce qui signifie que ses coefficients sont positifs et que la somme de chaque ligne est égale à 1 :

$$\forall i \in E, \quad \sum_{j \in E} p_{i,j} = 1.$$

Proposition 1.4.1.1 (Probabilités de transition à n étapes). *Les probabilités de transition en n étapes sont données par la puissance n -ième de la matrice de transition. En notant $p_{i,j}^{(n)} = \mathbb{P}(X_n = j \mid X_0 = i)$, nous avons :*

$$p_{i,j}^{(n)} = (P^n)_{i,j},$$

où $(P^n)_{i,j}$ est le terme (i, j) de la matrice P élevée à la puissance n . Ce résultat découle directement de l'équation de Chapman-Kolmogorov, qui s'écrit, pour tous $m, n \geq 0$:

$$p_{i,j}^{(m+n)} = \sum_{k \in E} p_{i,k}^{(m)} p_{k,j}^{(n)}.$$

Démonstration. Par récurrence sur n . Pour $n = 1$, c'est la définition. Supposons la propriété vraie au rang n . Alors :

$$p_{i,j}^{(n+1)} = \mathbb{P}(X_{n+1} = j \mid X_0 = i) = \sum_{k \in E} \mathbb{P}(X_n = k \mid X_0 = i) \mathbb{P}(X_{n+1} = j \mid X_n = k) = \sum_{k \in E} p_{i,k}^{(n)} p_{k,j}.$$

Ceci est exactement la formule de multiplication matricielle : $(P^{n+1})_{i,j} = (P^n P)_{i,j}$. \square

Loi de probabilité et évolution. Soit $\mu_n = (\mu_n(i))_{i \in E}$ la distribution de probabilité de X_n , où $\mu_n(i) = \mathbb{P}(X_n = i)$. La loi initiale est notée μ_0 . L'évolution de la chaîne est régie par l'équation :

$$\mu_{n+1} = \mu_n P, \quad \text{soit en composantes : } \mu_{n+1}(j) = \sum_{i \in E} \mu_n(i) p_{i,j}.$$

Par récurrence, on obtient directement que la loi à l'instant n est $\mu_n = \mu_0 P^n$.

1.4.2 Classification des états

La classification des états d'une chaîne de Markov permet de comprendre son comportement asymptotique.

Définition 1.4.2.1 (Communication et irréductibilité). *Un état j est accessible depuis un état i , noté $i \rightarrow j$, s'il existe $n \geq 0$ tel que $p_{ij}^{(n)} > 0$. Deux états i et j communiquent, noté $i \leftrightarrow j$, si chacun est accessible depuis l'autre. La relation \leftrightarrow est une relation d'équivalence. Si tous les états communiquent entre eux, la chaîne est dite irréductible.*

Définition 1.4.2.2 (Périodicité). *La période d'un état i , notée $d(i)$, est le plus grand commun diviseur de l'ensemble $\{n \geq 1 : p_{ii}^{(n)} > 0\}$. Si $d(i) = 1$, l'état est dit apériodique. La périodicité est une propriété de classe : tous les états d'une même classe communicante ont la même période. Une chaîne irréductible est dite apériodique si ses états sont apériodiques.*

Définition 1.4.2.3 (Récurrence et transient). *Soit $T_i = \inf\{n \geq 1 : X_n = i\}$ le temps de premier retour à l'état i (avec la convention $\inf \emptyset = +\infty$).*

- L'état i est dit récurrent si $\mathbb{P}_i(T_i < \infty) = 1$, c'est-à-dire si la probabilité d'y revenir un jour est certaine.
- Il est dit transient si $\mathbb{P}_i(T_i < \infty) < 1$.
- Un état récurrent i est dit récurrent positif si le temps moyen de retour $\mathbb{E}_i[T_i]$ est fini. Si le temps de retour moyen est infini, l'état est dit récurrent nul.

Dans une chaîne de Markov à espace d'états fini et irréductible, tous les états sont récurrents positifs.

Théorème 1.4.2.1 (Caractérisation de la récurrence). *Un état i est transient si et seulement si $\sum_{n=1}^{\infty} p_{i,i}^{(n)} < \infty$. Il est récurrent si et seulement si $\sum_{n=1}^{\infty} p_{i,i}^{(n)} = \infty$.*

Démonstration. Soit $N_i = \sum_{n=1}^{\infty} \mathbf{1}_{\{X_n=i\}}$ le nombre de visites à l'état i . On a $\mathbb{E}_i[N_i] = \sum_{n=1}^{\infty} p_{i,i}^{(n)}$. D'autre part, par la propriété de Markov, le nombre de visites suit une loi géométrique : $\mathbb{P}_i(N_i \geq k) = (\mathbb{P}_i(T_i < \infty))^k$. Donc $\mathbb{E}_i[N_i] = \frac{\mathbb{P}_i(T_i < \infty)}{1 - \mathbb{P}_i(T_i < \infty)}$. Cette quantité est finie si et seulement si $\mathbb{P}_i(T_i < \infty) < 1$, c'est-à-dire si l'état est transient. \square

1.4.3 Comportement stationnaire et asymptotique

L'étude du régime stationnaire est fondamentale pour les applications, notamment pour calculer les performances d'un système à l'équilibre.

Définition 1.4.3.1 (Mesure et probabilité stationnaire). *Une mesure de probabilité $\pi = (\pi_i)_{i \in E}$ sur l'espace d'états est dite stationnaire (ou invariante) pour la chaîne de matrice P si elle vérifie :*

$$\pi = \pi P, \quad \text{soit en composantes : } \forall j \in E, \quad \pi_j = \sum_{i \in E} \pi_i p_{i,j}.$$

Si la loi initiale μ_0 est stationnaire, alors toutes les lois marginales μ_n sont identiques à π : la chaîne est en régime stationnaire.

Théorème 1.4.3.1 (Existence et unicité pour une chaîne irréductible). *Pour une chaîne de Markov irréductible, il existe au plus une probabilité stationnaire. De plus, on a équivalence entre les trois propositions suivantes :*

1. Il existe une probabilité stationnaire π .
2. Il existe un état récurrent positif.
3. Tous les états sont récurrents positifs.

Si ces conditions sont remplies, la probabilité stationnaire est donnée par :

$$\pi_i = \frac{1}{\mathbb{E}_i[T_i]}, \quad \forall i \in E.$$

Démonstration. Esquisse : Si π est stationnaire, alors pour tout état i , on a $\pi_i = \frac{1}{\mathbb{E}_i[T_i]}$ (théorème du renouvellement pour les chaînes de Markov). L'unicité découle de l'irréductibilité. La réciproque utilise le fait que si un état est récurrent positif, alors la mesure définie par $\pi_j = \mathbb{E}_i \left[\sum_{n=1}^{T_i} \mathbf{1}_{\{X_n=j\}} \right] / \mathbb{E}_i[T_i]$ est une probabilité stationnaire. \square

Théorème 1.4.3.2 (Convergence vers l'équilibre). *Si une chaîne de Markov est irréductible, apériodique et récurrente positive (on dit qu'elle est ergodique), alors la distribution μ_n de X_n converge vers la probabilité stationnaire π , quelle que soit la distribution initiale μ_0 :*

$$\lim_{n \rightarrow \infty} \mu_n(i) = \pi_i, \quad \forall i \in E.$$

Dans le cas d'un espace d'états fini, cette convergence est exponentiellement rapide : il existe $K > 0$ et $\alpha \in]0, 1[$ tels que $\|\mu_n - \pi\| \leq K\alpha^n$.

Démonstration. Pour une chaîne finie irréductible et apériodique, la matrice P a une valeur propre 1 simple et toutes les autres valeurs propres sont de module strictement inférieur à 1. La décomposition spectrale donne alors la convergence exponentielle. \square

Théorème 1.4.3.3 (Loi forte des grands nombres (Théorème ergodique)). *Soit $\{X_n\}$ une chaîne de Markov irréductible admettant une probabilité stationnaire π . Pour toute fonction f telle que $\sum_i |f(i)|\pi_i < \infty$, la moyenne empirique de f le long de la trajectoire converge presque sûrement vers la moyenne spatiale sous la loi stationnaire :*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} f(X_k) = \sum_{i \in E} f(i)\pi_i, \quad p.s.$$

Ce théorème justifie l'estimation des probabilités stationnaires par l'observation d'une seule trajectoire sur un long horizon temporel.

Lien avec les processus de naissance et de mort à temps discret. Les processus de naissance et de mort à temps discret sont un cas particulier de CMTD où les transitions sont limitées aux états voisins. Leur matrice de transition P est tri-diagonale :

$$p_{i,i+1} = \lambda_i, \quad p_{i,i-1} = \mu_i \text{ (pour } i \geq 1), \quad p_{i,i} = 1 - \lambda_i - \mu_i,$$

avec $\mu_0 = 0$. L'étude de la récurrence et de la stationnarité de ces processus se fait via les équations de bilan, qui sont la version discrète des équations différentielles que nous verrons dans la section suivante.

1.4.4 Processus de Markov à temps continu et générateur infinitésimal

Revenons maintenant aux processus de Markov à temps continu, qui sont au cœur des modèles de files d'attente classiques.

Définition 1.4.4.1 (Générateur infinitésimal). *Soit $\{X(t), t \geq 0\}$ un processus de Markov à temps continu homogène, à espace d'états discret E . On définit le générateur infinitésimal $Q = (q_{ij})_{i,j \in E}$ par :*

$$q_{ij} = \lim_{h \rightarrow 0^+} \frac{\mathbb{P}(X(h) = j \mid X(0) = i)}{h}, \quad i \neq j,$$

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

Les quantités q_{ij} sont les taux de transition de l'état i vers l'état j .

Définition 1.4.4.2 (Équations de Kolmogorov). *Les probabilités de transition $P_{ij}(t) = \mathbb{P}(X(t) = j \mid X(0) = i)$ vérifient les équations différentielles de Kolmogorov :*

— *Équations forward (pour $t \geq 0$) :*

$$P'_{ij}(t) = \sum_{k \neq j} P_{ik}(t)q_{kj} - P_{ij}(t) \sum_{k \neq j} q_{jk},$$

ou sous forme matricielle : $P'(t) = P(t)Q$.

— *Équations backward* :

$$P'_{ij}(t) = \sum_{k \neq i} q_{ik} P_{kj}(t) - \left(\sum_{k \neq i} q_{ik} \right) P_{ij}(t),$$

soit : $P'(t) = QP(t)$.

Démonstration.[Équation forward de Kolmogorov] Par la propriété de Markov et l'homogénéité :

$$P_{ij}(t+h) = \sum_{k \in E} P_{ik}(t) P_{kj}(h).$$

En soustrayant $P_{ij}(t)$ et en divisant par h :

$$\frac{P_{ij}(t+h) - P_{ij}(t)}{h} = \sum_{k \neq j} P_{ik}(t) \frac{P_{kj}(h) - 1}{h} + P_{ij}(t) \frac{P_{jj}(h) - 1}{h}.$$

En passant à la limite $h \rightarrow 0^+$ et en utilisant les définitions de q_{kj} et q_{jj} , on obtient :

$$P'_{ij}(t) = \sum_{k \neq j} P_{ik}(t) q_{kj} + P_{ij}(t) q_{jj},$$

soit : $P'_{ij}(t) = \sum_{k \in E} P_{ik}(t) q_{kj}$. D'où la forme matricielle $P'(t) = P(t)Q$. La condition initiale est $P(0) = I$. \square

Remarque 1.4.4.1 (Solution formelle). *Si Q est indépendant du temps, la solution de $P'(t) = P(t)Q$ avec $P(0) = I$ est donnée par l'exponentielle de matrice :*

$$P(t) = e^{tQ} = \sum_{n=0}^{\infty} \frac{t^n Q^n}{n!}.$$

Définition 1.4.4.3 (Distribution stationnaire). *Une distribution de probabilité $\pi = (\pi_i)_{i \in E}$ est dite stationnaire (ou invariante) pour le processus de Markov si :*

$$\pi_j = \sum_{i \in E} \pi_i P_{ij}(t), \quad \forall t \geq 0, \forall j \in E.$$

En termes du générateur, cela équivaut à :

$$\pi Q = 0, \quad \sum_{i \in E} \pi_i = 1.$$

Démonstration.[Condition d'équilibre] Si π est stationnaire, alors pour tout j :

$$\pi_j = \sum_i \pi_i P_{ij}(t).$$

En dérivant par rapport à t et en évaluant en $t = 0$:

$$0 = \sum_i \pi_i P'_{ij}(0) = \sum_i \pi_i q_{ij}.$$

Donc $\pi Q = 0$. Réciproquement, si $\pi Q = 0$, alors $\pi e^{tQ} = \pi$ pour tout t , donc π est stationnaire. \square

1.5 Processus de naissance et de mort

Les processus de naissance et de mort constituent une classe particulière de processus de Markov à temps continu, extrêmement utile pour modéliser les files d'attente.

Définition 1.5.1 (Processus de naissance et de mort). *Un processus de naissance et de mort est un processus de Markov à temps continu $\{X(t), t \geq 0\}$ à valeurs dans $\{0, 1, 2, \dots\}$ dont les transitions sont limitées aux états voisins [7, 9] :*

- De l'état n vers l'état $n + 1$: c'est une naissance (arrivée d'un client), avec taux λ_n .
- De l'état n vers l'état $n - 1$ (pour $n \geq 1$) : c'est une mort (départ d'un client), avec taux μ_n .

Aucune autre transition n'est possible.

Définition 1.5.2 (Taux de transition). *Les paramètres d'un processus de naissance et de mort sont [1] :*

- λ_n : taux de naissance (arrivée) quand le système est dans l'état n .
- μ_n : taux de mort (départ) quand le système est dans l'état n , avec $\mu_0 = 0$ (aucun départ possible si le système est vide).

Pendant un intervalle de temps infinitésimal dt , les probabilités de transition sont :

$$\begin{aligned}\mathbb{P}(X(t + dt) = n + 1 \mid X(t) = n) &= \lambda_n dt + o(dt) \\ \mathbb{P}(X(t + dt) = n - 1 \mid X(t) = n) &= \mu_n dt + o(dt) \\ \mathbb{P}(X(t + dt) = n \mid X(t) = n) &= 1 - (\lambda_n + \mu_n)dt + o(dt)\end{aligned}$$

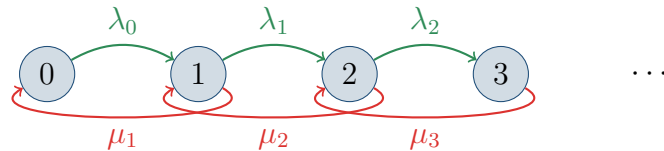


Figure : Diagramme des transitions d'un processus de naissance et de mort

Équations d'équilibre. Pour un processus de naissance et de mort, les équations de Kolmogorov deviennent des équations de bilan :

$$\frac{d\pi_n(t)}{dt} = \lambda_{n-1}\pi_{n-1}(t) + \mu_{n+1}\pi_{n+1}(t) - (\lambda_n + \mu_n)\pi_n(t).$$

À l'équilibre, $\frac{d\pi_n}{dt} = 0$, d'où le système :

$$\lambda_{n-1}\pi_{n-1} + \mu_{n+1}\pi_{n+1} = (\lambda_n + \mu_n)\pi_n, \quad n \geq 1,$$

et pour $n = 0$: $\lambda_0\pi_0 = \mu_1\pi_1$.

Équations de bilan détaillé. Pour un processus de naissance et de mort, on a en fait une relation plus forte appelée bilan détaillé (ou réversibilité) :

$$\lambda_n\pi_n = \mu_{n+1}\pi_{n+1}, \quad \forall n \geq 0.$$

Ces équations expriment qu'à l'équilibre, le flux de naissances de l'état n vers $n + 1$ est égal au flux de morts de l'état $n + 1$ vers n .

Démonstration. En sommant les équations d'équilibre pour $n = 0, 1, \dots, k$, on obtient par télescopage $\lambda_k \pi_k = \mu_{k+1} \pi_{k+1}$. \square

Solution stationnaire. En résolvant les équations de bilan détaillé par récurrence, on obtient :

$$\pi_n = \pi_0 \prod_{k=0}^{n-1} \frac{\lambda_k}{\mu_{k+1}}, \quad n \geq 1,$$

où π_0 est déterminé par la condition de normalisation $\sum_{n=0}^{\infty} \pi_n = 1$, soit :

$$\pi_0 = \left(1 + \sum_{n=1}^{\infty} \prod_{k=0}^{n-1} \frac{\lambda_k}{\mu_{k+1}} \right)^{-1}.$$

1.6 Définition et composants d'un système de file d'attente

Avant d'analyser des modèles spécifiques comme M/M/1 ou M/G/1, il est essentiel de définir formellement ce qu'est un système de file d'attente.

Définition 1.6.1 (Système de file d'attente). *Un système de file d'attente (ou chaîne) est un modèle mathématique décrivant un système où des entités appelées "clients" arrivent au fil du temps pour recevoir un service d'un ou plusieurs "serveurs". Si le serveur est occupé au moment de l'arrivée, le client fait la queue (attend) jusqu'à ce que le serveur soit libre.*

La théorie des files d'attente analyse les phénomènes d'attente afin d'optimiser les performances (temps d'attente, longueur de la file, taux d'occupation, etc.).

Proposition 1.6.1 (Composants fondamentaux). *Tout système de file d'attente peut être décrit par les six éléments suivants :*

1. **Processus d'arrivée** : La manière dont les clients arrivent (distribution des temps inter-arrivées).
2. **Distribution des temps de service** : La durée du traitement de chaque client.
3. **Nombre de serveurs** : Le nombre de postes de service fonctionnant en parallèle.
4. **Capacité du système** : Le nombre maximal de clients autorisés dans le système (file + serveurs).
5. **Taille de la population source** : Le nombre total de clients pouvant potentiellement arriver (infini ou fini).
6. **Discipline de service** : L'ordre dans lequel les clients sont servis (FIFO, LIFO, Priorité, etc.).

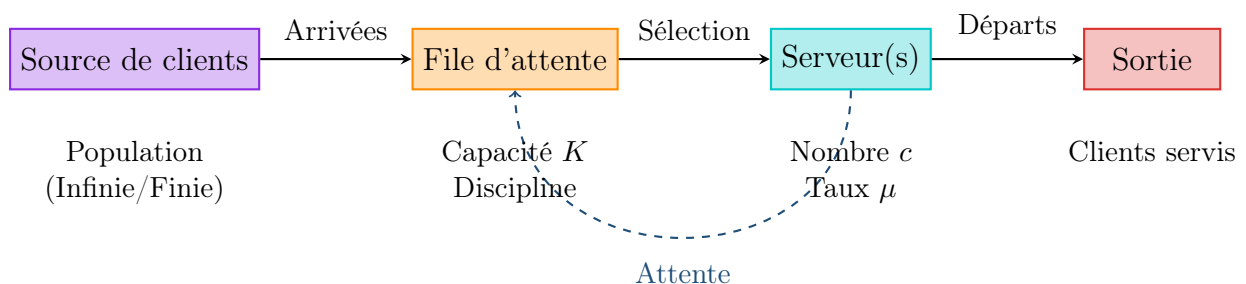


Figure : Schéma fonctionnel des composants d'un système de file d'attente

Remarque 1.6.1 (Hypothèses classiques). *En général, on suppose que :*

- Les clients sont indépendants entre eux.
- Les temps de service sont indépendants des temps inter-arrivées.
- Le système est en régime stationnaire après un certain temps (si les conditions de stabilité sont remplies).

1.7 Notation de Kendall pour les files d'attente

Afin de classer efficacement ces systèmes, David G. Kendall a proposé en 1953 une notation standardisée. Cette notation, introduite par David G. Kendall en 1953, permet de décrire de manière concise les caractéristiques d'un système de file d'attente.

Définition 1.7.1 (Notation de Kendall). *Un système de file d'attente est généralement décrit par la notation $A/S/c/K/N/D$, où :*

- A : distribution des temps entre arrivées (processus d'arrivée)
- S : distribution des temps de service
- c : nombre de serveurs
- K : capacité du système (taille maximale de la file + serveurs)
- N : taille de la population source
- D : discipline de service

Les symboles les plus courants pour les distributions sont :

- M : Markovien (exponentiel)
- D : Déterministe (constant)
- E_k : Erlang (distribution Gamma avec paramètre de forme k)
- H_k : Hyper-exponentiel d'ordre k
- G : Général (distribution quelconque)

Lorsque $K = \infty$, $N = \infty$ et $D = FIFO$ (First In First Out), on les omet souvent, écrivant simplement $A/S/c$.

Symbole	Distribution	Propriétés
M	Exponentielle	Absence de mémoire
D	Déterministe	Variance nulle
E_k	Erlang	Somme de k exponentielles
H_k	Hyper-exponentielle	Mélange de k exponentielles
G	Générale	Distribution quelconque

Figure : Symboles courants dans la notation de Kendall

1.8 Application aux modèles de files d'attente

Les processus de naissance et de mort permettent de modéliser une grande variété de systèmes de files d'attente [8, 19]. Les cas particuliers suivants sont essentiels.

1.8.1 Modèle M/M/1

Définition 1.8.1.1 (Modèle M/M/1). *Le modèle M/M/1 est un processus de naissance et de mort avec [11] :*

- Arrivées poissonniennes de taux constant : $\lambda_n = \lambda$ pour tout $n \geq 0$.
- Service exponentiel de taux constant : $\mu_n = \mu$ pour tout $n \geq 1$, et $\mu_0 = 0$.
- Un seul serveur, capacité infinie.

Proposition 1.8.1.1 (Distribution stationnaire de M/M/1). *Pour le modèle M/M/1 avec $\rho = \lambda/\mu < 1$, la distribution stationnaire est géométrique [1] :*

$$\pi_n = (1 - \rho)\rho^n, \quad n = 0, 1, 2, \dots$$

Les mesures de performance sont [8] :

$$L = \frac{\rho}{1 - \rho}, \quad L_q = \frac{\rho^2}{1 - \rho}, \quad W = \frac{1}{\mu - \lambda}, \quad W_q = \frac{\rho}{\mu - \lambda}$$

où L est le nombre moyen de clients dans le système, L_q le nombre moyen de clients en file d'attente, W le temps moyen passé dans le système et W_q le temps moyen d'attente.

Démonstration. Les équations d'équilibre (équations de bilan détaillé) s'écrivent : $\lambda\pi_n = \mu\pi_{n+1}$, $n \geq 0$. On en déduit $\pi_n = \left(\frac{\lambda}{\mu}\right)^n \pi_0 = \rho^n \pi_0$. La condition de normalisation $\sum_{n=0}^{\infty} \pi_n = 1$ donne $\pi_0 \sum_{n=0}^{\infty} \rho^n = \pi_0 \frac{1}{1-\rho} = 1$, donc $\pi_0 = 1 - \rho$ et $\pi_n = (1 - \rho)\rho^n$.

Pour L , on a $L = \sum_{n=0}^{\infty} n\pi_n = (1 - \rho) \sum_{n=0}^{\infty} n\rho^n = (1 - \rho) \frac{\rho}{(1-\rho)^2} = \frac{\rho}{1-\rho}$. Par la formule de Little, $W = L/\lambda = \frac{\rho}{\lambda(1-\rho)} = \frac{1}{\mu-\lambda}$. De même, $L_q = \sum_{n=1}^{\infty} (n-1)\pi_n = L - (1 - \pi_0) = \frac{\rho}{1-\rho} - \rho = \frac{\rho^2}{1-\rho}$, et $W_q = L_q/\lambda = \frac{\rho^2}{\lambda(1-\rho)} = \frac{\rho}{\mu-\lambda}$. \square

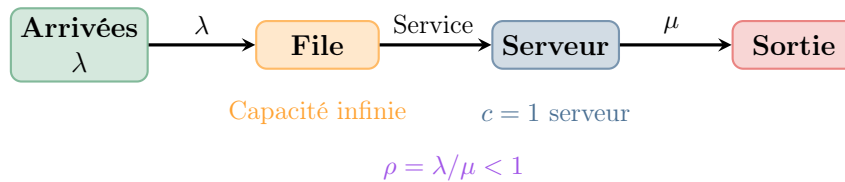


Figure : Schéma du modèle M/M/1

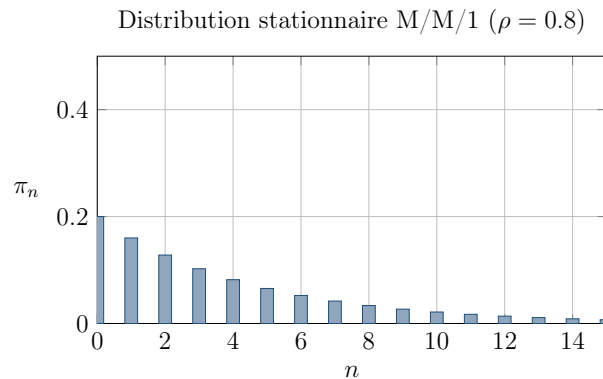


Figure : Distribution stationnaire M/M/1 ($\rho = 0.8$)

1.8.2 Modèle M/M/1/K (capacité finie)

Définition 1.8.2.1 (Modèle M/M/1/K). *Le modèle M/M/1/K est un processus de naissance et de mort avec [19] :*

- $\lambda_n = \lambda$ pour $0 \leq n < K$, et $\lambda_K = 0$ (aucune arrivée quand le système est plein).
- $\mu_n = \mu$ pour $1 \leq n \leq K$.
- Un seul serveur, capacité maximale K (incluant le client en service).

Proposition 1.8.2.1 (Distribution stationnaire de M/M/1/K). *Pour le modèle M/M/1/K avec $\rho = \lambda/\mu \neq 1$ [8] :*

$$\pi_n = \frac{1 - \rho}{1 - \rho^{K+1}} \rho^n, \quad n = 0, 1, \dots, K$$

Si $\rho = 1$, la distribution est uniforme : $\pi_n = \frac{1}{K+1}$ pour tout n .

La probabilité de rejet est $P_{\text{rejet}} = \pi_K$.

Démonstration. Les équations d'équilibre sont $\lambda\pi_n = \mu\pi_{n+1}$ pour $0 \leq n \leq K - 1$. On a donc $\pi_n = \rho^n \pi_0$ pour $0 \leq n \leq K$. La normalisation donne :

$$\pi_0 \sum_{n=0}^K \rho^n = \pi_0 \frac{1 - \rho^{K+1}}{1 - \rho} = 1 \quad (\rho \neq 1)$$

d'où $\pi_0 = \frac{1 - \rho}{1 - \rho^{K+1}}$. Si $\rho = 1$, alors $\pi_n = \pi_0$ pour tout n , et $(K + 1)\pi_0 = 1$, donc $\pi_0 = 1/(K + 1)$. □

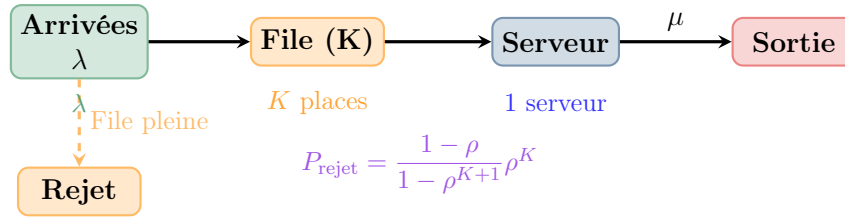


Figure : Modèle M/M/1/K avec rejet

1.8.3 Modèle M/M/c (serveurs multiples)

Définition 1.8.3.1 (Modèle M/M/c). *Le modèle M/M/c (ou M/M/∞ pour $c = \infty$) est un processus de naissance et de mort avec [11] :*

- $\lambda_n = \lambda$ pour tout $n \geq 0$ (arrivées indépendantes du nombre de clients).
- $\mu_n = \min(n, c)\mu$ pour $n \geq 1$ (taux de service proportionnel au nombre de serveurs occupés).

Proposition 1.8.3.1 (Distribution stationnaire de M/M/c). *Pour le modèle M/M/c avec $\rho = \lambda/(c\mu) < 1$ [1] :*

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

$$\pi_n = \begin{cases} \frac{(c\rho)^n}{n!} \pi_0, & 0 \leq n \leq c \\ \frac{c^n \rho^n}{c!} \pi_0, & n \geq c \end{cases}$$

La formule d'Erlang-C donne la probabilité d'attente [6] :

$$\mathbb{P}(\text{attente}) = \frac{(c\rho)^c}{c!(1-\rho)}\pi_0$$

Démonstration. Les équations d'équilibre détaillé s'écrivent $\lambda\pi_{n-1} = \mu_n\pi_n$, soit :

— Pour $1 \leq n \leq c$: $\lambda\pi_{n-1} = n\mu\pi_n$, donc $\pi_n = \frac{c\rho}{n}\pi_{n-1} = \frac{(c\rho)^n}{n!}\pi_0$.

— Pour $n > c$: $\lambda\pi_{n-1} = c\mu\pi_n$, donc $\pi_n = \rho\pi_{n-1} = \rho^{n-c}\pi_c = \rho^{n-c}\frac{(c\rho)^c}{c!}\pi_0 = \frac{c^c\rho^n}{c!}\pi_0$.

La condition de normalisation $\sum_{n=0}^{\infty} \pi_n = 1$ donne l'expression de π_0 . La probabilité d'attente est $\sum_{n=c}^{\infty} \pi_n = \pi_c \sum_{n=c}^{\infty} \rho^{n-c} = \frac{\pi_c}{1-\rho} = \frac{(c\rho)^c}{c!(1-\rho)}\pi_0$. \square

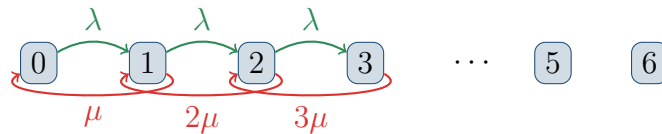


Figure : Diagramme de transitions pour M/M/c ($c = 3$)

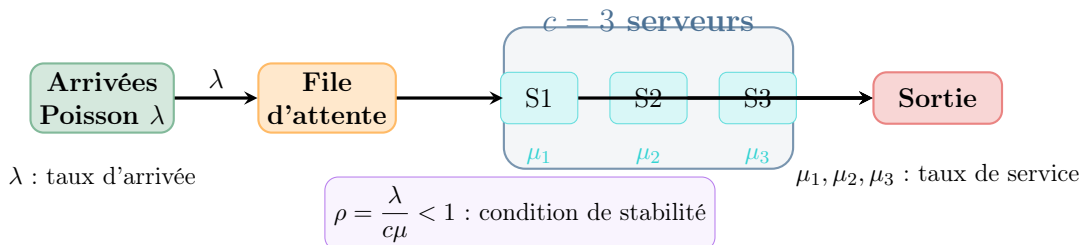


Figure : Schéma du modèle M/M/c à serveurs multiples ($c = 3$)

1.8.4 Modèle M/M/c/c (système avec pertes)

Définition 1.8.4.1 (Modèle M/M/c/c). *Le modèle M/M/c/c (ou système d'Erlang-B) est un processus de naissance et de mort avec [6] :*

- $\lambda_n = \lambda$ pour $0 \leq n < c$, et $\lambda_c = 0$.
- $\mu_n = n\mu$ pour $1 \leq n \leq c$.
- Capacité c (pas de file d'attente, les clients sont perdus si tous les serveurs sont occupés).

Proposition 1.8.4.1 (Formule d'Erlang-B). *Pour le modèle M/M/c/c, la distribution stationnaire est [19] :*

$$\pi_n = \frac{(\lambda/\mu)^n/n!}{\sum_{k=0}^c (\lambda/\mu)^k/k!}, \quad n = 0, 1, \dots, c$$

La probabilité de rejet (formule d'Erlang-B) est [6] :

$$P_{\text{rejet}} = \pi_c = \frac{(\lambda/\mu)^c/c!}{\sum_{k=0}^c (\lambda/\mu)^k/k!}$$

Démonstration. Les équations d'équilibre sont $\lambda\pi_{n-1} = n\mu\pi_n$ pour $1 \leq n \leq c$. On en déduit $\pi_n = \frac{(\lambda/\mu)^n}{n!}\pi_0$. La normalisation donne $\pi_0 \sum_{n=0}^c \frac{(\lambda/\mu)^n}{n!} = 1$, donc $\pi_0 = 1/\sum_{n=0}^c \frac{(\lambda/\mu)^n}{n!}$. \square

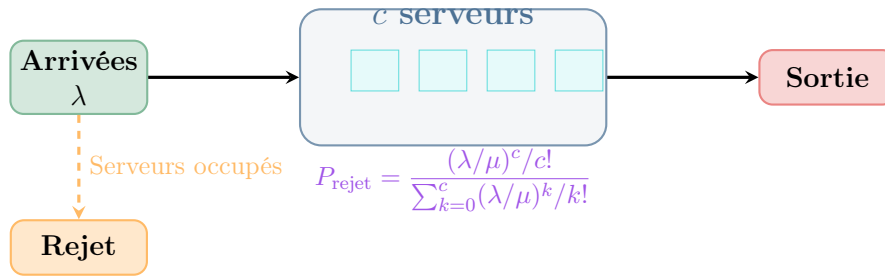


Figure : Modèle M/M/c/c (système d'Erlang-B)

1.8.5 Modèle M/M/∞ (nombre infini de serveurs)

Définition 1.8.5.1 (Modèle M/M/∞). *Le modèle M/M/∞ est un processus de naissance et de mort avec [18] :*

- $\lambda_n = \lambda$ pour tout $n \geq 0$.
- $\mu_n = n\mu$ pour $n \geq 1$ (chaque client est servi immédiatement, pas d'attente).

Ce modèle correspond à un système où il y a toujours un serveur disponible (self-service).

Proposition 1.8.5.1 (Distribution stationnaire de M/M/∞). *Pour le modèle M/M/∞, la distribution stationnaire est une loi de Poisson [11] :*

$$\pi_n = \frac{(\lambda/\mu)^n}{n!} e^{-\lambda/\mu}, \quad n = 0, 1, 2, \dots$$

Le nombre moyen de clients est $L = \lambda/\mu$, et il n'y a pas d'attente ($W_q = 0$, $W = 1/\mu$).

Démonstration. Les équations d'équilibre sont $\lambda\pi_{n-1} = n\mu\pi_n$ pour $n \geq 1$. On obtient $\pi_n = \frac{\lambda}{n\mu}\pi_{n-1} = \frac{(\lambda/\mu)^n}{n!}\pi_0$. La normalisation donne $\pi_0 \sum_{n=0}^{\infty} \frac{(\lambda/\mu)^n}{n!} = \pi_0 e^{\lambda/\mu} = 1$, donc $\pi_0 = e^{-\lambda/\mu}$ et $\pi_n = \frac{(\lambda/\mu)^n}{n!} e^{-\lambda/\mu}$. La moyenne d'une loi de Poisson est λ/μ . \square

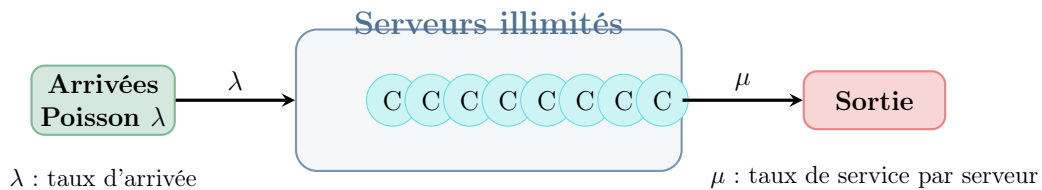


Figure : Schéma du modèle M/M/∞

1.9 Loi de Little

Théorème 1.9.1 (Loi de Little). *Considérons un système de file d'attente en régime stationnaire. Soit :*

- L : le nombre moyen de clients dans le système,
- λ : le taux d'arrivée moyen des clients dans le système,
- W : le temps moyen passé par un client dans le système.

Alors, sous des conditions de régularité très générales, on a la relation fondamentale [12] :

$$\boxed{L = \lambda W}$$

De plus, cette relation s'applique également à la file d'attente seule :

$$L_q = \lambda W_q$$

où L_q est le nombre moyen de clients en file d'attente et W_q le temps moyen d'attente dans la file.

Remarque 1.9.1. La puissance de la loi de Little réside dans sa généralité. Elle ne dépend d'aucune hypothèse sur les distributions des temps d'inter-arrivées ou des temps de service, ni sur la discipline de service, ni sur le nombre de serveurs. Elle s'applique à tout système en régime stationnaire pourvu que les moyennes soient bien définies.

Démonstration.[Approche heuristique] Considérons un intervalle de temps long $[0, T]$. Soit $A(T)$ le nombre d'arrivées dans cet intervalle et $D(T)$ le nombre de départs. Le nombre moyen d'arrivées par unité de temps est $\lambda_T = A(T)/T$.

Pour chaque client i , soit W_i son temps de séjour dans le système. L'aire sous la courbe du nombre de clients dans le système, notée $S(T)$, peut être calculée de deux façons :

1. Par intégration directe : $S(T) = \int_0^T N(t)dt$, où $N(t)$ est le nombre de clients présents à l'instant t .
2. En sommant les temps de séjour de tous les clients arrivés pendant $[0, T]$ (en négligeant les clients présents au début et à la fin pour T grand) : $S(T) \approx \sum_{i=1}^{A(T)} W_i$.

En égalant ces deux expressions, on obtient : $\int_0^T N(t)dt \approx \sum_{i=1}^{A(T)} W_i$. Divisons par T :

$$\frac{1}{T} \int_0^T N(t)dt \approx \frac{A(T)}{T} \cdot \frac{1}{A(T)} \sum_{i=1}^{A(T)} W_i.$$

Lorsque $T \rightarrow \infty$, sous réserve de convergence, le terme de gauche converge vers L (moyenne temporelle du nombre de clients), $A(T)/T$ converge vers λ (taux d'arrivée moyen), et $\frac{1}{A(T)} \sum_{i=1}^{A(T)} W_i$ converge vers W (temps de séjour moyen). D'où $L = \lambda W$. \square

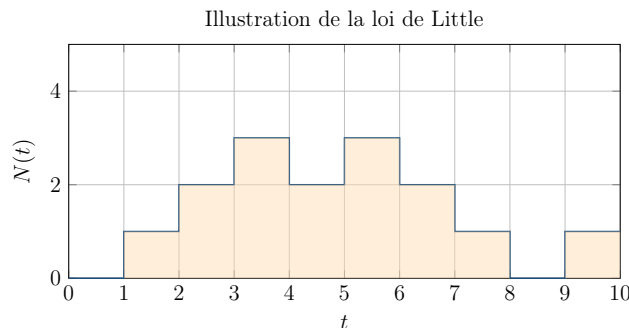


Figure : Illustration de la loi de Little : l'aire sous la courbe = \sum temps de séjour

1.10 Résumé des modèles et leurs caractéristiques

Le tableau suivant récapitule les principaux modèles issus des processus de naissance et de mort [19] :

Modèle	λ_n	μ_n	Caractéristiques
M/M/1	λ	μ	Un serveur, capacité infinie
M/M/1/K	$\lambda \mathbf{1}_{\{n < K\}}$	$\mu \mathbf{1}_{\{n > 0\}}$	Capacité finie K
M/M/c	λ	$\min(n, c)\mu$	c serveurs en parallèle
M/M/c/c	$\lambda \mathbf{1}_{\{n < c\}}$	$n\mu$	Pas de file, pertes
M/M/ ∞	λ	$n\mu$	Infini serveurs

TABLE 1.1 – Caractéristiques des modèles de naissance et de mort

Modèle	Distribution stationnaire
M/M/1	$\pi_n = (1 - \rho)\rho^n$
M/M/1/K	$\pi_n = \frac{1 - \rho}{1 - \rho^{K+1}}\rho^n$
M/M/c	$\pi_n = \begin{cases} \frac{(c\rho)^n}{n!}\pi_0, & n \leq c \\ \frac{c^c \rho^n}{c!}\pi_0, & n \geq c \end{cases}$
M/M/c/c	$\pi_n = \frac{(\lambda/\mu)^n / n!}{\sum_{k=0}^c (\lambda/\mu)^k / k!}$
M/M/ ∞	$\pi_n = \frac{(\lambda/\mu)^n}{n!} e^{-\lambda/\mu}$

TABLE 1.2 – Distributions stationnaires des modèles

Chapitre 2

Analyse exacte du système M/G/1 : chaînes de Markov incluses

L'analyse des systèmes de files d'attente avec des temps de service généraux constitue une avancée majeure dans la théorie des files d'attente. Le modèle M/G/1, où les arrivées sont poissonniennes mais les temps de service suivent une loi générale, est d'une importance considérable pour les applications réelles. Ce chapitre présente l'approche fondamentale des chaînes de Markov incluses (imbedded Markov chains) pour l'analyse exacte de ce système.

2.1 Limites des processus markoviens à temps continu

Définition 2.1 (Système M/G/1). *Le système M/G/1 est une file d'attente avec les caractéristiques suivantes :*

- **M** (Markovien) : Les arrivées sont poissonniennes de taux $\lambda > 0$.
- **G** (Général) : Les temps de service sont indépendants et identiquement distribués selon une loi générale $B(t)$ de moyenne $\mathbb{E}[S] = 1/\mu$ et de variance σ^2 .
- **1** : Un seul serveur.

La discipline de service est FIFO (First In, First Out) par défaut.

Remarque 2.1 (Problème de la non-markovianité). *Pour le système M/M/1, le processus $\{X(t), t \geq 0\}$ (nombre de clients dans le système) est un processus de Markov à temps continu car les temps de service sont exponentiels et possèdent la propriété d'absence de mémoire.*

Dans le cas M/G/1, les temps de service suivent une loi générale. Soit S_n le temps de service du n -ième client. La propriété suivante est fondamentale :

$$\mathbb{P}(S > t + s \mid S > s) \neq \mathbb{P}(S > t) \quad \text{pour une loi générale.}$$

Par conséquent, le temps restant de service dépend du temps déjà écoulé. Le processus $\{X(t), t \geq 0\}$ n'est plus markovien car l'évolution future dépend non seulement du nombre de clients mais aussi de la quantité de service déjà effectuée sur le client en cours.

Exemple 2.1 (Contre-exemple de non-markovianité). *Considérons un système M/G/1 avec des temps de service déterministes $S = c$ constante. Si à l'instant t le serveur est occupé, le temps restant de service dépend de l'instant où le service a commencé. Deux situations avec le même nombre de clients peuvent avoir des évolutions futures différentes :*

- Situation A : Le service a commencé il y a $0.1c$ secondes. Le temps restant est $0.9c$.
- Situation B : Le service a commencé il y a $0.8c$ secondes. Le temps restant est $0.2c$.

Le processus $\{X(t)\}$ n'est donc pas markovien. Il faut un espace d'états plus riche incluant l'âge ou le temps restant du service en cours.

2.2 Chaînes de Markov incluses (Imbedded Markov Chains)

L'idée fondamentale de l'approche par chaînes de Markov incluses est de n'observer le système qu'à des instants particuliers où la propriété de Markov est rétablie.

2.2.1 Points de régénération

Définition 2.2 (Points de régénération). On appelle points de régénération (ou instants de renouvellement) des instants aléatoires t_1, t_2, \dots tels que :

1. Le comportement futur du système après t_n ne dépend que de l'état du système à t_n .
2. Le processus recommence probabilistiquement à neuf à chaque point de régénération.

Définition 2.3 (Instants de départ). Pour le système $M/G/1$, on choisit comme points de régénération les instants de départ des clients, c'est-à-dire les instants où un client termine son service et quitte le système. Soit D_n l'instant du n -ième départ. On note :

$X_n = X(D_n^+) =$ nombre de clients laissés dans le système par le n -ième client sortant.

C'est le nombre de clients juste après le départ du n -ième client.

Proposition 2.1 (Propriété de Markov aux instants de départ). Le processus $\{X_n, n \geq 1\}$ est une chaîne de Markov à temps discret à valeurs dans $\{0, 1, 2, \dots\}$. Cette chaîne est appelée **chaîne de Markov incluse** (imbedded Markov chain).

En effet, à l'instant d'un départ, le serveur devient libre (si $X_n = 0$) ou commence immédiatement à servir le client suivant (si $X_n > 0$). L'état futur ne dépend que de X_n et des arrivées futures, qui sont indépendantes du passé par la propriété de Poisson.

Démonstration. Soit X_n le nombre de clients juste après le n -ième départ. Le $(n+1)$ -ième départ se produit après :

- Un temps de service S_{n+1} du client suivant (si $X_n > 0$),
- Ou un temps d'inactivité (période vide) suivi d'un temps de service (si $X_n = 0$).

Pendant le temps de service, le nombre d'arrivées suit une loi de Poisson de paramètre λS_{n+1} . Par la propriété de Poisson, ces arrivées sont indépendantes du passé. Donc X_{n+1} ne dépend de l'histoire avant D_n qu'à travers X_n . \square

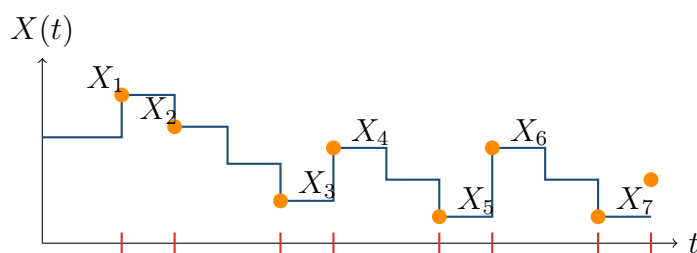


Figure : Instants de départ et chaîne incluse X_n

2.2.2 Relation de récurrence fondamentale

Théorème 2.1 (Équation de récurrence pour la chaîne incluse). *Soit X_n le nombre de clients juste après le n -ième départ. Alors :*

$$X_{n+1} = X_n - \mathbf{1}_{\{X_n > 0\}} + A_{n+1},$$

où A_{n+1} est le nombre d'arrivées pendant le service du $(n+1)$ -ième client.

Si $X_n = 0$, le serveur est inactif. Le $(n+1)$ -ième départ correspond au premier client de la période d'occupation suivante. Dans ce cas, on peut écrire :

$$X_{n+1} = A_{n+1}^{(0)},$$

où $A_{n+1}^{(0)}$ est le nombre d'arrivées pendant une période d'inactivité (qui suit une loi exponentielle de taux λ) suivie du temps de service.

Remarque 2.2. *La variable A_{n+1} est indépendante de X_n et de l'histoire passée. Sa distribution est donnée par :*

$$a_k = \mathbb{P}(A = k) = \int_0^\infty \frac{(\lambda t)^k}{k!} e^{-\lambda t} dB(t), \quad k = 0, 1, 2, \dots$$

où $B(t)$ est la fonction de répartition des temps de service.

2.2.3 Matrice de transition

Définition 2.4 (Matrice de transition de la chaîne incluse). *La matrice de transition $P = (p_{ij})_{i,j \geq 0}$ de la chaîne $\{X_n\}$ est donnée par :*

$$p_{ij} = \mathbb{P}(X_{n+1} = j \mid X_n = i).$$

Pour $i \geq 1$ (serveur occupé) :

$$p_{ij} = \mathbb{P}(i - 1 + A = j) = \mathbb{P}(A = j - i + 1) = a_{j-i+1},$$

avec $a_k = 0$ si $k < 0$.

Pour $i = 0$ (serveur inactif) :

$$p_{0j} = \mathbb{P}(X_{n+1} = j \mid X_n = 0) = \mathbb{P}(A^{(0)} = j) = a_j^{(0)},$$

où $a_j^{(0)}$ est la probabilité que j clients arrivent pendant une période d'inactivité suivie d'un service.

Proposition 2.2 (Forme de la matrice de transition). *La matrice de transition P est de la forme :*

$$P = \begin{pmatrix} a_0^{(0)} & a_1^{(0)} & a_2^{(0)} & a_3^{(0)} & \cdots \\ a_0 & a_1 & a_2 & a_3 & \cdots \\ 0 & a_0 & a_1 & a_2 & \cdots \\ 0 & 0 & a_0 & a_1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

C'est une matrice de Toeplitz par blocs.

2.2.4 Distribution stationnaire de la chaîne incluse

Théorème 2.2 (Condition de stabilité). *La chaîne de Markov incluse $\{X_n\}$ est récurrente positive (ergodique) si et seulement si :*

$$\rho = \lambda \mathbb{E}[S] = \frac{\lambda}{\mu} < 1.$$

C'est la condition de stabilité du système $M/G/1$. Dans ce cas, il existe une distribution stationnaire unique $\pi = (\pi_0, \pi_1, \pi_2, \dots)$.

Démonstration. La condition de stabilité découle du théorème de Loynes pour les files d'attente $G/G/1$. Pour une file $M/G/1$, le processus de la charge de travail est récurrent si et seulement si $\lambda \mathbb{E}[S] < 1$. La chaîne incluse hérite de cette propriété car les instants de départ sont des points de régénération. \square

Théorème 2.3 (Équations d'équilibre de la chaîne incluse). *La distribution stationnaire π de la chaîne incluse satisfait le système d'équations :*

$$\pi_j = \sum_{i=0}^{\infty} \pi_i p_{ij}, \quad j \geq 0.$$

En développant, on obtient pour $j \geq 0$:

$$\pi_j = \pi_0 a_j^{(0)} + \sum_{i=1}^{j+1} \pi_i a_{j-i+1}.$$

Ces équations sont à la base de la dérivation de la fonction génératrice des probabilités stationnaires.

2.3 Fonction génératrice et transformée

2.3.1 Définition des fonctions génératrices

Définition 2.5 (Fonction génératrice de la distribution stationnaire). *Soit $\Pi(z)$ la fonction génératrice de la distribution stationnaire $\{\pi_j\}_{j \geq 0}$:*

$$\Pi(z) = \sum_{j=0}^{\infty} \pi_j z^j, \quad |z| \leq 1.$$

Définition 2.6 (Fonction génératrice des arrivées pendant un service). *Soit $A(z)$ la fonction génératrice du nombre d'arrivées pendant un temps de service :*

$$A(z) = \sum_{k=0}^{\infty} a_k z^k = \sum_{k=0}^{\infty} \left(\int_0^{\infty} \frac{(\lambda t)^k}{k!} e^{-\lambda t} dB(t) \right) z^k.$$

En intervertissant somme et intégrale, on obtient une expression simple :

$$A(z) = \int_0^{\infty} e^{-\lambda t} \sum_{k=0}^{\infty} \frac{(\lambda t z)^k}{k!} dB(t) = \int_0^{\infty} e^{-\lambda t(1-z)} dB(t).$$

C'est donc la transformée de Laplace-Stieltjes de la distribution des temps de service évaluée en $\lambda(1-z)$:

$$A(z) = \tilde{B}^*(\lambda(1-z)),$$

où $\tilde{B}^(s) = \int_0^{\infty} e^{-st} dB(t)$ est la transformée de Laplace-Stieltjes de $B(t)$.*

2.3.2 Dérivation de la formule de Pollaczek-Khinchine

Théorème 2.4 (Formule de Pollaczek-Khinchine pour la fonction génératrice). *Pour le système $M/G/1$ en régime stationnaire avec $\rho < 1$, la fonction génératrice $\Pi(z)$ de la distribution stationnaire du nombre de clients juste après les départs est donnée par :*

$$\Pi(z) = \frac{(1 - \rho)(1 - z)A(z)}{A(z) - z}$$

Démonstration. Les équations d'équilibre de la chaîne incluse peuvent être transformées en une équation fonctionnelle pour $\Pi(z)$.

Multiplions l'équation $\pi_j = \pi_0 a_j^{(0)} + \sum_{i=1}^{j+1} \pi_i a_{j-i+1}$ par z^j et sommions sur j de 0 à ∞ :

$$\sum_{j=0}^{\infty} \pi_j z^j = \pi_0 \sum_{j=0}^{\infty} a_j^{(0)} z^j + \sum_{j=0}^{\infty} \sum_{i=1}^{j+1} \pi_i a_{j-i+1} z^j.$$

Le terme de gauche est $\Pi(z)$. Le premier terme de droite est $\pi_0 A^{(0)}(z)$, où $A^{(0)}(z)$ est la fonction génératrice des $a_j^{(0)}$.

Pour le deuxième terme, effectuons le changement d'indice $k = j - i + 1$. Alors $j = i - 1 + k$, et la somme devient :

$$\sum_{i=1}^{\infty} \pi_i z^{i-1} \sum_{k=0}^{\infty} a_k z^k = \frac{1}{z} \sum_{i=1}^{\infty} \pi_i z^i A(z) = \frac{A(z)}{z} (\Pi(z) - \pi_0).$$

On a donc l'équation fonctionnelle :

$$\Pi(z) = \pi_0 A^{(0)}(z) + \frac{A(z)}{z} (\Pi(z) - \pi_0).$$

Il reste à déterminer $A^{(0)}(z)$ et π_0 . On montre que $A^{(0)}(z) = \frac{1 - A(z)}{\lambda \mathbb{E}[S](1 - z)}$ (relation de renouvellement). De plus, en utilisant $\Pi(1) = 1$ et la règle de L'Hôpital, on obtient $\pi_0 = 1 - \rho$.

En substituant, on obtient :

$$\Pi(z) = \pi_0 \frac{1 - A(z)}{\lambda \mathbb{E}[S](1 - z)} + \frac{A(z)}{z} (\Pi(z) - \pi_0).$$

En résolvant pour $\Pi(z)$ et en utilisant $\lambda \mathbb{E}[S] = \rho$, on trouve :

$$\Pi(z) = \frac{(1 - \rho)A(z)(1 - z)}{A(z) - z} \cdot \frac{1}{1 - z} = \frac{(1 - \rho)(1 - z)A(z)}{A(z) - z}.$$

D'où le résultat. □

2.3.3 La formule de Pollaczek-Khinchine pour le nombre moyen de clients

Théorème 2.5 (Formule de Pollaczek-Khinchine pour L). *Le nombre moyen de clients dans le système $M/G/1$ en régime stationnaire est donné par :*

$$L = \rho + \frac{\lambda^2 \mathbb{E}[S^2]}{2(1 - \rho)} = \rho + \frac{\rho^2(1 + \sigma^2 \mu^2)}{2(1 - \rho)}$$

où $\mathbb{E}[S^2]$ est le moment d'ordre 2 du temps de service et $\sigma^2 = \mathbb{E}[S^2] - (\mathbb{E}[S])^2$ sa variance.

Démonstration. La fonction génératrice $\Pi(z)$ nous donne le nombre moyen de clients par $L = \Pi'(1)$. Calculons la dérivée de $\Pi(z)$ en $z = 1$.

Posons $N(z) = (1 - \rho)(1 - z)A(z)$ et $D(z) = A(z) - z$. Alors $\Pi(z) = N(z)/D(z)$.

En $z = 1$, on a $A(1) = 1$ (car c'est une fonction génératrice). Donc $D(1) = 0$ et $N(1) = 0$. Utilisons un développement limité.

Développons $A(z)$ autour de $z = 1$:

$$A(z) = 1 + A'(1)(z - 1) + \frac{A''(1)}{2}(z - 1)^2 + o((z - 1)^2).$$

Avec $A'(1) = \rho$ et $A''(1) = \lambda^2 \mathbb{E}[S^2]$.

Posons $u = 1 - z$. Alors $z = 1 - u$, et pour $u \rightarrow 0$:

$$\begin{aligned} A(z) &= 1 - \rho u + \frac{A''(1)}{2}u^2 + o(u^2), \\ A(z) - z &= \left(1 - \rho u + \frac{A''(1)}{2}u^2\right) - (1 - u) + o(u^2) = (1 - \rho)u + \frac{A''(1)}{2}u^2 + o(u^2), \\ (1 - z)A(z) &= u \left(1 - \rho u + \frac{A''(1)}{2}u^2\right) = u - \rho u^2 + \frac{A''(1)}{2}u^3. \end{aligned}$$

Donc :

$$\Pi(z) = \frac{(1 - \rho) \left(u - \rho u^2 + \frac{A''(1)}{2}u^3\right)}{(1 - \rho)u + \frac{A''(1)}{2}u^2 + o(u^2)} = (1 - \rho) \frac{1 - \rho u + \frac{A''(1)}{2}u^2}{(1 - \rho) + \frac{A''(1)}{2}u + o(u)}.$$

Pour $u \rightarrow 0$, on a :

$$\Pi(z) = \frac{1 - \rho}{1 - \rho} \left(1 - \rho u + \frac{A''(1)}{2}u^2\right) \left(1 - \frac{A''(1)}{2(1 - \rho)}u + o(u)\right).$$

Le développement à l'ordre 1 donne :

$$\Pi(z) = 1 - \left(\rho + \frac{A''(1)}{2(1 - \rho)}\right)u + o(u).$$

Puisque $u = 1 - z$, on a $\Pi(z) = 1 - L(1 - z) + o(1 - z)$. Par identification :

$$L = \rho + \frac{A''(1)}{2(1 - \rho)} = \rho + \frac{\lambda^2 \mathbb{E}[S^2]}{2(1 - \rho)}.$$

□

2.3.4 La formule de Pollaczek-Khinchine pour le temps d'attente moyen

Théorème 2.6 (Formule de Pollaczek-Khinchine pour W). *Par la loi de Little, le temps moyen passé dans le système $M/G/1$ est :*

$$W = \frac{L}{\lambda} = \frac{1}{\mu} + \frac{\lambda \mathbb{E}[S^2]}{2(1 - \rho)} = \frac{1}{\mu} + \frac{\rho(1 + \sigma^2 \mu^2)}{2\mu(1 - \rho)}.$$

Le temps moyen d'attente dans la file (hors service) est :

$$W_q = W - \frac{1}{\mu} = \frac{\lambda \mathbb{E}[S^2]}{2(1 - \rho)} = \frac{\rho(1 + \sigma^2 \mu^2)}{2\mu(1 - \rho)}.$$

Remarque 2.3. La formule de Pollaczek-Khinchine montre que le temps d'attente moyen dépend du moment d'ordre 2 du temps de service. Plus la variance des temps de service est grande, plus le temps d'attente moyen est grand. Cela illustre l'effet négatif de la variabilité sur les performances du système.

2.4 Analyse de la période d'occupation (Busy Period)

2.4.1 Définition et propriétés fondamentales

Définition 2.7 (Période d'occupation). On appelle période d'occupation (busy period) un intervalle de temps pendant lequel le serveur reste continuellement occupé. Elle commence à l'instant où un client arrive dans un système vide, et se termine à l'instant où le système redevient vide.

La période d'occupation est notée B . On distingue également :

- La période d'occupation initiale : commence avec un client qui arrive dans un système vide.
- Le cycle : période d'occupation suivie d'une période d'inactivité (idle period).

Proposition 2.3 (Structure de la période d'occupation). Soit B la durée d'une période d'occupation commençant avec un seul client. La période d'occupation peut être décomposée en :

$$B = S_1 + \sum_{i=1}^{A_1} B_i,$$

où :

- S_1 est le temps de service du premier client,
- A_1 est le nombre d'arrivées pendant ce service,
- B_i sont des périodes d'occupation indépendantes et identiquement distribuées, chacune ayant la même distribution que B .

Cette équation est l'équation de renouvellement de la période d'occupation.

Démonstration. Lorsque le serveur traite le premier client de la période d'occupation, des arrivées peuvent survenir. Si k clients arrivent pendant le service du premier client, alors à la fin de ce service, il y a k clients dans le système. Le serveur va ensuite traiter ces k clients, ce qui génère k sous-périodes d'occupation. Par la propriété de renouvellement, ces sous-périodes sont indépendantes et ont la même distribution que la période d'occupation initiale. \square

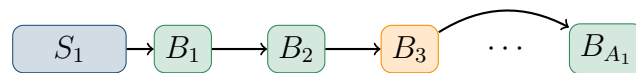


Figure : Décomposition de la période d'occupation

2.4.2 Transformée de Laplace de la période d'occupation

Théorème 2.7 (Équation fonctionnelle pour $B^*(s)$). Soit $B^*(s) = \mathbb{E}[e^{-sB}]$ la transformée de Laplace de la période d'occupation. Alors $B^*(s)$ satisfait l'équation fonctionnelle :

$$B^*(s) = \tilde{B}^*(s + \lambda - \lambda B^*(s))$$

où $\tilde{B}^*(s)$ est la transformée de Laplace du temps de service.

Démonstration. En utilisant la décomposition de la période d'occupation :

$$B = S + \sum_{i=1}^A B_i,$$

où $S \sim B(t)$, A est le nombre d'arrivées pendant S , et B_i sont i.i.d. de même loi que B , indépendants de S et de A .

Conditionnellement à $S = t$, A suit une loi de Poisson de paramètre λt . Donc :

$$\begin{aligned} \mathbb{E}[e^{-sB} \mid S = t] &= \mathbb{E}\left[e^{-s(t + \sum_{i=1}^A B_i)} \mid S = t\right] \\ &= e^{-st} \sum_{k=0}^{\infty} \mathbb{P}(A = k \mid S = t) (\mathbb{E}[e^{-sB}])^k \\ &= e^{-st} \sum_{k=0}^{\infty} \frac{(\lambda t)^k}{k!} e^{-\lambda t} (B^*(s))^k \\ &= e^{-st} e^{-\lambda t} e^{\lambda t B^*(s)} = e^{-t(s + \lambda - \lambda B^*(s))}. \end{aligned}$$

En intégrant par rapport à la distribution de S :

$$B^*(s) = \int_0^{\infty} e^{-t(s + \lambda - \lambda B^*(s))} dB(t) = \tilde{B}^*(s + \lambda - \lambda B^*(s)).$$

□

2.4.3 Moments de la période d'occupation

Théorème 2.8 (Moyenne et variance de la période d'occupation). *Sous la condition de stabilité $\rho < 1$, la période d'occupation moyenne est donnée par :*

$$\mathbb{E}[B] = \frac{\mathbb{E}[S]}{1 - \rho} = \frac{1/\mu}{1 - \rho}.$$

Le moment d'ordre 2 de la période d'occupation est :

$$\mathbb{E}[B^2] = \frac{\mathbb{E}[S^2]}{(1 - \rho)^3}.$$

Démonstration. Dérivons l'équation fonctionnelle $B^*(s) = \tilde{B}^*(s + \lambda - \lambda B^*(s))$ par rapport à s .

Posons $u = s + \lambda - \lambda B^*(s)$. Alors $B^*(s) = \tilde{B}^*(u)$. En dérivant :

$$\frac{dB^*}{ds} = \frac{d\tilde{B}^*}{du} \cdot \frac{du}{ds} = \frac{d\tilde{B}^*}{du} \left(1 - \lambda \frac{dB^*}{ds}\right).$$

En $s = 0$, $B^*(0) = 1$, et $\frac{dB^*}{ds}(0) = -\mathbb{E}[B]$. Soit $m_1 = -\frac{dB^*}{ds}(0)$ et $\beta_1 = -\frac{d\tilde{B}^*}{ds}(0) = \mathbb{E}[S]$. L'équation devient :

$$m_1 = \beta_1(1 + \lambda m_1).$$

En effet, $\frac{du}{ds} = 1 - \lambda(-m_1) = 1 + \lambda m_1$. Donc :

$$m_1 = \beta_1(1 + \lambda m_1) \quad \Rightarrow \quad m_1(1 - \lambda\beta_1) = \beta_1 \quad \Rightarrow \quad m_1 = \frac{\beta_1}{1 - \lambda\beta_1} = \frac{\mathbb{E}[S]}{1 - \rho}.$$

Pour le moment d'ordre 2, dérivons une seconde fois l'équation. Soit $m_2 = \mathbb{E}[B^2]$. En développant, on obtient :

$$m_2 = \frac{\mathbb{E}[S^2]}{(1 - \rho)^3}.$$

□

2.5 Comparaison mathématique selon le type de service

2.5.1 Cas particuliers

Définition 2.8 (Trois cas particuliers importants). *On considère les trois cas suivants de distribution des temps de service :*

- $M/M/1$: service exponentiel de taux μ . Alors $B(t) = 1 - e^{-\mu t}$.
- $M/D/1$: service déterministe constant $S = 1/\mu$. Alors $B(t) = \mathbf{1}_{\{t \geq 1/\mu\}}$.
- $M/H_2/1$: service hyper-exponentiel d'ordre 2 (mélange de 2 exponentielles).

2.5.2 Comparaison des formules de Pollaczek-Khinchine

Proposition 2.4 (Formules spécifiques). *Pour chacun des trois cas, le temps d'attente moyen est :*

- $M/M/1$: $\sigma^2 = 1/\mu^2$, donc :

$$W_q^{MM1} = \frac{\lambda(1/\mu^2)}{2(1 - \rho)} = \frac{\rho}{\mu(1 - \rho)}.$$

- $M/D/1$: $\sigma^2 = 0$, donc :

$$W_q^{MD1} = \frac{\lambda(1/\mu^2)}{2(1 - \rho)} = \frac{\rho}{2\mu(1 - \rho)} = \frac{1}{2}W_q^{MM1}.$$

Le temps d'attente moyen dans $M/D/1$ est la moitié de celui dans $M/M/1$ pour le même taux d'utilisation ρ .

- $M/H_2/1$ ($k=2$) : la loi hyper-exponentielle à deux branches a une variance supérieure à celle de l'exponentielle. En général :

$$W_q^{MH2/1} = \frac{\rho(1 + \sigma^2\mu^2)}{2\mu(1 - \rho)} > W_q^{MM1} \quad \text{si } \sigma^2 > 1/\mu^2.$$

2.5.3 Impact de la variabilité sur les performances

Théorème 2.9 (Effet de la variabilité). *Pour un taux d'utilisation ρ fixé, le temps d'attente moyen W_q est une fonction croissante de la variance σ^2 des temps de service :*

$$\frac{\partial W_q}{\partial \sigma^2} = \frac{\lambda}{2(1 - \rho)} > 0.$$

Ainsi, à charge égale, plus la distribution des temps de service est variable, plus les temps d'attente sont longs.

Remarque 2.4 (Ordre de convexité). On peut ordonner les distributions selon leur coefficient de variation $C_V = \sigma/\mathbb{E}[S]$:

- $M/D/1$: $C_V = 0$ (service déterministe) \rightarrow temps d'attente minimal,
- $M/M/1$: $C_V = 1$ (service exponentiel) \rightarrow temps d'attente intermédiaire,
- $M/H_2/1$: $C_V > 1$ (service hyper-exponentiel) \rightarrow temps d'attente maximal pour un ρ donné.

2.5.4 Tableau récapitulatif

Modèle	C_V^2	W_q	$\mathbb{E}[B]$	$\mathbb{E}[B^2]$
M/D/1	0	$\frac{\rho}{2\mu(1-\rho)}$	$\frac{1}{\mu(1-\rho)}$	$\frac{1/\mu^2}{(1-\rho)^3}$
M/M/1	1	$\frac{\rho}{\mu(1-\rho)}$	$\frac{1}{\mu(1-\rho)}$	$\frac{2/\mu^2}{(1-\rho)^3}$
M/H ₂ /1	> 1	$> \frac{\rho}{\mu(1-\rho)}$	$\frac{1}{\mu(1-\rho)}$	$> \frac{2/\mu^2}{(1-\rho)^3}$

TABLE 2.1 – Comparaison des mesures de performance pour différents types de service

Exemple 2.2 (Application numérique). Prenons $\lambda = 0.8$ clients par seconde, $\mu = 1$ client par seconde ($\rho = 0.8$), $\mathbb{E}[S] = 1$ seconde.

- $M/D/1$: $\sigma^2 = 0$, $W_q = \frac{0.8}{2 \times 1 \times (1-0.8)} = \frac{0.8}{0.4} = 2$ secondes.
- $M/M/1$: $\sigma^2 = 1$, $W_q = \frac{0.8}{1 \times (1-0.8)} = \frac{0.8}{0.2} = 4$ secondes.
- $M/H_2/1$: $\sigma^2 = 4$ (par exemple), $W_q = \frac{0.8 \times (1+4)}{2 \times 1 \times 0.2} = \frac{0.8 \times 5}{0.4} = \frac{4}{0.4} = 10$ secondes.

2.6 Résumé du chapitre

Ce chapitre a présenté l'analyse exacte du système M/G/1 par la méthode des chaînes de Markov incluses. Les points clés sont :

1. La non-markovianité du processus continu justifie l'introduction des instants de départ comme points de régénération.
2. La chaîne de Markov incluse $\{X_n\}$ a une matrice de transition triangulaire qui permet le calcul explicite.
3. La formule de Pollaczek-Khinchine établit la relation fondamentale entre le temps d'attente moyen et les moments du service :

$$W_q = \frac{\lambda \mathbb{E}[S^2]}{2(1-\rho)}.$$

4. La période d'occupation satisfait une équation de renouvellement et sa moyenne est $\mathbb{E}[B] = \mathbb{E}[S]/(1-\rho)$.

5. La comparaison M/M/1, M/D/1, M/H_k/1 montre l'impact crucial de la variabilité des services sur les performances.

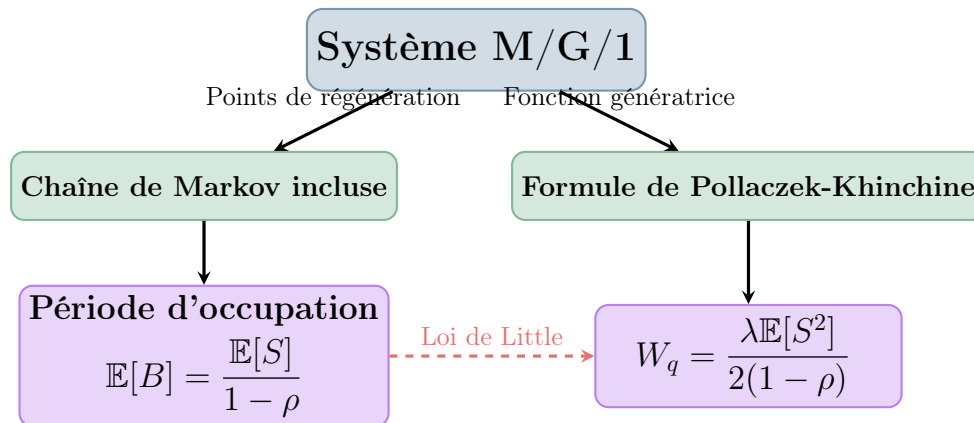


Figure : Synthèse de l'analyse M/G/1

Chapitre 3

File d'attente M/G/1 non fiable avec temps de rappel général, vacances actives et temps de configuration

La théorie des files d'attente est une branche fondamentale des mathématiques appliquées avec des applications répandues dans divers domaines tels que les télécommunications, les systèmes informatiques, les transports et la fabrication. Lorsque les serveurs sont occupés, les clients entrent souvent dans une orbite de rappel, attendant la disponibilité du service selon des politiques prédéfinies.

Les systèmes de files d'attente, en particulier ceux impliquant des rappels avec vacances, jouent un rôle crucial dans l'équilibrage entre l'utilisation des ressources et l'efficacité du service.

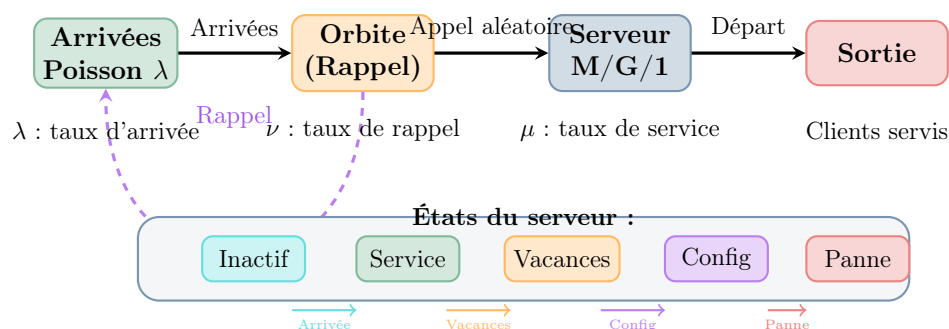


Figure 1 : Schéma général du modèle de file d'attente avec rappel

Définition 3.1 (Vacances actives (Working vacation)). Une **vacance active** est une période pendant laquelle le serveur, bien qu'étant dans un état d'inactivité relative, continue de traiter les clients à un taux de service réduit μ_w (avec $\mu_w < \mu_n$, où μ_n est le taux de service normal). Cette période débute lorsque l'orbite devient vide et sa durée suit une loi exponentielle de paramètre $\omega > 0$. Pendant cette période, les arrivées sont servies immédiatement mais à un rythme plus lent. À la fin de la vacance active, si l'orbite est toujours vide, le serveur s'éteint complètement.

Définition 3.2 (Temps de configuration (Setup time)). Le **temps de configuration** est la durée nécessaire au serveur pour passer de l'état éteint à l'état opérationnel. Lorsqu'un client arrive alors que le serveur est éteint, ce dernier doit subir une phase de configuration

avant de pouvoir le servir. Cette durée est modélisée par une variable aléatoire T_s de fonction de répartition générale $T_s(x)$ et de transformée de Laplace-Stieltjes $T_s^*(\theta)$. Durant cette phase, les clients arrivants rejoignent l'orbite.

Définition 3.3 (Rappel (Retrial)). Le **rappel** est le mécanisme par lequel un client bloqué (c'est-à-dire arrivé alors que le serveur est occupé, en vacances actives, en configuration ou en panne) quitte temporairement le système et entre dans une **orbite**. Depuis cette orbite, il effectue des tentatives aléatoires pour accéder au serveur. Les temps inter-rappel sont indépendants et identiquement distribués selon une loi générale $N(x)$ de moyenne $1/\nu$ et de transformée de Laplace-Stieltjes $N^*(\theta)$.

Définition 3.4 (Serveur non fiable (Unreliable server)). Le serveur est dit **non fiable** car il peut tomber en panne pendant la phase de configuration. Lorsque le serveur entre en configuration, cette opération réussit avec probabilité α ($0 < \alpha < 1$) et échoue avec probabilité $\bar{\alpha} = 1 - \alpha$. En cas d'échec, le serveur est envoyé en réparation. La durée de réparation est une variable aléatoire S_f de fonction de répartition générale $S_f(x)$ et de transformée de Laplace-Stieltjes $S_f^*(\theta)$. Pendant la réparation, tout client arrivant rejoint l'orbite.

3.1 Description du modèle

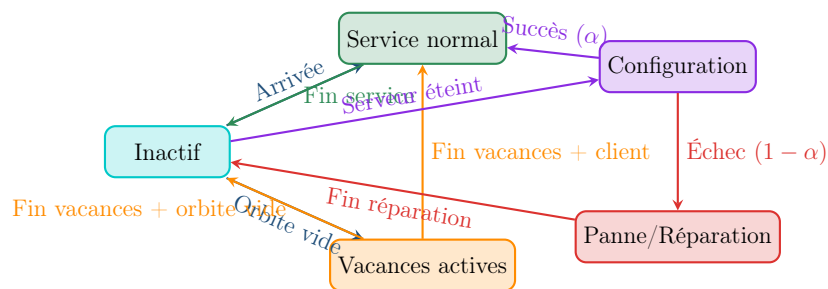


Figure 2 : Diagramme de transition des états du serveur

Dans cette partie, nous considérons une file d'attente M/G/1 non fiable avec rappel, vacances actives unique et temps de configuration. Le modèle est décrit en détail comme suit :

1. L'arrivée des consommateurs suit un processus de Poisson avec un taux $\lambda > 0$. Si un consommateur arrive et que le serveur est inactif, il commence le service immédiatement. D'autre part, s'il le trouve occupé, en vacances actives, en configuration ou en panne, alors le consommateur quitte la zone de service et entre dans un pool de consommateurs bloqués appelé orbite selon une discipline Premier Arrivé Premier Servi. Pendant qu'il est en orbite, le consommateur attend un temps aléatoire avant de réessayer. Les temps inter-rappel suivent une fonction de distribution de probabilité arbitraire $N(x)$ avec sa fonction LST correspondante $N^*(\theta)$.
2. Le serveur initie promptement le service normal pour les nouveaux consommateurs ou ceux en rappel lors de leur arrivée alors qu'il est inactif. Le temps de service normal est distribué selon une fonction de distribution générale $G(x)$ ayant une LST $G^*(\theta)$.

3. Dans le cas où l'orbite est vide, le serveur commence automatiquement une seule vacance active, qui suit une distribution exponentielle avec un taux ω . Si des consommateurs arrivent pendant la vacance active, ils sont servis à un taux réduit. Lorsque la vacance active se termine, le serveur reprend le service normal. Pendant la période de vacances actives, le temps de service est distribué selon une fonction de distribution générale $W_v(x)$ ayant une LST $W_v^*(\theta)$.
4. À la fin de la vacance active, si aucun consommateur n'attend son tour, le serveur est éteint directement pour économiser l'énergie.
5. Dans l'état éteint du serveur, si un client arrive, il attendra son tour en position avancée devant le serveur jusqu'à ce qu'il soit activé (temps de configuration). Le temps de configuration est supposé suivre une distribution générale avec une fonction de distribution de probabilité $T_s(x)$ et sa fonction LST correspondante $T_s^*(\theta)$. Les clients arrivant dans l'état de configuration doivent rejoindre l'orbite.
6. Lorsque le serveur passe à l'état de configuration, cette opération peut échouer avec une probabilité $\bar{\alpha} = (1 - \alpha)$. Ensuite, le serveur est envoyé en réparation et le temps de réparation du serveur a une fonction de distribution de probabilité arbitraire $S_f(x)$ avec sa fonction LST correspondante $S_f^*(\theta)$.

3.2 Analyse à l'état stationnaire

Soient $\chi_1(t), \chi_2(t), \chi_3(t), \chi_4(t), \chi_5(t)$ les temps écoulés respectivement dans le rappel, le service régulier, les vacances actives, la réparation et la configuration au temps t .

Supposons que :

$$N(0) = 0, N(\infty) = 1, G(0) = 0, G(\infty) = 1, W_v(0) = 0, W_v(\infty) = 1,$$

$$T_s(0) = 0, T_s(\infty) = 1, S_f(0) = 0, S_f(\infty) = 1$$

sont continues en $x = 0$.

Par conséquent, nous spécifions les fonctions de taux de hasard $f(x), \mu_n(x), \mu_w(x), v(x), \delta(x)$, respectivement pour le rappel, le service normal, le service à taux réduit, la réparation et la configuration :

$$f(x)dx = \frac{dN(x)}{1 - N(x)}, \quad \mu_n(x)dx = \frac{dG(x)}{1 - G(x)}, \quad \mu_w(x)dx = \frac{dW_v(x)}{1 - W_v(x)}$$

$$v(x)dx = \frac{dS_f(x)}{1 - S_f(x)}, \quad \delta(x)dx = \frac{dT_s(x)}{1 - T_s(x)}$$

3.2.1 Description des états du système

L'état du système au temps t peut être défini par le processus de Markov $\{N(t); t \geq 0\} = \{D(t), X(t), \chi_1(t), \chi_2(t), \chi_3(t), \chi_4(t), \chi_5(t) \geq 0\}$, où $X(t)$ désigne le nombre de clients dans l'orbite au temps t et

$$D(t) = \begin{cases} 0, & \text{si le serveur est inactif en période normale} \\ 1, & \text{si le serveur est inactif en période de vacances actives} \\ 2, & \text{si le serveur est occupé en service normal} \\ 3, & \text{si le serveur est occupé en période de vacances actives} \\ 4, & \text{si le serveur est en réparation} \\ 5, & \text{si le serveur est en état de configuration} \end{cases}$$

respectivement. Si $D(t) = 0$ et $X(t) > 0$, alors $\chi_1(t)$ représente le temps de rappel écoulé, si $D(t) = 2$, alors $\chi_2(t)$ représente le temps de service écoulé pendant la période normale de service, si $D(t) = 3$ et $X(t) \geq 0$ alors $\chi_3(t)$ représente le temps de vacances actives écoulé au temps t , et si $D(t) = 4$ et $X(t) \geq 1$ alors $\chi_4(t)$ représente le temps de réparation écoulé au temps t , et si $D(t) = 5$ et $X(t) \geq 1$ alors $\chi_5(t)$ représente le temps de configuration écoulé au temps t .

État	Description
$D(t) = 0$	Serveur inactif (période normale)
$D(t) = 1$	Serveur inactif (vacances actives)
$D(t) = 2$	Serveur occupé (service normal)
$D(t) = 3$	Serveur occupé (vacances actives)
$D(t) = 4$	Serveur en réparation
$D(t) = 5$	Serveur en configuration

Tableau 1 : Description des états du système

3.2.2 Condition de stabilité et d'ergodicité

Soit $\{t_n; n \in \mathbb{N}\}$ la séquence des époques de fins de service ou de terminaison de vacances. La séquence de vecteurs aléatoires $Z_n = \{D(t_n^+), X(t_n^+)\}$ forme une chaîne de Markov qui est la chaîne de Markov imbriquée pour notre système de file d'attente. Son espace d'états est $S = \{0, 1, 2, 3, 4, 5\} \times \mathbb{N}$.

Théorème 3.1 (Condition de stabilité). *La chaîne de Markov imbriquée $\{Z_n; n \in \mathbb{N}\}$ est ergodique si et seulement si :*

$$\rho = \lambda \left(\frac{1}{\mu_n} + \frac{1}{\mu_w} + \omega \mathcal{W}_v^*(\omega) \right) < \mathcal{N}^*(\lambda)$$

3.2.3 Équations régissant le système

Pour le processus de Markov $\{N(t); t \geq 0\}$, nous définissons la probabilité :

$$P_{00}(t) = \mathbb{P}\{D(t) = 0, X(t) = 0\}, \quad Q_0(t) = \mathbb{P}\{D(t) = 1, X(t) = 0\}$$

et les densités de probabilité :

$$\begin{aligned} P_n(x, t)dx &= \mathbb{P}\{D(t) = 0, X(t) = n, x \leq \chi_1(t) < x + dx\}, & x \geq 0, n \geq 1 \\ M_{n,b}(x, t)dx &= \mathbb{P}\{D(t) = 2, X(t) = n, x \leq \chi_2(t) < x + dx\}, & x \geq 0, n \geq 0 \\ G_{n,v}(x, t)dx &= \mathbb{P}\{D(t) = 3, X(t) = n, x \leq \chi_3(t) < x + dx\}, & x \geq 0, n \geq 0 \\ U_n(x, t)dx &= \mathbb{P}\{D(t) = 4, X(t) = n, x \leq \chi_4(t) < x + dx\}, & x \geq 0, n \geq 1 \\ K_n(x, t)dx &= \mathbb{P}\{D(t) = 5, X(t) = n, x \leq \chi_5(t) < x + dx\}, & x \geq 0, n \geq 1 \end{aligned}$$

Nous supposons que la condition de stabilité est satisfaite dans la suite et ainsi nous pouvons poser :

$$P_{00} = \lim_{t \rightarrow \infty} P_{00}(t), \quad Q_0 = \lim_{t \rightarrow \infty} Q_0(t)$$

et les densités limites pour $x > 0$ et $n \geq 0$:

$$\begin{aligned} M_{n,b}(x) &= \lim_{t \rightarrow \infty} M_{n,b}(x, t), & P_n(x) &= \lim_{t \rightarrow \infty} P_n(x, t), & G_{n,v}(x) &= \lim_{t \rightarrow \infty} G_{n,v}(x, t) \\ \text{et } U_n(x) &= \lim_{t \rightarrow \infty} U_n(x, t) & \text{et } K_n(x) &= \lim_{t \rightarrow \infty} K_n(x, t). \end{aligned}$$

Basé sur les hypothèses et notations ci-dessus, notre modèle est régi par l'ensemble d'équations différentielles aux différences suivant :

$$\lambda P_{00} = \omega Q_0 \tag{1}$$

$$(\lambda + \omega)Q_0 = \int_0^\infty M_{0,b}(x)\mu_n(x)dx + \int_0^\infty G_{0,v}(x)\mu_w(x)dx \tag{2}$$

$$\frac{d}{dx}P_n(x) + (\lambda + f(x))P_n(x) = 0, \quad x > 0, \quad n \geq 1 \tag{3}$$

$$\frac{d}{dx}M_{0,b}(x) + (\lambda + \mu_n(x))M_{0,b}(x) = 0, \quad x > 0 \tag{4}$$

$$\frac{d}{dx}M_{n,b}(x) + (\lambda + \mu_n(x))M_{n,b}(x) = \lambda M_{n-1,b}(x), \quad x > 0, \quad n \geq 1 \tag{5}$$

$$\frac{d}{dx}G_{0,v}(x) + (\lambda + \omega + \mu_w(x))G_{0,v}(x) = 0, \quad x > 0 \tag{6}$$

$$\frac{d}{dx}G_{n,v}(x) + (\lambda + \omega + \mu_w(x))G_{n,v}(x) = \lambda G_{n-1,v}(x), \quad x > 0, \quad n \geq 1 \tag{7}$$

$$\frac{d}{dx}U_0(x) + (\lambda + v(x))U_0(x) = 0, \quad x > 0 \tag{8}$$

$$\frac{d}{dx}U_n(x) + (\lambda + v(x))U_n(x) = \lambda U_{n-1}(x), \quad x > 0, \quad n \geq 1 \tag{9}$$

$$\frac{d}{dx}K_0(x) + (\lambda + \delta(x))K_0(x) = 0, \quad x > 0 \tag{10}$$

$$\frac{d}{dx}K_n(x) + (\lambda + \delta(x))K_n(x) = \lambda K_{n-1}(x), \quad x > 0, \quad n \geq 1 \tag{11}$$

Les conditions aux limites en $x = 0$ incluent $P_n(0)$, $M_{0,b}(0)$, $M_{n,b}(0)$, $G_{0,v}(0)$, $U_0(0)$,

$U_n(0)$, $K_0(0)$. La description de ces termes en $x = 0$ est la suivante :

$$P_n(0) = \int_0^\infty G_{n,v}(x)\mu_w(x)dx + \int_0^\infty M_{n,b}(x)\mu_n(x)dx, \quad n \geq 1 \quad (12)$$

$$M_{0,b}(0) = \alpha \int_0^\infty K_0(x)\delta(x)dx + \int_0^\infty P_1(x)f(x)dx + \omega \int_0^\infty G_{0,v}(x)dx + \int_0^\infty U_0(x)v(x)dx \quad (13)$$

$$M_{n,b}(0) = \alpha \int_0^\infty K_n(x)\delta(x)dx + \int_0^\infty P_{n+1}(x)f(x)dx + \omega \int_0^\infty G_{n,v}(x)dx \\ + \lambda \int_0^\infty P_n(x)dx + \int_0^\infty U_n(x)v(x)dx, \quad n \geq 1 \quad (14)$$

$$G_{0,v}(0) = \lambda Q_0 \quad (15)$$

$$G_{n,v}(0) = 0, \quad n \geq 1 \quad (16)$$

$$K_0(0) = \lambda P_{00} \quad (17)$$

$$K_n(0) = 0, \quad n \geq 1 \quad (18)$$

$$U_0(0) = \bar{\alpha} \int_0^\infty K_0(x)\delta(x)dx \quad (19)$$

$$U_n(0) = \bar{\alpha} \int_0^\infty K_n(x)\delta(x)dx, \quad n \geq 1 \quad (20)$$

La condition de normalisation est donnée par :

$$P_{00} + Q_0 + \sum_{n=1}^{\infty} \int_0^\infty P_n(x)dx + \sum_{n=0}^{\infty} \int_0^\infty M_{n,b}(x)dx + \sum_{n=0}^{\infty} \int_0^\infty G_{n,v}(x)dx \\ + \sum_{n=0}^{\infty} \int_0^\infty U_n(x)dx + \sum_{n=0}^{\infty} \int_0^\infty K_n(x)dx = 1$$

Afin de résoudre l'ensemble d'équations ci-dessus, nous définissons les fonctions géné-

matrices comme suit :

$$\begin{aligned}
P(x, z) &= \sum_{n=1}^{\infty} z^n P_n(x) \quad \text{pour } |z| \leq 1 \text{ et } x > 0, \\
P(0, z) &= \sum_{n=1}^{\infty} z^n P_n(0) \quad \text{pour } |z| \leq 1, \\
M_b(x, z) &= \sum_{n=0}^{\infty} z^n M_{n,b}(x) \quad \text{pour } |z| \leq 1 \text{ et } x > 0, \\
M_b(0, z) &= \sum_{n=0}^{\infty} z^n M_{n,b}(0), \\
G_v(x, z) &= \sum_{n=0}^{\infty} z^n G_{n,v}(x) \quad \text{pour } |z| \leq 1 \text{ et } x > 0, \\
G_v(0, z) &= \sum_{n=0}^{\infty} z^n G_{n,v}(0), \\
U(x, z) &= \sum_{n=0}^{\infty} z^n U_n(x) \quad \text{pour } |z| \leq 1, \\
U(0, z) &= \sum_{n=0}^{\infty} z^n U_n(0) \quad \text{pour } |z| \leq 1, \\
K(x, z) &= \sum_{n=0}^{\infty} z^n K_n(x) \quad \text{pour } |z| \leq 1, \\
K(0, z) &= \sum_{n=0}^{\infty} z^n K_n(0) \quad \text{pour } |z| \leq 1.
\end{aligned}$$

En multipliant les équations par les puissances appropriées de z et en sommant sur n , nous obtenons l'ensemble d'équations aux dérivées partielles suivant :

$$\frac{\partial P(x, z)}{\partial x} + (\lambda + f(x))P(x, z) = 0 \quad (21)$$

$$\frac{\partial M_b(x, z)}{\partial x} + (\lambda - \lambda z + \mu_n(x))M_b(x, z) = 0 \quad (22)$$

$$\frac{\partial G_v(x, z)}{\partial x} + (\lambda - \lambda z + \omega + \mu_w(x))G_v(x, z) = 0 \quad (23)$$

$$\frac{\partial U(x, z)}{\partial x} + (\lambda - \lambda z + v(x))U(x, z) = 0 \quad (24)$$

$$\frac{\partial K(x, z)}{\partial x} + (\lambda - \lambda z + \delta(x))K(x, z) = 0 \quad (25)$$

En résolvant les équations aux dérivées partielles (22) à (26), nous obtenons :

$$P(x, z) = P(0, z)[1 - N(x)]e^{-\lambda x} \quad (26)$$

$$M_b(x, z) = M_b(0, z)[1 - G(x)]e^{-I(z)x} \quad (27)$$

$$G_v(x, z) = G_v(0, z)[1 - W_v(x)]e^{-(I(z)+\eta)x} \quad (28)$$

$$U(x, z) = U(0, z)[1 - S_f(x)]e^{-I(z)x} \quad (29)$$

$$K(x, z) = K(0, z)[1 - T_s(x)]e^{-I(z)x} \quad (30)$$

où $I(z) = \lambda(1 - z)$.

En multipliant l'équation (12) par les puissances appropriées de z , en sommant sur n de 1 à ∞ et en utilisant les équations (1) et (2) après quelques manipulations algébriques, nous obtenons :

$$P(0, z) = \int_0^\infty G_v(x, z)\mu_w(x)dx + \int_0^\infty M_b(x, z)\mu_n(x)dx - \left(\frac{\lambda + \omega}{\omega}\right) \lambda P_{00} \quad (31)$$

De même, en multipliant les équations (14)-(21) par les puissances appropriées de z , en sommant sur n et après quelques manipulations algébriques, nous obtenons :

$$\begin{aligned} M_b(0, z) &= \frac{\alpha}{z} \int_0^\infty K(x, z)\delta(x)dx + \lambda \int_0^\infty P(x, z)dx + \frac{1}{z} \int_0^\infty P(x, z)f(x)dx \\ &\quad + \omega \int_0^\infty G_v(x, z)dx + \int_0^\infty U(x, z)v(x)dx \end{aligned} \quad (32)$$

$$G_v(0, z) = G_{0,v}(0) \quad (33)$$

$$U(0, z) = \bar{\alpha} \int_0^\infty K(x, z)\delta(x)dx \quad (34)$$

$$K(0, z) = K_0(0) \quad (35)$$

En insérant les équations (28)-(29) dans (32), nous obtenons :

$$P(0, z) = G_v(0, z)W_v^*(I(z) + \eta) + M_b(0, z)G^*(I(z)) - \left(\frac{\lambda + \omega}{\omega}\right) \lambda P_{00} \quad (36)$$

De manière similaire, en insérant les équations (26)-(30) dans (32), nous obtenons :

$$\begin{aligned} M_b(0, z) &= \frac{1}{z} P(0, z)N^*(\lambda) + \alpha T_s^*(I(z))K(0, z) + V(z)G_v(0, z) \\ &\quad + P(0, z)(1 - N^*(\lambda)) + U(0, z)S_f^*(I(z)) \end{aligned} \quad (37)$$

où $V(z) = \frac{\omega}{I(z)+\omega}(1 - W_v^*(I(z) + \omega))$.

En utilisant les équations (16) et (1), nous obtenons :

$$G_v(0, z) = \frac{\lambda^2}{\omega} P_{00} \quad (38)$$

En utilisant l'équation (18), nous obtenons :

$$K(0, z) = \lambda P_{00} \quad (39)$$

En utilisant l'équation (31) dans (35), nous obtenons :

$$U(0, z) = \bar{\alpha} T_s^*(I(z))K(0, z) \quad (40)$$

En utilisant l'équation (40) dans (41), nous obtenons :

$$U(0, z) = \bar{\alpha} \lambda P_{00} T_s^*(I(z)) \quad (41)$$

En utilisant les équations (39), (40) et (42) dans (37) et (38), nous obtenons

$$P(0, z) = \frac{N_r(z)}{D_r(z)} \quad (42)$$

où

$$\begin{aligned}
N_r(z) &= \lambda P_{00} z G^*(I(z)) T_s^*(I(z)) [(\alpha - 1) \omega S_f^*(I(z)) - \alpha \omega] + \omega \\
&\quad - \lambda G^*(I(z)) V(z) - \lambda (W_v^*(I(z) + \omega) - 1) \\
D_r(z) &= \omega G^*(I(z)) [z(1 - N^*(\lambda)) + N^*(\lambda)] - z \\
M_b(0, z) &= \frac{-\lambda P_{00}}{D_r(z)} \left\{ \begin{array}{l} \omega z (-\alpha S_f^*(I(z)) + \alpha + S_f^*(I(z))) T_s^*(I(z)) \\ + \lambda z V(z) + z(N^*(\lambda) - 1)(\omega - \lambda W_v^*(I(z) + \omega) + \lambda) \\ - (\omega - \lambda W_v^*(I(z) + \omega) + \lambda) N^*(\lambda) \end{array} \right\} \quad (43)
\end{aligned}$$

En substituant les équations (39-44) dans (27-31).

3.3 Résultats en régime permanent

Si le système est en régime permanent avec $\rho < N^*(\lambda)$, les fonctions génératrices sont les suivantes :

(I) Le nombre de clients dans l'orbite lorsque le serveur est inactif :

$$P(z) = \int_0^\infty P(x, z) dx = \frac{P(0, z)(1 - N^*(\lambda))}{\lambda} \quad (44)$$

(II) Le nombre de clients dans l'orbite lorsque le serveur est occupé en service normal :

$$M_b(z) = \int_0^\infty M_b(x, z) dx = \frac{M_b(0, z)(1 - G^*(I(z)))}{I(z)} \quad (45)$$

(III) Le nombre de clients dans l'orbite lorsque le serveur est en service à vitesse réduite :

$$G_v(z) = \int_0^\infty G_v(x, z) dx = \frac{G_v(0, z)(1 - W_v^*(I(z) + \omega))}{I(z) + \omega} \quad (46)$$

(IV) Le nombre de clients dans l'orbite lorsque le serveur est en réparation :

$$U(z) = \int_0^\infty U(x, z) dx = \frac{U(0, z)(1 - S_f^*(I(z)))}{I(z)} \quad (47)$$

(V) Le nombre de clients dans l'orbite lorsque le serveur est en configuration :

$$K(z) = \int_0^\infty K(x, z) dx = \frac{K(0, z)(1 - T_s^*(I(z)))}{I(z)} \quad (48)$$

À partir des équations ci-dessus, la seule inconnue est P_{00} qui peut être obtenue en utilisant la condition de normalisation $P_{00} + Q_0 + P(1) + M_b(1) + G_v(1) + U(1) + K(1) = 1$:

$$P_{00} = \frac{\omega^2(N^*(\lambda) - \lambda E(G))}{\left\{ \begin{array}{l} \omega^2 (\lambda (E(S_f)(1 - \alpha) + E(T_s)) + N^*(\lambda)) \\ + \omega (\lambda^2 \{2W_v^{*'}(\omega)(\lambda E(N_b) - N^*(\lambda) + 1) - E(N_b)W_v^*(\omega)\}) \\ + \lambda N^*(\lambda) + \lambda^2(1 - W_v^*(\omega)) \end{array} \right\}} \quad (49)$$

Corollaire 1. Si le système satisfait la condition de régime permanent, la fonction génératrice du nombre de clients dans le système ($K_s(z)$) est obtenue par :

$$K_s(z) = P_{00} + Q_0 + P(z) + zM_b(z) + zG_v(z) + U(z) + K(z) \quad (50)$$

La fonction génératrice du nombre de clients dans l'orbite ($K_o(z)$) est obtenue par :

$$K_o(z) = P_{00} + Q_0 + P(z) + M_b(z) + G_v(z) + U(z) + K(z) \quad (51)$$

où $P(z)$, $M_b(z)$, $G_v(z)$, $U(z)$ et $K(z)$ sont définies dans les résultats en régime permanent (44)-(48).

3.4 Mesures de performance du système

3.4.1 Probabilités des états du système

1. La probabilité à l'état stationnaire que le serveur soit inactif pendant le temps de rappel est donnée par :

$$I = P(1) = \frac{-\lambda P_{00}(\mathcal{N}^*(\lambda) - 1)}{\omega^2(\mathcal{N}^*(\lambda) - \lambda E(N_b))} \left\{ \begin{array}{l} \omega(-E(N_b)\lambda(\mathcal{W}_v^*(\omega) - 1) + \omega(-\alpha E(S_f) + E(N_b) + E(T_s))) \\ + \omega E(S_f) + \lambda \omega \mathcal{W}_v^*(\omega) + \lambda(\omega \mathcal{W}_v^*(\omega) - \mathcal{W}_v^*(\omega) + 1) \end{array} \right\}$$

2. La probabilité à l'état stationnaire que le serveur soit occupé en période de service normal est donnée par :

$$U = M_b(1) = \frac{\lambda P_{00} E(N_b)}{\omega^2(\mathcal{N}^*(\lambda) - \lambda E(N_b))} \left\{ \begin{array}{l} \omega^2(-\alpha E(S_f) + E(T_s) + E(S_f) + \mathcal{N}^*(\lambda)) \\ + \omega(2\lambda^2 \mathcal{W}_v^*(\omega) - \lambda \mathcal{W}_v^*(\omega) \mathcal{N}^*(\lambda) + \lambda \mathcal{N}^*(\lambda)) \\ - \lambda^2 \mathcal{W}_v^*(\omega) + \lambda^2 \end{array} \right\}$$

3. La probabilité à l'état stationnaire que le serveur soit occupé en période de vacances actives est donnée par :

$$V = G_v(1) = \frac{\lambda^2 P_{00} \cdot (1 - W_v^*(\omega))}{\omega^2}$$

4. La probabilité à l'état stationnaire que le serveur soit en période de réparation est donnée par :

$$S = U(1) = E(S_f) \lambda P_{00} \cdot (1 - \alpha)$$

5. La probabilité à l'état stationnaire que le serveur soit en période de configuration est donnée par :

$$K = K(1) = E(T_s) \lambda P_{00}$$

3.4.2 Taille moyenne du système et de l'orbite

1. Le nombre attendu de clients dans l'orbite (L_q) est obtenu en différenciant l'équation (52) par rapport à z et en évaluant à $z = 1$:

$$L_q = K'_o(1) = \lim_{z \rightarrow 1} \frac{d}{dz} K_o(z)$$

2. Le nombre attendu de clients dans le système (L_s) est obtenu en différenciant l'équation (51) par rapport à z et en évaluant à $z = 1$:

$$L_s = K'_s(1) = \lim_{z \rightarrow 1} \frac{d}{dz} K_s(z)$$

3. Le temps moyen qu'un client passe dans le système (W_s) et le temps moyen qu'un client passe dans la file d'attente (W_q) sont trouvés en utilisant la formule de Little :

$$W_s = \frac{L_s}{\lambda} \quad \text{et} \quad W_q = \frac{L_q}{\lambda}.$$

3.5 Mesures de fiabilité

Dans le système de file d'attente avec rappel et serveur non fiable, les mesures de fiabilité fournissent les informations nécessaires à l'amélioration du système.

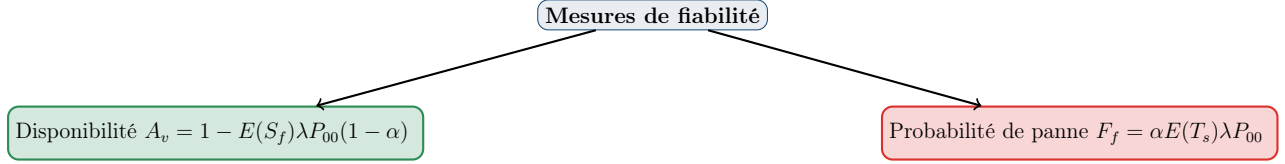


Figure 4 : Mesures de fiabilité du système

1. La disponibilité à l'état stationnaire A_v , qui est la probabilité que le serveur soit soit en train de travailler pour un client positif soit en période d'inactivité, est donnée par :

$$A_v = 1 - U(1) = 1 - E(S_f)\lambda P_{00} \cdot (1 - \alpha) \quad (52)$$

2. Soit F_f la probabilité à l'état stationnaire de panne du serveur :

$$F_f = \alpha K(1) = \alpha E(T_s)\lambda P_{00}$$

Soient $E(T_b)$ et $E(T_c)$ la longueur attendue de la période occupée et du cycle occupé sous les conditions d'état stationnaire, nous avons :

$$E(T_b) = \frac{1}{\omega^2(A^*(\lambda) - \lambda E(N_b))} \begin{cases} \omega^2(-\alpha E(S_f) + E(N_b) + E(T_s) + E(S_f)) \\ \omega(-E(N_b)\lambda W_v^*(w) + 2\lambda W_v'(w)(E(N_b)\lambda - N^*(w) + 1)) \\ +\lambda^2\omega E(N_b)W_v^*(w) + \lambda N^*(w) + \lambda^2(1 - W_v^*(w)) \end{cases} \quad (53)$$

$$E(T_c) = \frac{1}{\omega^2\lambda(N^*(w) - \lambda E(N_b))} \begin{cases} \omega^2(\lambda(E(S_f)(1 - \alpha) + E(T_s)) + N^*(w)) \\ +\omega(\lambda^2\{2W_v'(w)(\lambda E(N_b) - N^*(w) + 1)\}) \\ -\lambda^2\omega E(N_b)W_v^*(w) + \lambda N^*(w) + \lambda^2(1 - W_v^*(w)) \end{cases} \quad (54)$$

Démonstration. Le résultat découle directement de l'argument d'un processus de renouvellement alterné qui conduit à :

$$P_{00} = \frac{E(T_0)}{E(T_b) + E(T_0)}; \quad E(T_b) = \frac{1}{\lambda} \left(\frac{1}{P_{00}} - 1 \right); \quad E(T_c) = E(T_0) + E(T_b) \quad (55)$$

où T_0 est la durée pendant laquelle le système est vide. Puisque le temps inter-arrivée entre deux clients suit une distribution exponentielle de paramètre λ , nous avons $E(T_0) = 1/\lambda$. En insérant l'équation (49) dans (55) et par calculs directs, nous obtenons (53) et (54). \square

3.6 Périodes d'occupation et de cycle

Soient $E(T_b)$ et $E(T_c)$ la longueur attendue de la période occupée et du cycle occupé sous les conditions d'état stationnaire. En appliquant l'argument d'un processus de renouvellement alterné :

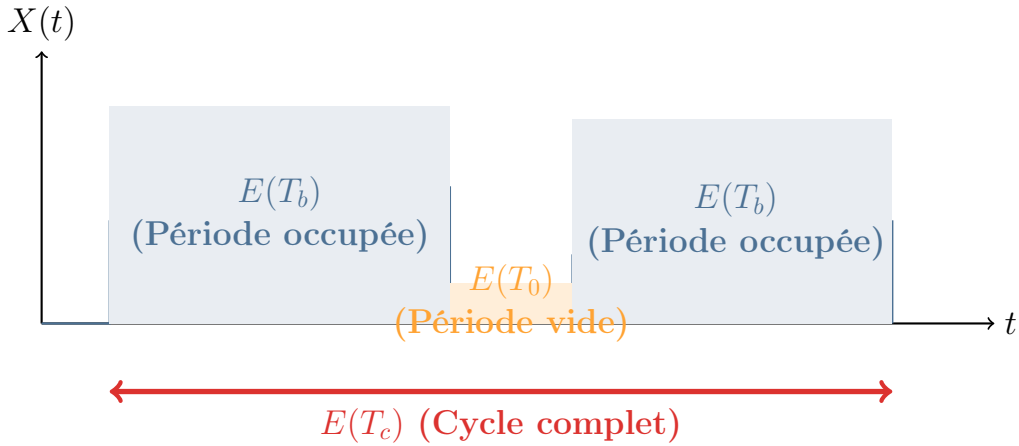


Figure 5 : Périodes d'occupation et de cycle

Remarque 3.1. Les expressions explicites de $E(T_b)$ et $E(T_c)$ sont données respectivement par les équations (53) et (54) dans la section des mesures de fiabilité.

3.6.1 Application pratique : Centre de tri de déchets recyclables

Afin d'illustrer concrètement le modèle mathématique présenté dans ce chapitre, nous considérons un centre de tri de déchets recyclables. Ce système présente une analogie directe avec notre modèle de file d'attente M/G/1 non fiable avec rappel, vacances actives et temps de configuration.



Figure : Application au centre de tri de déchets recyclables

Correspondance entre le modèle mathématique et le système réel

1. Arrivée des cargaisons de déchets (loi de Poisson)

Les cargaisons de déchets arrivent au centre de tri. Chaque cargaison est transportée par un camion, mais seule la cargaison est modélisée comme client.

- **Loi** : processus de Poisson de taux λ
- **Paramètre** : $\lambda = 6$ cargaisons par heure
- **Rôle dans le modèle** : c'est l'arrivée des clients dans le système

Élément du modèle	Élément réel	Paramètres
Client	Cargaison de déchets	Une cargaison = un camion
Serveur	Machine de tri	Une seule machine
Orbite	Zone de stockage des déchets	Capacité illimitée
Service normal	Tri à pleine vitesse	Loi générale $G(x)$
Vacances actives	Machine tournant à vitesse réduite	Loi exponentielle $\mathcal{E}(\omega)$
Service ralenti	Tri à vitesse réduite	Loi générale $W_v(x)$
Configuration	Redémarrage après extinction	Loi générale $T_s(x)$
Panne	Défaillance lors du redémarrage	Probabilité d'échec $1 - \alpha$
Réparation	Intervention technique	Loi générale $S_f(x)$
Rappel	Nouvel essai depuis le stockage	Loi générale $N(x)$

TABLE 3.1 – Correspondance entre le modèle mathématique et le centre de tri

2. La machine de tri (le serveur)

Une machine unique traite les cargaisons une par une.

— **Rôle** : c'est le serveur du modèle M/G/1

États possibles (comme dans le document) :

Élément du modèle	Élément réel	Loi / Paramètres
Client	Cargaison de déchets	Une cargaison = un camion
Serveur	Machine de tri	Une seule machine
Orbite	Zone de stockage des déchets	Capacité illimitée
Service normal	Tri à pleine vitesse	Loi générale $G(x)$
Vacances actives	Machine tournant à vitesse réduite	Loi exponentielle de taux ω
Service ralenti	Tri à vitesse réduite	Loi générale $W_v(x)$
Configuration	Redémarrage après extinction	Loi générale $T_s(x)$
Panne	Défaillance lors du redémarrage	Probabilité d'échec $1 - \alpha$
Réparation	Intervention technique	Loi générale $S_f(x)$
Rappel	Nouvel essai depuis le stockage	Loi générale $N(x)$

TABLE 3.2 – Correspondance entre le modèle mathématique et le centre de tri (distribution de Weibull)

— *Paramètres* : $k = 1,5$ (paramètre de forme), $\beta = 8$ minutes (paramètre d'échelle)

— *Transformée de Laplace-Stieltjes* : $N^*(\theta)$

— *Fonction de taux de hasard* : $f(x)$ utilisée dans l'équation (3)

5. Service normal (état 2)

Quand la machine est en service normal, elle trie une cargaison à pleine vitesse.

— **Loi** : générale $G(x)$ (loi gamma)

— **Paramètres** : $\alpha = 2$ (forme), $\beta = 0,125$ par minute \rightarrow moyenne = 16 minutes

— **Transformée** : $G^*(\theta)$

— **Taux de hasard** : $\mu_n(x)$ utilisé dans les équations (4), (5), (12)

6. Vacances actives (état 1 et état 3)

Quand le système est vide (aucune cargaison dans la machine ni dans l'orbite), la machine entre en **vacances actives** :

- Elle tourne à vitesse réduite (économie d'énergie)
- La durée des vacances actives est exponentielle de taux ω
- **Loi** : exponentielle
- **Paramètre** : $\omega = 1/20$ par minute (moyenne 20 minutes)

Service pendant les vacances actives (état 3)

Une cargaison arrivant pendant une vacance active est servie immédiatement, mais à vitesse réduite.

- **Loi** : générale $W_v(x)$ (loi log-normale)
- **Paramètres** : moyenne = 35 minutes, écart-type = 8 minutes
- **Transformée** : $W_v^*(\theta)$
- **Taux de hasard** : $\mu_w(x)$ utilisé dans les équations (6), (7), (12)

7. Extinction complète de la machine (fin des vacances)

Quand une vacance active se termine et que l'orbite est toujours vide, la machine s'éteint complètement.

- **Objectif** : économie d'énergie maximale

8. Temps de configuration (état 5)

Si une cargaison arrive alors que la machine est éteinte :

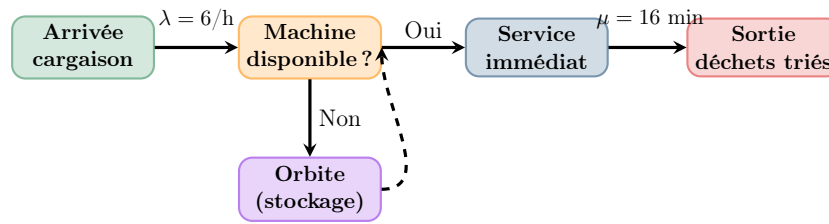
- La cargaison va dans l'orbite
- La machine doit se configurer avant de pouvoir servir
- **Loi** : générale $T_s(x)$ (loi exponentielle)
- **Paramètre** : $\gamma = 1/10$ par minute (moyenne 10 minutes)
- **Transformée** : $T_s^*(\theta)$
- **Taux de hasard** : $\delta(x)$ utilisé dans les équations (10), (11), (13)-(16)

9. Panne pendant la configuration (état 4)

La configuration peut échouer.

- **Probabilité de succès** : α
- **Probabilité d'échec** : $1 - \alpha$
- En cas d'échec, la machine part en réparation
- **Loi du temps de réparation** : générale $S_f(x)$ (loi exponentielle)
- **Paramètre** : $\nu = 1/25$ par minute (moyenne 25 minutes)
- **Transformée** : $S_f^*(\theta)$
- **Taux de hasard** : $v(x)$ utilisé dans les équations (8), (9), (21), (22)
- **Paramètre** : $\alpha = 0,93$ (93% de succès)

10. Synthèse du cycle de vie d'une cargaison



Rappel : loi de Weibull $k = 1,5$, $\beta = 8$ min

11. Résumé des paramètres pour l'étude numérique

Symbole	Signification	Loi	Paramètres
λ	Taux d'arrivée	Poisson	6 cargaisons/h
ω	Taux de vacances actives	Exponentielle	1/20 par min
μ	Taux de service normal	Gamma	$\alpha = 2$, $\beta = 0,125$ /min
μ_w	Taux de service ralenti	Log-normale	Moyenne=35 min, $\sigma = 8$ min
γ	Taux de configuration	Exponentielle	1/10 par min
ν	Taux de réparation	Exponentielle	1/25 par min
α	Probabilité de succès	Bernoulli	0,93
k, β	Rappel	Weibull	$k = 1,5$, $\beta = 8$ min

TABLE 3.3 – Récapitulatif des paramètres pour le centre de tri

12. Indicateurs de performance à calculer

1. P_{00} : Probabilité que la machine soit inactive et prête
2. L_q : Nombre moyen de cargaisons dans l'orbite (stockage)
3. L_s : Nombre moyen de cargaisons dans le système (orbite + machine)
4. W_q : Temps moyen d'attente dans l'orbite
5. W_s : Temps moyen de séjour dans le système
6. A_v : Disponibilité de la machine (probabilité qu'elle fonctionne)
7. F_f : Probabilité de panne lors de la configuration
8. $E(T_b)$: Durée moyenne d'une période d'occupation
9. $E(T_c)$: Durée moyenne d'un cycle complet

3.7 Analyse de sensibilité et exemples numériques

L'analyse de sensibilité représente une étape cruciale dans l'étude des modèles stochastiques, en particulier pour les systèmes d'attente avec rappels (*Retrial Queueing Systems*) intégrant des mécanismes complexes tels que les pannes, les réparations, les phases de préparation (*Setup*) et les vacances du serveur. Cette analyse permet d'évaluer la robustesse du système et de comprendre comment les variations des paramètres d'entrée influencent de manière majeure les mesures de performance globales.

L'objectif de ce chapitre est de présenter une étude quantitative et qualitative rigoureuse de l'impact des paramètres clés sur les performances du système. Nous examinerons méticuleusement leurs effets sur la longueur moyenne de l'orbite ($E[L_q]$) ainsi que sur les différentes probabilités d'état du serveur (l'oisiveté P_{00} , les vacances P_v , la préparation P_s , et la réparation P_r).

3.8 Configuration des Paramètres et Discussion des Courbes

3.8.1 Choix des paramètres numériques

Afin d'illustrer le comportement dynamique du modèle et d'exécuter l'analyse graphique, nous établissons un ensemble de valeurs de base pour les paramètres du système. Ces valeurs, choisies pour garantir la condition de stabilité stricte du modèle, sont récapitulées dans le Tableau 3.4.

Paramètre	Description	Valeur Numérique
λ	Taux d'arrivée des clients primaires	0.5
μ_n	Taux de service régulier du serveur	2.0
ξ	Taux de rappel depuis l'orbite	0.2
ω	Taux de déclenchement des vacances	0.3
η	Taux de réparation du serveur en panne	0.5
β	Taux de préparation du serveur (<i>Setup</i>)	0.4
α	Taux de transition vers l'état secondaire	0.1

TABLE 3.4 – Valeurs numériques des paramètres fondamentaux du système.

Chacune des sections suivantes présente l'effet de la variation d'un paramètre spécifique tout en maintenant les autres paramètres constants aux valeurs de base définies dans le Tableau 3.4.

3.8.2 Impact du taux de rappel (ξ) sur la taille de l'orbite

Dans cette première analyse, nous étudions le comportement du nombre moyen de clients dans l'orbite ($E[L_q]$) en fonction du taux de rappel (ξ), pour différentes valeurs du taux d'arrivée (λ).

La Figure 3.1 illustre une décroissance prononcée de $E[L_q]$ à mesure que le taux de rappel ξ augmente. Un taux de rappel plus élevé signifie que les clients en attente dans l'orbite tentent d'accéder au serveur plus fréquemment, ce qui accélère leur prise en charge et réduit mécaniquement leur temps de séjour dans l'orbite dès que le serveur se libère. De plus, l'augmentation du taux d'arrivée λ décale visiblement la courbe vers le haut, car un flux de trafic plus intense accroît la congestion initiale et alimente davantage l'orbite.

3.8.3 Effet du taux de service régulier (μ_n) sur la taille de l'orbite

Cette section met en évidence l'impact de la vitesse de traitement du serveur sur l'encombrement de l'orbite de rappel.

La courbe de la Figure 3.2 met en évidence la décroissance asymptotique de $E[L_q]$ lorsque le taux de service μ_n augmente. Un serveur plus rapide libère les unités plus

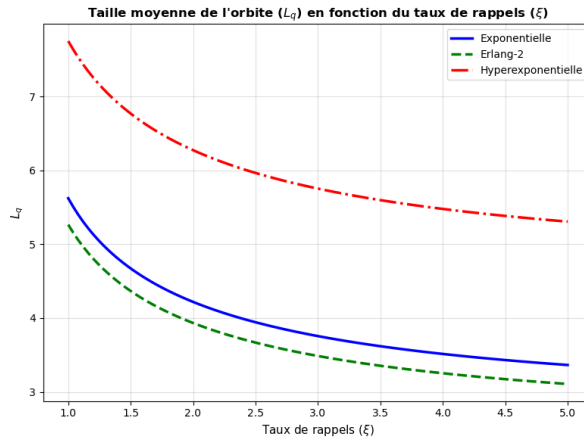


FIGURE 3.1 – Évolution de $E[L_q]$ en fonction de ξ pour différentes valeurs de λ .

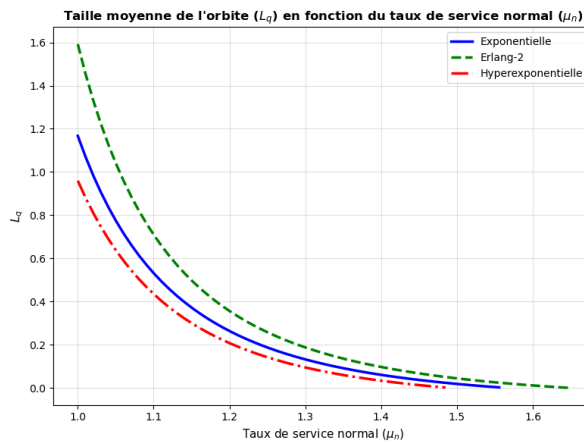


FIGURE 3.2 – Impact du taux de service μ_n sur la longueur moyenne de l'orbite $E[L_q]$.

efficacement, diminuant ainsi la probabilité que les nouveaux clients arrivants ou en rappel trouvent le serveur occupé. L'écart important entre les courbes pour de faibles valeurs de μ_n souligne que le système est extrêmement sensible à la charge globale (λ) lorsqu'il opère près de sa limite de saturation.

3.8.4 Variation de la probabilité de vacances en fonction de son taux (ω)

Nous examinons ici comment la propension du serveur à prendre des vacances (P_v) réagit aux variations du paramètre de fréquence des vacances ω .

Comme le montre la Figure 3.3, une augmentation de ω se traduit par une hausse naturelle de la probabilité globale de vacances P_v . Plus la politique opérationnelle favorise le déclenchement des vacances, plus le serveur passera une proportion importante du temps total dans cet état d'absence. En revanche, lorsque le flux d'arrivée λ est plus intense, le serveur est plus sollicité par les clients, ce qui restreint ses opportunités de partir en vacances et abaisse visiblement la courbe.

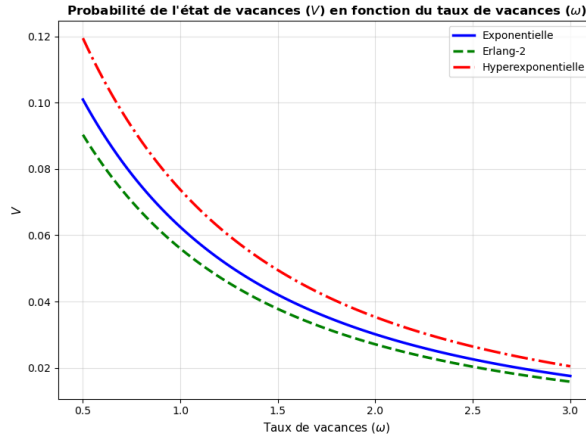


FIGURE 3.3 – Variation de la probabilité de vacances P_v en fonction de ω .

3.8.5 Sensibilité de la probabilité de préparation (*Setup*) par rapport au taux de service

Cette figure évalue la probabilité que le serveur soit en phase de préparation (P_s) face aux changements du taux de service μ_n .

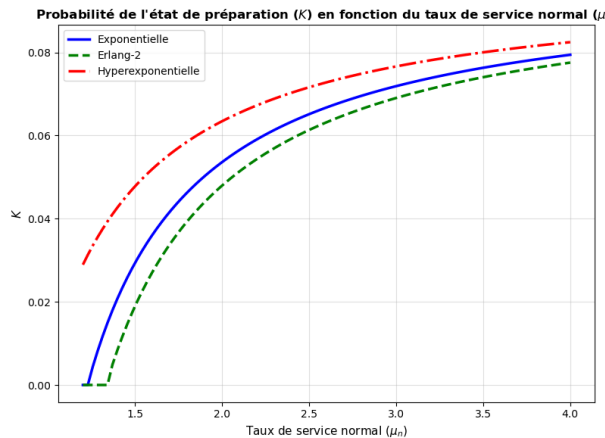


FIGURE 3.4 – Effet du taux de service μ_n sur la probabilité de phase de préparation P_s .

L'augmentation de la rapidité de service (μ_n) permet au système d'écouler les clients plus vite, favorisant ainsi les retours à l'état vide. Toutefois, la dynamique de la phase de préparation (*Setup*) dépend intimement de la fréquence à laquelle le serveur doit se réactiver pour servir un client après une période d'absence ou de vacances. Les courbes de la Figure 3.4 démontrent l'interaction non linéaire complexe entre la charge du système (λ) et l'efficacité du service sur l'occurrence de ces phases transitoires.

3.8.6 Impact du taux de réparation (η) sur la probabilité de réparation

Nous analysons ici la résilience et la capacité de récupération du système face aux pannes accidentelles, en observant la probabilité d'état de réparation (P_r).

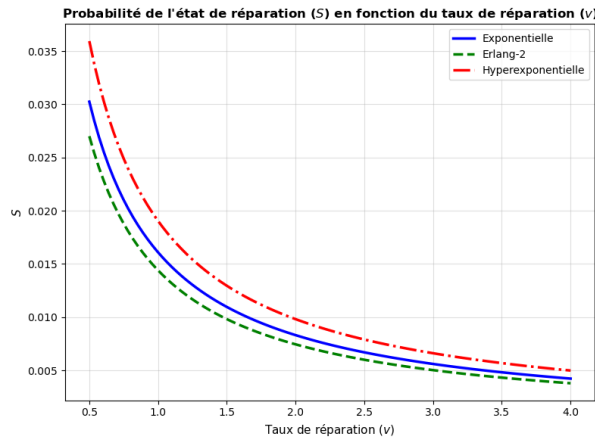


FIGURE 3.5 – Sensibilité de la probabilité de réparation P_r en fonction de η .

Le graphique 3.5 illustre une décroissance logique et rapide de la probabilité P_r à mesure que le taux de réparation η augmente. Mathématiquement, une maintenance plus efficace (valeur élevée de η) minimise la durée moyenne d'indisponibilité du serveur suite à une défaillance. Le système passe donc moins de temps bloqué dans l'état de panne, ce qui optimise la disponibilité globale pour les clients en attente et stabilise l'orbite.

3.8.7 Évolution de la probabilité d'oisiveté (P_{00}) en fonction du taux de service

Enfin, cette dernière figure adopte une perspective globale sur la disponibilité du serveur en étudiant la probabilité d'oisiveté absolue (P_{00}).

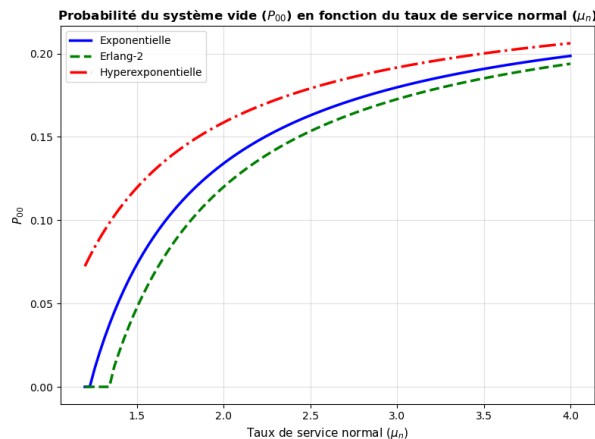


FIGURE 3.6 – Probabilité d'oisiveté du système P_{00} en fonction du taux de service μ_n .

La probabilité que le système soit totalement vide et le serveur oisif (P_{00}) croît de manière monotone avec l'augmentation du taux de service μ_n (Figure 3.6). Un service expéditif permet de liquider rapidement les demandes des arrivées primaires ainsi que les flux de rappel, laissant le serveur inactif en attente de nouveaux événements. Les courbes correspondant aux valeurs de λ les plus fortes se situent logiquement en dessous, car un afflux constant réduit mécaniquement ces périodes d'oisiveté.

3.9 Conclusion

Dans ce chapitre, nous avons présenté l'analyse d'un modèle de file d'attente avec rappel à serveur unique non fiable avec temps de rappel général, vacances actives et temps de configuration. Si certaines conditions nécessaires et suffisantes sont satisfaites, le système peut être stabilisé. En utilisant l'approche des variables supplémentaires et l'approche de la fonction génératrice de probabilité (PGF), nous avons déterminé la PGF du nombre de clients dans le système et dans son orbite.

Les principaux résultats obtenus sont :

1. La condition de stabilité du système : $\rho = \frac{\lambda}{\omega}(1 + \frac{\lambda}{\omega})(1 - W_v^*(\omega)) < N^*(\omega)$
2. Les expressions explicites des probabilités des états du système à l'équilibre
3. Les formules pour les mesures de performance : taille moyenne de l'orbite (L_q), taille moyenne du système (L_s), temps moyen d'attente (W_q) et temps moyen de séjour (W_s)
4. Les mesures de fiabilité : disponibilité du serveur (A_v) et probabilité de panne (F_f)
5. Les expressions des périodes moyennes d'occupation ($E(T_b)$) et de cycle ($E(T_c)$)
6. Un modèle de coût avec optimisation du taux de service par la méthode QFSM

La performance du modèle est illustrée à l'aide de Python. Le coût opérationnel du système de file d'attente est optimisé en ajustant le taux de service du serveur.

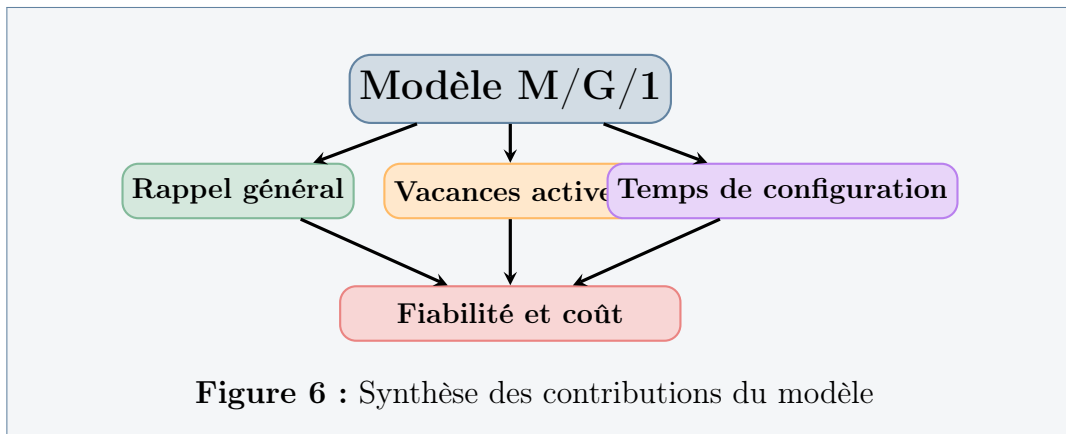


Figure 6 : Synthèse des contributions du modèle

Conclusion générale

Dans ce mémoire, nous avons développé une analyse complète et rigoureuse des systèmes de files d'attente, en partant des bases théoriques jusqu'à l'étude d'un modèle avancé intégrant plusieurs caractéristiques réalistes.

Le premier chapitre a établi les fondements probabilistes nécessaires : processus de comptage, processus de renouvellement, processus de Poisson avec ses propriétés fondamentales d'indépendance des accroissements et d'absence de mémoire, puis processus de Markov à temps discret et à temps continu, pour aboutir aux processus de naissance et de mort. Ces concepts ont été illustrés par les modèles classiques M/M/1, M/M/1/K, M/M/c, M/M/c/c et M/M/ ∞ , dont les distributions stationnaires ont été explicitées.

Le deuxième chapitre a abordé l'analyse exacte du système M/G/1, où les temps de service suivent une loi générale. Nous avons mis en évidence la non-markovianité du processus continu et introduit la méthode fondamentale des chaînes de Markov incluses aux instants de départ. La dérivation de la formule de Pollaczek-Khinchine a permis d'obtenir l'expression explicite du nombre moyen de clients dans le système :

$$L = \rho + \frac{\lambda^2 \mathbb{E}[S^2]}{2(1 - \rho)},$$

ainsi que le temps d'attente moyen $W_q = \lambda \mathbb{E}[S^2] / [2(1 - \rho)]$. L'étude de la période d'occupation a révélé sa structure de renouvellement et sa moyenne $\mathbb{E}[B] = \mathbb{E}[S] / (1 - \rho)$. Une comparaison entre les cas M/D/1, M/M/1 et M/H₂/1 a montré l'impact crucial de la variabilité des temps de service sur les performances.

Le troisième chapitre a proposé un modèle original de file d'attente M/G/1 non fiable avec rappel général, vacances actives et temps de configuration. Ce modèle intègre six états du serveur (inactif, service normal, vacances actives, configuration, panne) et un mécanisme d'orbite pour les clients bloqués. En utilisant la méthode des variables supplémentaires, nous avons établi le système d'équations différentielles aux différences régissant le processus, puis dérivé les fonctions génératrices et les probabilités stationnaires. Les mesures de performance clés (taille moyenne de l'orbite L_q , temps moyen d'attente W_q , disponibilité A_v , probabilité de panne F_f , périodes d'occupation $E(T_b)$ et de cycle $E(T_c)$) ont été explicitées. Une application au centre de tri de déchets recyclables a illustré la correspondance entre le modèle mathématique et un système réel, avec des paramètres numériques concrets. Enfin, une analyse de sensibilité a permis d'étudier l'influence des paramètres μ_n , μ_w , ξ et λ sur les performances.

Les principaux résultats obtenus sont les suivants :

1. La condition de stabilité du système : $\rho = \frac{\lambda}{\omega}(1 + \frac{\lambda}{\omega})(1 - W_v^*(\omega)) < N^*(\omega)$.
2. Les expressions explicites des probabilités des états du système à l'équilibre.
3. Les formules pour la taille moyenne de l'orbite L_q , le temps moyen d'attente W_q , la disponibilité A_v et la probabilité de panne F_f .

4. Les expressions des périodes moyennes d'occupation $E(T_b)$ et de cycle $E(T_c)$.
5. Une illustration numérique via Python montrant l'impact de la variabilité des services sur les performances.

Perspectives. Ce travail ouvre plusieurs pistes pour des recherches futures :

- L'introduction de politiques de vacances multiples ou de vacances avec interruption.
- L'extension au cas de plusieurs serveurs (modèle M/G/c).
- La prise en compte de clients impatients (renégation, abandon).
- L'optimisation conjointe des taux de service et des paramètres de rappel pour minimiser un coût global.
- La validation du modèle sur des données réelles provenant d'un centre de tri ou d'un système industriel.

En conclusion, ce mémoire a contribué à l'analyse des files d'attente complexes en proposant un modèle unifié intégrant la non-fiabilité, les vacances actives, le temps de configuration et les rappels généraux, avec une validation sur une application pratique.

Bibliographie

- [1] Arivudainambi, D., Godhandaraman, P., and Rajadurai, P. (2014). Performance analysis of a single server retrial queue with working vacation. *OPSEARCH*, 51, 434-462.
- [2] Arrar, N., Djellab, N., and Baillon, J.-B. (2012). On the asymptotic behavior of M/G/1 retrial queue with batch arrivals and impatience phenomenon. *Mathematical and Computer Modeling*, 55, 654-665.
- [3] Arrar, N., Djellab, N., and Baillon, J.-B. (2017). On the stochastic decomposition of single server retrial queueing systems. *Turkish Journal of Mathematics*, 41(4), 918-932.
- [4] Arrar, N., Derouiche, L., and Djellab, N. (2018). On the asymptotic behaviour of M/G/1 retrial queue with priority customers, Bernoulli schedule and general retrial times. *International Journal of Applied Mathematics*, 48(2), 206-213.
- [5] Banik, A.D., Gupta, U.C., and Pathak, S.S. (2007). On the GI/M/1/N queue with multiple working vacations-analytic analysis and computation. *Applied Mathematical Modelling*, 31, 1701-1710.
- [6] Chandrasekaran, V.M., Indhira, K., Saravanarajan, M.C., and Rajadurai, P. (2016). A survey on working vacation queueing models. *International Journal of Pure and Applied Mathematics*, 106, 33-41.
- [7] Falin, G.I., and Templeton, J.G.C. (1997). *Retrial Queues*. Chapman and Hall, London, UK.
- [8] Gomez-Corral, A. (1999). Stochastic analysis of a single server retrial queue with general retrial time. *Naval Research Logistics*, 46(5), 561-581.
- [9] Gupta, P., and Kumar, N. (2021). Performance Analysis of Retrial Queueing Model with Working Vacation, Interruption, Waiting Server, Breakdown, and Repair. *Journal of Scientific Research*, 13(3), 833-844.
- [10] Gupta, P., and Kumar, N. (2021). Study of feedback retrial queueing system with working vacation, setup time and perfect repair. *Ratio Mathematica*, 41, 291-307.
- [11] Hadjadj Houari, Arrar Nawel, Lahcene Yahiaoui. (2025). UNRELIABLE M/G/1 QUEUE WITH GENERAL RETRIAL TIME, WORKING VACATION AND SETUP TIME. *RT&A*, No 1 (82), Volume 20, March 2025, p. 548.
- [12] Keilson, J., and Servi, L.D. (1986). Oscillating random walk models for GI/G/1 vacation systems with Bernoulli schedule. *Journal of Applied Probability*, 23(3), 790-802.
- [13] Laxmi, P.V., Goswami, V., and Jyothsha, K. (2013). Optimization of balking and reneging queue with vacation interruption under N-policy. *Journal of Optimization*, 683708, 9.

- [14] Levy, Y., and Yechiali, U. (1975). Utilization of idle time in an M/G/1 queueing system. *Management Science*, 22, 202-211.
- [15] Li, J., and Tian, N. (2007). The M/M/1 queue with working vacations and vacation interruptions. *Journal of Systems Science and Systems Engineering*, 16, 121-127.
- [16] Manoharan, P., and Jeeva, T. (2019). Impatient customers in an M/M/1 working vacation queue with a waiting server and setup time. *Journal of Computer and Mathematical Sciences*, 10(5), 1189-1196.
- [17] Manoharan, P., and Jeeva, T. (2020). Impatient customers in a Markovian queue with Bernoulli schedule working vacation interruption and setup time. *Applications and Applied Mathematics*, 15(2), 725-739.
- [18] Phung-Duc, T. (2015). M/M/1/1 retrial queues with setup time. *Advances in Intelligent Systems and Computing*, 383, 93-104.
- [19] Phung-Duc, T. (2017). Single server retrial queues with setup time. *Journal of Industrial and Management Optimization*.
- [20] Rardin, R.L. (1997). *Optimization in Operations Research*. Prentice-Hall, Upper Saddle River.
- [21] Servi, L.D., and Finn, S.G. (2002). M/M/1 queues with working vacations (M/M/1/WV). *Performance Evaluation*, 50, 41-52.
- [22] Takagi, H. (1991). *Queueing Analysis : Vacation and Priority Systems, Volume I*. North-Holland, Amsterdam.
- [23] Tian, N.-S., Li, J.-H., and Zhang, Z.G. (2009). Matrix analytic method and working vacation queues-a survey. *International Journal of Information Management Science*, 20, 603-633.
- [24] Zhang, M., and Hou, Z. (2010). Performance analysis of M/G/1 queue with working vacations and vacation interruption. *Journal of Computational and Applied Mathematics*, 234, 2977-2985.