

Table des matières

1	Processus Stochastiques	11
1.1	Processus de comptage	12
1.2	Processus de renouvellement	15
1.3	Processus de Poisson	15
1.4	Processus de naissance et de mort	17
1.4.1	Processus de naissance	18
1.4.2	Processus de mort	18
2	Systèmes de Files d'Attente Classiques	21
2.1	Les différents types de files d'attente	23
2.2	File d'attente simple	24
2.3	Notation de Kendall :	26
2.4	Loi de Little	28
2.5	Quelque Modèles Sur Les Files d'Attente	28
2.5.1	Modèle d'attente $M/M/1$	28
2.5.2	Modèle d'attente $M/M/1/K$	30
2.5.3	Modèle d'attente $M/M/s$	32
2.5.4	Modèle d'attente $M/M/\infty$	34
3	Les Systèmes De File d'Attente Avec Dérobade	37
3.1	Le Modèle d'attente $M/M/1$ avec dérobade	38
3.2	Le Modèle d'attente $M/M/s$ avec dérobade	40
	Bibliographie	47

Remerciements

Avant toute chose, je tiens à remercier Dieu le tout puissant.

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

Je remercie Dr. ROUANE R. d'avoir accepté à présider le jury de mon mémoire.

Je souhaite également remercier mademoiselle BENZIADI F. et Dr. RAHMANI S. de m'avoir fait l'honneur d'accepter d'être examinatrices de mon mémoire. Leurs remarques et précieux conseils ont abouti à une amélioration du document de mémoire.

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

J'adresse mes sincères remerciements à tous les professeurs qui m'ont enseigné. je remercie toute ma famille pour leur soutien qu'ils m'ont apporté pendant ces années d'études.

Enfin je remercie mes amis pour les bons moments passés ensemble et aussi pour leur soutiens et leur encouragements.

Et pour finir, merci à toutes les personnes que j'ai oubliées de citer et qui m'ont permis de mener à bien ce mémoire .

Dédicaces

Je dédie ce travail à : Mes parents, qui ont dépensé pour moi sans compter. En reconnaissance de tous les sacrifices consentis par tous et chacun pour me permettre d'atteindre cette étape de ma vie. Avec toute ma tendresse.

A mon chère frère "Brahim el khallil" mes chères soeurs "Souhila", "Zahira", et "Safia" sans oublier mes beaux frères et leurs enfants : "Seddik", "Fatima zohra", "Abdelmalek", "Yassine", et "Anes".

A mes oncles, tantes, cousins et cousines.

Spécial dédicace a ma chère soeur "Hafida" qui ma bien aidé le long de mon parcours universitaire et qui ma donné beaucoup de succès.

A tous les membres de ma promotion.

A mes amis "Khaoula", "Shafika", "Halima", "Khalida", "Soumia", "Amel", "Latifa", et "Khadidja".

Introduction générale

La théorie des files d'attente fournit un outil très puissant et efficace pour la modélisation des systèmes admettant un phénomène d'attente. Cette théorie datent du début du XXème siècle par les travaux de l'ingénieur danois Agner Krarup Erlang (1878, 1929). Ses études sur le trafic téléphonique de Copenhague pour le mieux gérer afin de déterminer le nombre de circuits nécessaires pour fournir un service téléphonique acceptable, sont considérées comme la première brique dans cette théorie [11]. Ensuite, les files d'attente ont été intégrés dans la modélisation des systèmes informatiques et aux réseaux de communication. Cette intégration dans ces domaines et d'autre a permet une évolution de cette théorie surtout dans l'évaluation des paramètres de performances des systèmes informatiques et aux réseaux de communication. Actuellement ce sont les applications dans le domaine de l'analyse de performance des réseaux (téléphone mobile, Internet, multimédia, . . .) qui suscitent le plus de travaux. L'étude d'un système de files d'attente consiste à calculer ces paramètres de performance afin d'évaluer son rendement, et améliorer son fonctionnement (minimiser le temps d'attente et le temps d'inactivité de l'installation) de savoir par exemple si le nombre de serveurs dans le système est adéquat pour gérer le flux de demandes ou encore d'appréhender les effets d'une modification des conditions de fonctionnement, et ainsi prendre des décisions sur le nombre minimum de ressources nécessaires.

Depuis les travaux d'Erlang [2] Un grand nombre d'applications dans tous les domaines ont été mis en oeuvre et publiées. En 1953, David G. Kendall [2] a introduit la notation de Kendall pour décrire les caractéristiques d'un système de file d'attente. en 1957 d'une manière particulièrement élégante et efficace Jackson a traité certains réseaux de files d'attente. En 1961, Thomas L. Saaty [3], auteur de l'un des premiers livres complets sur la théorie des files d'attente. Ensuite c'est les contributions des mathématiciens Khintchine, Palm, Pollaczek et Kolmogorov qui ont vraiment poussés la théorie des files d'attente.

Mon mémoire est composé de trois chapitres :

Dans le premier chapitre, nous présentons les notions de bases de l'étude des systèmes de files d'attente, à savoir les processus stochastiques :

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

Dans le deuxième chapitre, nous introduisons la terminologie de la théorie des files d'attente. Certaines définitions et notations qui sont nécessaires dans l'étude des systèmes de files d'attente comme (Notation de Kendall, la loi de Little ,...etc) sont notamment données. Ensuite nous étudions quelque modèles de files d'attente ($M/M/1$, $M/M/1/K$, $M/M/s$, $M/M/\infty$) et l'évaluation de leurs paramètres de performance. Enfin dans le troisième chapitre nous présentons une étude de certains modèles d'attente avec clients impatientes. Nous traitons le cas de files d'attente $M/M/1$ avec dérobade et le modèle $M/M/s$ avec dérobade [9].

Chapitre 1

Processus Stochastiques

Les processus stochastiques décrivent l'évolution d'une grandeur aléatoire en fonction du temps (ou de l'espace). Il existe de nombreuses applications des processus aléatoires notamment en physique statistique [31] (par exemple le ferromagnétisme, les transitions phases, etc), en biologie (évolution, génétique et génétique des populations), médecine (croissance de tumeurs, épidémie), et bien entendu les sciences de l'ingénieur. Dans ce dernier domaine, les applications principales sont pour l'administration des réseaux, de l'internet, des télécommunications et bien entendu dans les domaines économique et financier.

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

Définitions et propriétés de base :

Définition 1.0.1. *Un processus stochastique est une famille de variables aléatoires X_t , $t \in T$ ou chaque variable aléatoire X_t est indexée par le paramètre $t \in T$, si T est un ensemble de \mathbb{R}_+ , alors t signifie temps.*

Généralement X_t représente l'état du processus stochastique au temps t [51].

- *Si T est dénombrable, i.e $T \subseteq \mathbb{N}$, alors nous disons que X_t , $t \in T$ est un processus à temps discret.*
- *Si T est un intervalle de $[0; \infty)$, alors le processus stochastique est dit un processus à temps continu.*

L'ensemble des valeurs de X_t est appelé l'espace d'état, qui peut également être soit discret (fini ou infini dénombrable) ou continu (un sous-ensemble de \mathbb{R} ou \mathbb{R}^n

), donc nous écrivons $(X_n)_{n \geq 0}$ pour le processus à temps discret et $(X_t)_{t \geq 0}$ pour le processus à temps continu.

1.1 Processus de comptage

Définition 1.1.1. (processus de comptage) Un processus stochastique $[N(t), t \in \mathbb{R}^+]$ est un processus de comptage si $N(t)$ représente le nombre total d'événements qui se sont produits entre 0 et t , il doit donc satisfaire

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

Un processus de comptage est un processus discret à temps continu. Un second processus peut être associé au processus des temps d'occurrence; le processus des temps d'inter-arrivées $\{W_n, n \in \mathbb{N}_0\}$ ou $\forall n \in \mathbb{N}_0$ la variable aléatoire W_n est le temps d'attente entre les $(n-1)^{ieme}$ et n^{ieme} occurrences [50], c-à-d :

$$W_n = T_n - T_{n-1} \text{ avec } T_n \text{ est le temps d'arrivé du } n^{ieme} \text{ client.}$$

Proposition 1.1.1. Les relations suivantes sont triviales tel que $T_0 = 0$ à vérifier :

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

demonstration : on a

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

Définition 1.1.2. (processus à accroissements indépendants)

Un processus $\{X_t\}$ tel que $X_0 = 0$ est à accroissements indépendants si pour tout suite finie $0 < t_1 < t_2 < t_3 \dots < t_n$ les variables aléatoires $X_{t_1}, X_{t_2} - X_{t_1}, \dots, X_{t_n} - X_{t_{n-1}}$ sont indépendantes.

Définition 1.1.3. *Un processus à accroissements indépendants est à accroissements stationnaires si la loi de l'accroissement $(X_{t+s} - X_t)$ ne dépend pas de t pour tout $t \geq 0$.*

Définition 1.1.4. *Un processus de comptage $\{N(t), t \geq 0\}$ est un processus de poisson d'intensité $\lambda > 0$ si :*

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

loi de Poisson et loi exponentielle :

Définition 1.1.5. *Une variable aléatoire X à valeurs entières suit une loi de Poisson de paramètre $\lambda > 0$ si :*

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

Définition 1.1.6. *Une variable aléatoire Y à valeurs réelles strictement positives suit une loi exponentielle de paramètre $\mu > 0$ si :*

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

Distribution exponentielle

Soit t une variable aléatoire avec $t \geq 0$ qui suit une distribution exponentielle. La densité de probabilité de t est $f(t) = \mu \exp^{-\mu t}$ et la distribution cumulée correspondante est $F(t) = 1 - \exp^{-\mu t}$. L'espérance et la variance de t sont :

$$E(t) = 1/\mu,$$

et

$$V(t) = 1/\mu^2, \text{ respectivement.}$$

Distribution de Poisson

Soit n une variable aléatoire discrète avec $n = 0, 1, \dots$ qui suit une distribution Poisson. La distribution de probabilité de n est $P_n = \frac{\lambda^n}{n!} \exp^{-\lambda}$. L'espérance et la variance de n sont :

$$E(n) = \lambda,$$

et

$V(n) = \lambda$, respectivement.

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

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

Relation entre la distribution Exponentielle et la distribution de Poisson

La densité de probabilité d'une distribution exponentielle $f(t) = \mu \exp^{-\mu t}$ Supposons τ est exponentielle avec une espérance $1/\mu$, et n est de Poisson de moyenne λ . on

a :

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

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

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

Définition 1.1.7. Une variable aléatoire X est dite **sans mémoire** (ou sans usure) si :

$$\forall s, t \geq 0$$

$$\mathbb{P}(X > t + s | X > t) = \mathbb{P}(X > s)$$

Si X est la durée de vie d'un matériel quelconque l'équation précédant s'interprète de la manière suivante, sachant le matériel en état de bon fonctionnement au temps t , la loi de probabilité de sa durée de vie future est la même que celle de sa durée de vie initiale. En d'autres termes, le matériel ne s'use pas.

Remarque 1.1. L'unique loi de probabilité continue à perte mémoire est la loi exponentielle, cette définition est similaire à la version discrète à l'exception des variable s et t sont réelles positives et non entières, plutôt que de compte le nombre d'essais jusqu'au premier succès on peut penser à l'heure d'arrivée du premier **appel téléphonique dans un centre d'appel** [28].

1.2 Processus de renouvellement

Inter-arrivées indépendantes :

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

Définition 1.2.1. (*processus de renouvellement*)

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

Définition 1.2.2. (*processus de renouvellement*)

Soit $(X_n)_{n \geq 0}$ une suite de variables aléatoire positives on note S_n la suite des sommes partielles : $S_0 = 0$ et $S_n = X_n + S_{n-1}$ pour tout $n \geq 1$ on considère alors le processus R_t défini comme suit :

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

Par exemple, si les X_n modélisent les durées de vie d'une ampoule R_t représente le nombre d'ampoules changées avant l'instant t ; les X_n peuvent également représenter le temps séparant deux ventes successives, ou deux sinistres successifs pour une compagnie d'assurance. R_t désignera alors, suivant le cas, le nombre d'articles vendus ou le nombre sinistres survenus au cours de l'intervalle de temps $[0, t]$, la suite (S_n) est appelée processus de renouvellement associé aux $(X_n)_{n \geq 0}$ et le processus (R_t) est le processus de comptage.

Par abus de langage, on appelle également R_t Processus de renouvellement.

1.3 Processus de Poisson

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

Exemples :

- arrivées d'appels à un central téléphonique.
- impacts de micrométéorites sur un satellite.
- passage de véhicules a un péage d'autoroute.
- arrivée des clients à un guichet, occurrence d'accidents dans une ville, pannes de machines dans une usine...ect.

De tels phénomènes peuvent se définir par la famille $(A_n)_{n \in \mathbb{N}^*}$ des temps d'arrivées qui sont des variables aléatoires [23]. mais on peut aussi le faire a partir du processus de comptage $(N_t)_{t \in \mathbb{R}}$. N_t est le nombre d'événements apparus jusqu'a l'instant t .

Définition 1.3.1. (*processus de Poisson*)

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

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

Ce processus s'appelle le processus de Poisson d'intensité λ

Définition 1.3.2. (*processus de Poisson*)

Un processus de Poisson $N = (N_t)_{t \geq 0}$ d'intensité λ est un processus de comptage à trajectoires continues à droite [43] tel que :

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

Remarque 1.2. On peut voir le processus de Poisson comme une mesure aléatoire sur $(\mathbb{R}, \mathbf{B}(\mathbb{R}))$: la mesure de l'intervalle $]s, t]$ est $N(]s, t]) = N_t - N_s$.

Proposition 1.3.1. $(R_t)_{t \geq 0}$ est un processus de Poisson si et seulement si les quatre propriétés suivantes sont vérifiées :

- $R_0 = 0$ ps
- pour tout $t > 0$, R_t une loi de poisson d'espérance λt
- pour tous $t > s \geq 0$, $R_t - R_s$ suit une loi de poisson d'espérance $\lambda(t - s)$
- pour tous $0 \leq t_1 < t_2 < \dots < t_n$ les variables aléatoire $R_{t_1}, R_{t_2} - R_{t_1}, \dots, R_{t_n} - R_{t_{n-1}}$ sont indépendantes.

1.4 Processus de naissance et de mort

Définition 1.4.1. (*processus de naissance et de mort*)

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

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

Si $\pi_n(t)$ est la probabilité pour qu'il y a n clients dans le système à l'instant t , l'équation de Kolomogorov [20] s'écrit, pour $n > 0$:

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

C'est-à-dire, en faisant tendre dt vers 0, pour $n > 0$:

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

De la même façon, on obtient pour $n = 0$:

$$\frac{d}{dt}\pi_0(t) = -\lambda_0\pi_0(t) + \mu_1\pi_1(t).$$

Le processus de Poisson est un cas particulier du processus de naissance et de mort pour lequel $\mu_n = 0$ et $\lambda_n = \lambda$ mais dans ce cas, il n'y a pas de régime stationnaire [14].

Les équations différentielles s'écrivent alors :

$$\frac{d}{dt}\pi_0(t) = -\lambda_0\pi_0(t)$$

d'où $\pi_0(t) = \exp^{-\lambda t}$

$$\frac{d}{dt}\pi_n(t) = -\lambda(\pi_n(t) - \pi_{n-1}(t))$$

dont la solution est $\pi_n(t) = \frac{(\lambda t)^n \exp^{-\lambda t}}{n!}$

1.4.1 Processus de naissance

Si la taille d'une population a une transition $n \rightarrow n + 1$ donc il correspond à une naissance.

1.4.2 Processus de mort

Si la taille d'une population a une transition $n \rightarrow n - 1$ donc il correspond à une mort.

Définition 1.4.2. (*processus de naissance*)

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

Définition 1.4.3. (*processus de naissance et de mort*)

C'est un cas particulier de chaîne de Markov [7] où seules les transitions d'un état à un état voisin sont permises, on s'intéresse au cas continu avec des taux de transitions.

C'est le point de départ de la théorie des files d'attente.

On introduit les données suivantes :

- λ_n : taux de naissances quand le nombre de population égale à n .
- μ_n : taux de morts quand le nombre de population égale à n .

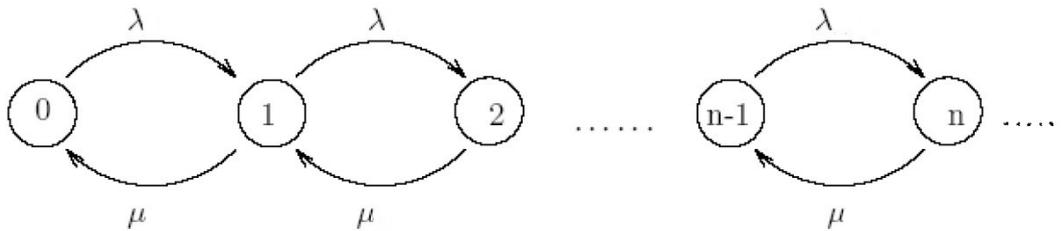


FIGURE 1.1 – Graphe de transition d'un processus de naissance et de mort

Exemple

Dans une file d'attente le temps s'écoulant entre deux arrivées consécutives est distribué exponentiellement ainsi que le temps de service.

Alors cette file d'attente peut être modéliser par un processus de naissance et de mort :

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

Chapitre 2

Systemes de Files d'Attente Classiques

La théorie des files d'attente a commencé en 1909 avec les travaux de recherches de l'ingénieur danois Agner Krarup Erlang (1878,1929) sur le trafic téléphonique de Copenhague pour déterminer le nombre de circuits nécessaires afin de fournir un service téléphonique acceptable. Par la suite, les files d'attente ont été intégrés dans la modélisation de divers domaines d'activité [19]. On assista alors à une évolution rapide de la théorie des files d'attente qu'on appliqua à l'évaluation des performances des systèmes informatiques et aux réseaux de communication. Les chercheurs oeuvrant dans cette branche d'activité ont élaboré plusieurs nouvelles méthodes qui ont été ensuite appliquées avec succès dans d'autres domaines, notamment dans le secteur de la fabrication. On a aussi constaté une résurgence des applications pratiques de la théorie des files d'attente dans des secteurs plus traditionnels de la recherche opérationnelle, un mouvement mené par Peter Kolesar et Richard Larson [8]. Grâce à tous ces développements, la théorie des files d'attente est aujourd'hui largement utilisé et ses applications sont multiples.

Définition 2.0.4. (*File d'attente*) [13]

Une file d'attente est un système dans lequel arrivent des clients auquel des serveurs fournissent un service. Ce formalisme peut être utilisé dans des situations diverses : guichet de poste trafic routier, traitement des instructions par un processeur, gestion de communications téléphoniques, ateliers de réparation,... etc.

On parle de phénomène d'attente chaque fois que certaines unités appelées clients se

présentent d'une manière aléatoire à de stations afin de recevoir un service dont la durée est généralement aléatoire.

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

Classification des files d'attente[40]

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

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

Terminologie et notations [4]

En lien avec la loi exponentielle :

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

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

En situation d'équilibre on note : [36]

- P_n : Probabilité qu'il y ait n clients dans le système.

- L : Nombre moyen (espérance mathématique) de client dans le système.
- L_q : Nombre de clients dans la file d'attente excluant ceux qui sont dans le service.
- W_q : Le temps moyen passé par un client dans la file (excluant le temps de service)

2.1 Les différents types de files d'attente

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

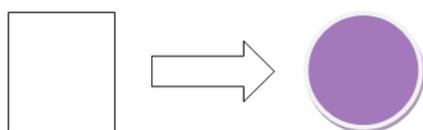


FIGURE 2.1 – File d'attente avec un seul espace d'attente et un seul serveur

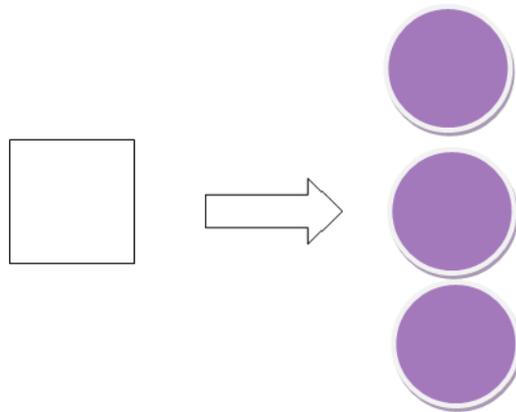


FIGURE 2.2 – File d'attente avec un seul espace d'attente et plusieurs serveurs

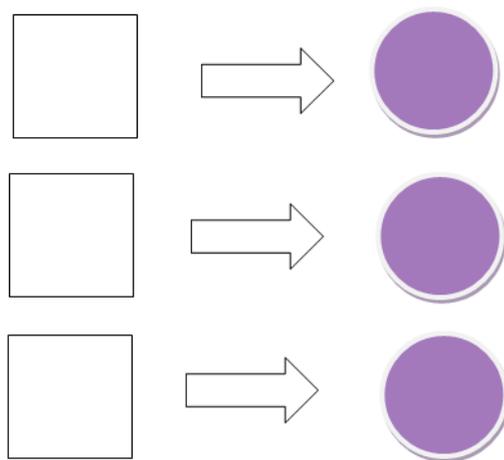


FIGURE 2.3 – File d'attente avec plusieurs espaces d'attente et plusieurs serveurs

2.2 File d'attente simple

La file simple

Une file d'attente simple est un système constitué d'un ou plusieurs serveurs et d'un espace d'attente. Les clients arrivent de l'extérieur, patientent éventuellement dans la file d'attente, reçoivent un service, puis quittent la station [27]. Afin de spécifier

complètement une file d'attente simple, on doit caractériser le processus d'arrivée des clients, le temps de service ainsi que la structure et la discipline de service de la file d'attente .

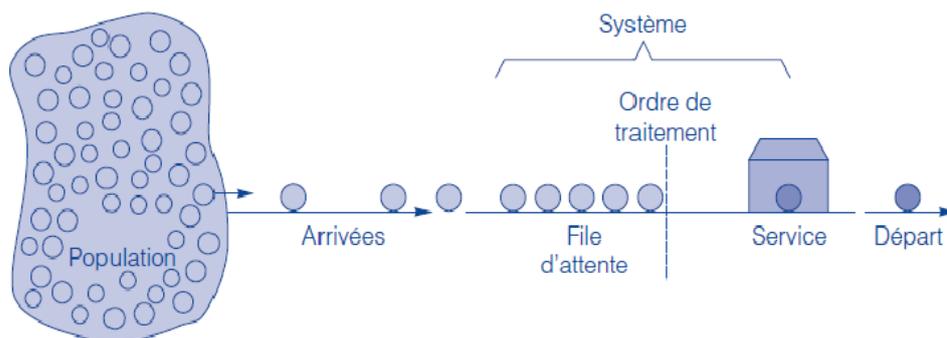


FIGURE 2.4 – File d'attente simple

Processus d'arrivée

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

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

Si T_n désigne la variable aléatoire mesurant le temps séparant l'arrivée du $(n - 1)^{\text{ième}}$ client et du $n^{\text{ième}}$ client [29], on a alors :

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

Temps de service

Considérons tout d'abord une file à serveur unique.

On note D_n la variable aléatoire mesurant l'instant de départ du $n^{\text{ième}}$ client du système et Y_n la variable aléatoire mesurant le temps de service du $n^{\text{ième}}$ client (le temps séparant le début et la fin du service). Un instant de départ correspond toujours à une fin de service, mais ne correspond pas forcément à un début de service [37]. Il se peut en effet qu'un client qui quitte la station laisse celle-ci vide. le serveur est alors inoccupé jusqu'à l'arrivée du prochain client.

On note μ le taux de service :

$1/\mu$ est la durée moyenne de service.

Structure de la file :*Nombre de serveurs*

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

Capacité de la file :

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

2.3 Notation de Kendall :

La notation suivante, appelée la notation de Kendall, est largement utilisée pour classer les différents systèmes de files d'attente [15] :

$$T/Y/C/K/m/Z$$

avec

1. T : indique le processus d'arrivée des clients. Les symboles utilisés sont :
 - M : Inter-arrivées des clients sont identiquement distribuées selon une loi exponentielle. Il correspond à un processus de Poisson ponctuel (propriété sans mémoire).
 - D : Les temps inter-arrivées des clients ou les temps de service sont constants et toujours les mêmes.
 - GI : Inter-arrivées des clients ont une distribution générale (il n'y a aucune hypothèse sur la distribution mais les inter-arrivées sont indépendantes et identiquement distribuées).
 - G : Inter-arrivées des clients ont une distribution générale et peuvent être dépendantes.

- E_k :Ce symbole désigne un processus où les intervalles de temps entre deux arrivées successives sont des variables aléatoires indépendantes et identiquement distribuées suivant une loi d'Erlang d'ordre k .
- 2. Y :décrit la distribution des temps de service d'un client. Les codes sont les mêmes que T .
- 3. C :nombre de serveurs.
- 4. K :capacité de la file c'est le nombre de places dans le système en d'autre terme c'est le nombre maximal de clients dans le système y compris ceux en service.
- 5. m :population des usagers.
- 6. Z :discipline de service c'est la façon dont les clients sont ordonnés pour être servi. Les codes utilisés sont les suivants :
 - FIFO (first in, first out) ou FCFS (first come, first served) :c'est la file standard dans laquelle les clients sont servis dans leur ordre d'arrivée. Notons que les disciplines FIFO et FCFS ne sont pas équivalentes lorsque la file contient plusieurs serveurs. Dans la première, le premier client arrivé sera le premier à quitter la file alors que la deuxième, il sera le premier à commencer son service. Rien n'empêche alors qu'un client qui commence son service après lui, dans un autre serveur, termine avant lui.
 - LIFO (last in, first out) ou LCFS (last come, first served). Cela correspond à une pile, dans laquelle le dernier client arrivé (donc posé sur la pile) sera le premier traité (retiré de la pile). A nouveau, les disciplines LIFO et LCFS ne sont équivalentes que pour une file mono serveur.
 - SIRO (Served In Random Order), les clients sont servis aléatoirement.
 - PNP (Priority service), les clients sont servis selon leur priorité. Tous les clients de la plus haute priorité sont servis premiers, puis les clients de priorité inférieur sont servis, et ainsi de suite.
 - PS (Processor Sharing), les clients sont servis de manière égale. La capacité du système est partagée entre les clients.

Remarque 2.1. :Dans sa version courte, seuls les trois premiers symboles $T/Y/C$ sont utilisés. dans un tel cas, on suppose que la file est régie par une discipline FIFO et que le nombre de places d'attente ainsi que celui des clients susceptibles d'accéder au système sont illimités.

2.4 Loi de Little

La loi de Little est une relation très général qui s'applique à une grande classe de systèmes.

Elle ne concerne que le régime permanent du système. Aucune hypothèse sur les variables aléatoires qui caractérisent le système (temps d'inter arrivées, temps de service,...) n'est nécessaire. La seule condition d'application de la loi de Little est que le système soit stable. le débit du système est alors indifféremment, soit le débit de sortie : $d_s = d_e = d$. La loi de Little s'exprime telle que dans la propriété suivante :

Théorème 2.4.1. *:(Formule de Little)Le nombre moyen de clients, le temps moyen passé dans le système et le débit moyen d'un système stable en régime permanent se relient de la façon suivante :*

$$L = W \times d$$

2.5 Quelques Modèles Sur Les Files d'Attente

2.5.1 Modèle d'attente $M/M/1$

La file d'attente $M/M/1$ est un modèle caractérisé par des arrivées suivants un processus de Poisson de taux λ , Temps de service exponentielle de paramètre μ et un seul serveur. Les clients arrivent à la station selon un processus de Poisson de taux λ , si le serveur est vide le client est pris en charge immédiatement sinon il rejoint la file d'attente (de capacité illimitée et discipline FIFO), les temps des inter-arrivées sont indépendants. Ces clients reçoivent le service selon une distribution exponentielle de paramètre μ . Ces temps de service sont également supposés indépendants [30]. En outre, toutes les variables aléatoires concernés sont censés être indépendants les uns des autres. Les clients sont servis dans leur ordre d'arrivée. Cette file est un cas particulier du processus de naissance et de mort

$$\forall n \in \mathbb{N}, \quad \lambda_n = \lambda, \quad \mu_n = \mu$$

ie indépendants du nombre de client dans le système.

sous l'hypothèse que $\lambda < \mu$ (le taux d'arrivée est plus petit que le taux de service)

on a :

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

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

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

Remarque 2.2. on note π_0 la probabilité d'être servi immédiatement à l'arrivée.

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

avec

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

π_n : c'est la proba d'avoir n client dans le système.

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

avec

$$\pi_0 = 1 - \rho$$

Le débit :

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

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

donc :

$$d = d_e = d_s$$

Taux d'utilisation du serveur U :

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

Nombre moyen de clients dans le système L :

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

Remarque 2.3. *Plus que l'intensité du trafic se rapproche de un, plus que la longueur moyenne de la file d'attente L tend vers l'infinie.*

Temps moyen de séjour W :

Ce paramètre est obtenu en utilisant la loi de Little :

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

qui peut se décomposer en :

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

La durée moyenne d'attente en régime stationnaire

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

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

Démonstration : [17] Temps moyen passé dans la file d'attente W_q :

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

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

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

2.5.2 Modèle d'attente $M/M/1/K$

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

- ★ le processus d'arrivée des clients dans la file est un processus de Poisson de taux λ .
- ★ le temps de service d'un client est une variable aléatoire exponentielle de taux μ .
- ★ Soit K la capacité de la file d'attente : c'est le nombre maximal de clients qui peuvent être présents dans le système, soit en attente, soit en service.

Quand un client arrive alors qu'il y a déjà K clients présents dans le système, il est perdu. Ce système est connu sous le nom de file $M/M/1/K$. L'espace d'états E est maintenant fini : $E = \{0, 1, 2, \dots, K\}$. La capacité de la file étant limitée, même si les clients arrivent en moyenne beaucoup plus vite que ce que le serveur de la file est capable de traiter, dès que celle-ci est pleine, les clients qui se présentent sont rejetés [40]. Le nombre de clients dans la file ne peut donc jamais "partir" à l'infini. De plus, dès qu'un client est autorisé à entrer, il sortira un jour et son temps de séjour dans la file est fini, puisqu'il correspond au temps de service de tous les clients devant lui et que ce nombre est limité par K . Sur un très long, le débit de sortie sera donc bien égal au débit d'entrée, ce qui correspond bien à la stabilité inconditionnelle du système. Donc ce processus est considéré comme un processus de naissance et de mort avec :

-un taux de naissance $\lambda_n = \lambda$, pour tout $n < K$

-et le taux de mortalité $\mu_i = \mu$, pour tout $i \neq 0$.

soit $\pi_{(i)}, i = 0, 1, 2, \dots, K$, la probabilité pour qu'il ait i clients dans le système à l'instant t .

$$\begin{aligned} \lambda\pi_0 &= \mu\pi_1 \\ (\lambda + \mu)\pi_i &= \lambda\pi_{i-1} + \mu\pi_{i+1} \quad \text{pour } i = 1, 2, \dots, K-1 \\ \lambda\pi_{K-1} &= \mu\pi_K. \end{aligned}$$

puisque la capacité est limitée, nous obtenons un régime stationnaire indépendant des conditions initiales quelle que soit la valeur de l'intensité de trafic ρ

Théorème 2.5.1. :[26]

Le calcul de π_i se fait de la manière suivante :

$$\pi_i = \begin{cases} \frac{(1-\rho)\rho^i}{1-\rho^{K+1}} & \text{pour } i \leq K \\ 0 & \text{pour } i > K \end{cases}$$

Cas particulier $\rho = 1$

dans ce cas

$$\begin{aligned} \pi_i &= \frac{1}{K+1} \\ \pi_i &= 0 \quad \text{pour } i > K \end{aligned}$$

Taux d'utilisation du serveur $U(K)$:

Le taux d'utilisation du serveur est la probabilité pour que le serveur de la file soit occupé donc au moins il y a un client dans la file

$$\begin{aligned} U(K) &= \sum_{i=1}^K \pi_i = 1 - \pi_0 \\ &= \rho \frac{1 - \rho^K}{1 - \rho^{K+1}} \end{aligned}$$

Remarque 2.4.

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

qui représente le taux d'utilisation du serveur Dans le cas $M/M/1$

Nombre moyen de clients L :

$$\begin{aligned} L &= \sum_{i=0}^K i \pi_i = \frac{1 - \rho}{1 - \rho^{K+1}} \sum_{i=0}^K i \rho^i \\ &= \frac{\rho(1 - \rho)}{1 - \rho^{K+1}} \sum_{i=0}^K i \rho^{i-1} \\ &= \frac{\rho}{1 - \rho} \frac{1 - (K+1)\rho^K + K\rho^{K+1}}{1 - \rho^{K+1}} \end{aligned}$$

Pareille que le taux d'utilisation du serveur, lorsque $K \rightarrow +\infty$ on obtient

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

qui représente le nombre moyen de clients pour la file $M/M/1$

Temps moyen de séjour W

Le calcul du temps moyen de séjour W ce fait on appliquant la formule de Little qui relie Nombre moyen de client L et débit d de la file :

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

2.5.3 Modèle d'attente $M/M/s$

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

le processus d'arrivée des clients dans la file est un processus de Poisson de taux λ et le temps de service d'un client est une variable aléatoire exponentielle de taux μ (pour chacun des serveurs). Ce système est connu sous le nom de file $M/M/s$ [21]. L'espace d'états E est comme pour la $M/M/1$ infini. Dans ce cas aussi, le processus modélisant le nombre de clients dans le système est un processus de naissance et de mort avec[1] :

$$\lambda_i = \lambda$$

$$\mu_i = \begin{cases} i\mu, & \text{pour } i = 1, 2, \dots, s-1; \\ s\mu, & \text{pour } i \geq s. \end{cases}$$

La condition de stabilité de ce modèle est :

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

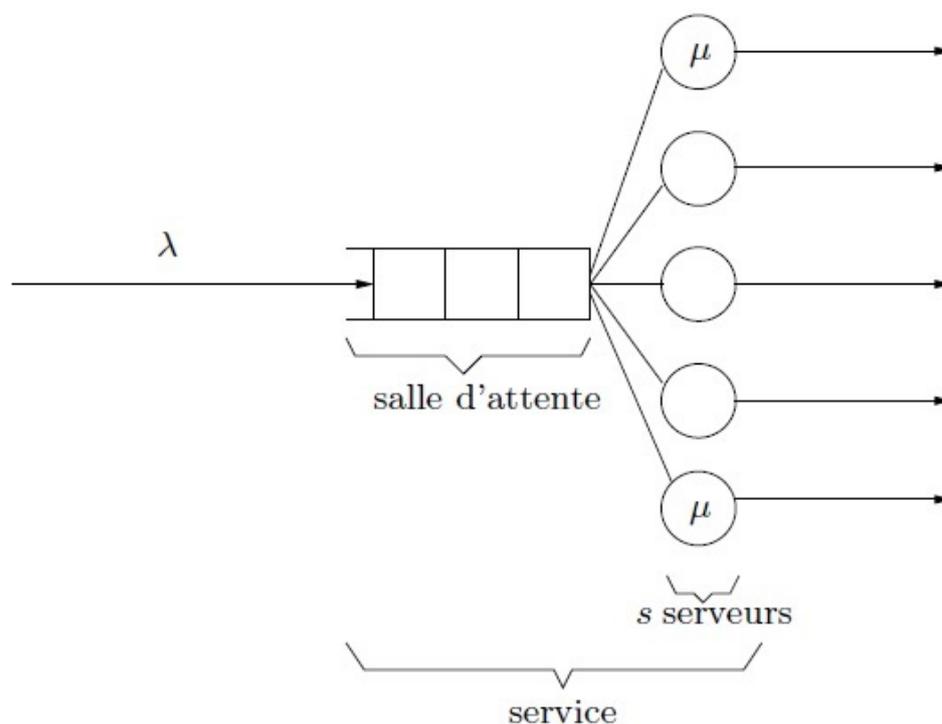


FIGURE 2.5 – Représentation schématique d'une file $M/M/s$

Théorème 2.5.2. :

pour calculer π_i le système doit être stable ($\lambda < s\mu$ donc $\rho < s$) qui exprime le fait que le nombre moyen de clients qui arrivent à la file par unité de temps doit être inférieur au nombre moyen de clients que les serveurs de la file sont capables de traiter par unité de temps, de la manière suivante :

Temps moyen de séjour W :

Le temps moyen de séjour d'un client se décompose en un temps moyen dans la file d'attente, plus un temps moyen de service. Il suffit alors d'appliquer la loi de Little à la seule file :

$$W = W_q + S = \frac{L_q}{d} + \frac{1}{\mu} = \frac{L_q}{\lambda} + \frac{1}{\mu}$$

avec $d = \lambda$

Il reste alors à calculer le nombre moyen de clients en attente dans la file, L_q :

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

On en déduit le temps moyen de séjour

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

Nombre moyen de clients dans le système L :

Le nombre moyen de clients s'obtient alors par application de la loi de Little à l'ensemble de la file :

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

2.5.4 Modèle d'attente $M/M/\infty$ **Description du modèle**

Pour ce modèle de file d'attente, le système est composé d'un nombre illimité de serveurs identiques et indépendants les uns des autres. Dès qu'un client arrive, il est immédiatement servi (c'est le cas où il n'y a pas d'attente). Dans cette file les clients arrivent à des instants $0 < t_1 < t_2 < \dots$ formant un processus de Poisson de taux λ et les temps de service sont exponentiels de taux μ . le taux de transition d'un état

n quelconque vers l'état $n-1$ est égal à $n\mu$ et correspond au taux de sortie d'un client parmi les n clients en service [20]. De même, le taux de transition d'un état n vers l'état $n+1$ est égal à λ qui correspond au taux d'arrivée d'un client, donc c'est un processus de naissance et de mort avec :

$$\forall n \in \mathbb{N}, \lambda_n = \lambda, \mu_n = n\mu$$

$$\pi_n = \frac{\rho^n}{n!} \pi_0$$

avec

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

Notons que la série $\sum_{n=1}^{+\infty} \frac{\rho^n}{n!}$ converge pour toutes valeurs de ρ , ce qui est cohérent avec la stabilité inconditionnelle de la file. On obtient finalement :

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

Le débit d :

Le service s'effectue avec un taux $n\mu$ dans chaque état où le système contient n clients :

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

On retrouve la stabilité inconditionnelle de la file.

Taux d'utilisation du serveur U :

$$U = 1 - \pi_0$$

Nombre moyen de clients dans le système L :

$$L = \sum_{n=1}^{+\infty} n\pi_n = e^{-\rho} \sum_{n=1}^{+\infty} \frac{\rho^n}{(n-1)!} = e^{-\rho} \rho e^\rho = \rho$$

Temps moyen de séjour W :

Ce paramètre est obtenu en utilisant la loi de Little :

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

Chapitre 3

Les Systèmes De File d'Attente Avec Dérobade

Dans divers domaines, les clients impatients, découragés par la longueur de la file, sont devenus le but de plusieurs études. Ces systèmes qui contiennent des clients impatients ont fait des pertes considérables à l'économie de plusieurs firmes [5].

Dans ce chapitre on s'intéresse aux systèmes de files d'attente avec clients impatients. Cette impatience est due à plusieurs facteurs, soit par insuffisance de nombre de serveurs, soit par une qualité de service médiocre, ou encore une mauvaise gestion du système [18].

Le dérobage : le client qui refuse d'entrer dans la file d'attente s'il voit que la file d'attente est très grande ou le temps d'attente est trop long.

Les files d'attente avec dérobage, abandon (Après un temps passé dans la file, le client impatient décide de quitter le système sans avoir le service) ou les deux ont fait l'intérêt de plusieurs études. Haight, 1957, est le premier qui a étudié le modèle d'attente $M/M/1$ avec dérobage [10].

Systèmes avec dérobage

Supposons qu'un client qui arrive à un système et trouve n clients devant lui, entre avec la probabilité b_n ou part avec la probabilité $q_n = 1 - b_n$. Si la longueur de la file d'attente décourage les clients, alors b_n serait une fonction décroissante de n . Comme un cas particulier, s'il y a une salle d'attente de capacité C , nous pourrions supposer que

$$b_n = \begin{cases} 1, & \text{pour } n < C; \\ 0, & \text{pour } n \geq C. \end{cases}$$

indiquant que, une fois la salle d'attente est remplie, les clients n'ont plus la possibilité d'entrer dans le système [6].

Soit $X(t)$ le nombre de clients dans le système à l'instant t .

3.1 Le Modèle d'attente $M/M/1$ avec dérobade

Le Modèle $M/M/1$ avec dérobade

A son arrivée au système, le client découragé par la longueur de la file d'attente décide de ne pas rejoindre la file, c'est le cas du dérobade. Cette impatience a poussé beaucoup de mathématiciens à l'étude de ce phénomène afin de trouver des solutions [12].

Description du modèle : Supposons qu'un client qui arrive à un système et trouve k clients devant lui, entre avec la probabilité b_k ou part avec la probabilité $q_k = 1 - b_k$. Si la longueur de la file d'attente décourage les clients, alors b_k serait une fonction décroissante de k [22].

Cas particulier : s'il y a une salle d'attente de capacité C , nous pourrions supposer que

$$b_k = \begin{cases} 1 & \text{pour } k < C \\ 0 & \text{pour } k \geq C \end{cases}$$

indiquant que, une fois la salle d'attente est remplie, les clients n'ont plus la possibilité d'entrer dans le système.

Soit $X(t)$ le nombre de clients dans le système à l'instant t , et le processus d'arrivée des clients dans le système est Poissonnier de taux λ [35] et le service est exponentiel de taux μ [34]. Un client qui arrive de l'extérieur et trouve k clients dans le système devant lui entre avec la probabilité b_k alors le processus de naissance appropriés est

$$\lambda_k = \lambda b_k \quad k = 0, 1, \dots$$

telque

$$b_k = \frac{1}{k+1} \quad k = 0, 1, \dots$$

et

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

c'est l'intensité du trafic. ainsi

$$\pi_k = \frac{\rho^k}{k!} \pi_0 \quad k = 0, 1, \dots,$$

(d'après Chapman Kolmogorov)[25]

puis en utilisant la condition de normalisation, nous obtenons

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

La condition de stabilité ici est $E(\delta) < \infty$, telque δ est la période d'occupation du serveur. On n'a pas besoin de la condition $\rho < 1$ comme dans la file $M/M/1$ [2].

Les paramètres de performance sont comme suit :

Taux d'utilisation du serveur U :

$$\begin{aligned} U_s &= 1 - \pi_0 = 1 - e^{-\rho} \\ &= \frac{E(\delta)}{\frac{1}{\lambda} + E(\delta)}, \end{aligned}$$

avec $E(\delta)$ est la période moyenne d'occupation du serveur.

Par conséquent

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

Nombre moyen de clients dans le système L :

$$L = \rho$$

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

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

Temps moyen de séjour W :

$$W = \frac{\rho}{\mu(1 - e^{-\rho})}$$

et on déduit le temps moyen passé dans la file d'attente

$$W_q = W - \frac{1}{\mu} = \frac{1}{\mu} \left(\frac{\rho + e^{-\rho} - 1}{1 - e^{-\rho}} \right)$$

Débit d : on a aussi

$$d = \sum_{k=0}^{\infty} \lambda_k \pi_k = \sum_{k=1}^{\infty} \mu_k \pi_k = \mu(1 - e^{-\rho})$$

d'où

$$d.W = \mu(1 - e^{-\rho}) \cdot \frac{\rho}{\mu(1 - e^{-\rho})} = \rho = L$$

qui est la formule de Little pour ce système [24].

3.2 Le Modèle d'attente $M/M/s$ avec dérobage

Nous considérons une file d'attente $M/M/s$ avec dérobage [1]. Le processus d'arrivée est de Poisson de taux λ et le service a une distribution exponentielle de moyenne $1/\mu$. Un client arrivant de l'extérieur, rejoint la file d'attente si et seulement s'il observe que le temps d'attente ne dépasse pas un temps fixe b .

Soit $N(t)$ le nombre de clients dans le système à l'instant t . La définition de $W(t)$ implique que $W(t) = 0$ si et seulement si $N(t) \leq s - 1$. Si $N = \{N(t), t \geq 0\}$ le processus subit une transition de l'état $s - 1$ à l'état s à l'instant t , alors il y a un saut dans le processus $W = \{W(t), t \geq 0\}$ en même temps. Ceci est illustré par la figure 3.1 qui montre un système à 2 serveurs.

On notera que, dans la trajectoire de l'échantillon représenté, le client qui arrive à l'instant T_3 dérobage. A l'instant T_1 (et T_4), le nombre de clients dans le système augmente de 1 à 2. Dans le même temps, le temps d'attente passe de 0 à un nombre positif. Il est clair que le processus de W se termine un cycle de régénération du T_1 au T_4 . Soit I une variable aléatoire représentant la période d'inactivité définie comme l'intervalle de temps pendant lequel $W(t) = 0$. En d'autres termes, soit t_1 l'instant de la fin du service tels que $N(t_1 -) = s$. Alors,

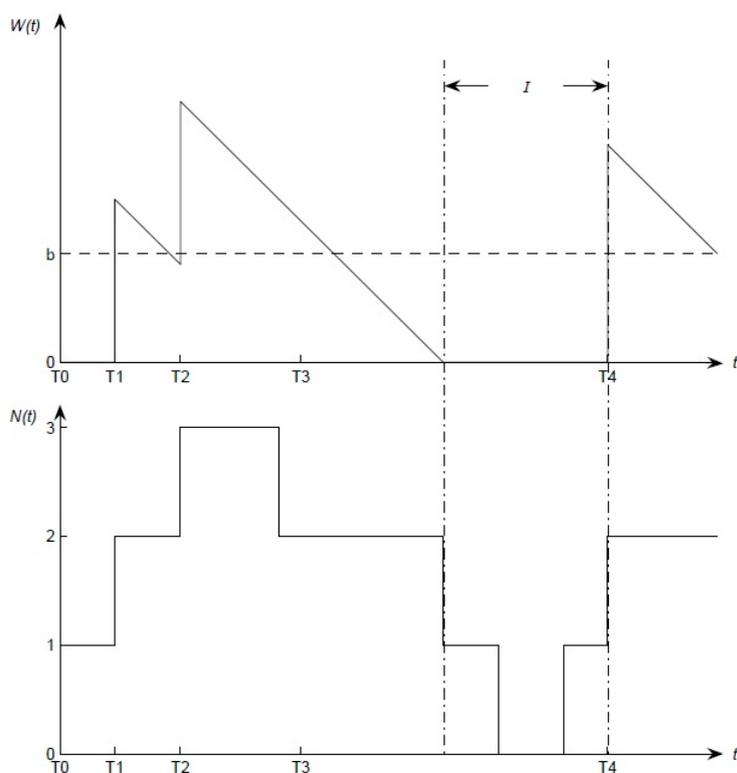
$$I = \min_{t>0} \{t : W(t_1 + t) > 0\}$$

et

$$I = \min_{t>0} \{t : N(t_1 + t) = s\}.$$

Théorème 3.2.1. [33] *Systèmes de files d'attente basé sur le temps d'attente : La durée prévue de I est donnée par*

$$E(I) = \frac{1 - p_s}{s\mu p_s}. \quad (3.1)$$

FIGURE 3.1 – Un chemin d'échantillon de $W(t)$ et $N(t)$ [33]

avec

$$p_s = \frac{(\lambda/\mu)^s}{s!} \sum_{i=0}^s \frac{(\lambda/\mu)^i}{i!}$$

Preuve

Considérons un système $M/M/s/s$ avec le taux d'arrivée λ et la moyenne des temps de service $1/\mu$. De toute évidence, le temps entre deux périodes consécutives lorsque le système est plein a la même distribution que I . La durée moyenne de chaque période du système complet est clairement $1/(s\mu)$ [46]. De la théorie du processus de renouvellement, nous obtenons :

$$\frac{1/(s\mu)}{1/(s\mu) + E(I)} = p_s$$

où p_s est la probabilité que le système $M/M/s/s$ est plein.

Tenir compte du cycle de régénération de T_1 à T_4 , comme il a été illustré sur la figure 3.1. Le cycle se compose d'une période d'activité intense (où $W(t) > 0$) et une

période de repos (où $W(t) = 0$). Il est facile de voir que l'évolution du processus de W pendant la période occupée est stochastiquement identique à celui d'un système avec dérobade à serveur unique avec un taux d'arrivée λ et des temps de service $\exp(s\mu)$ [18]. Soit $\tilde{W}(t)$ le temps d'attente et $\tilde{N}(t)$ le nombre de clients à l'instant t dans ce système. La même règle de dérobade appliquée, à savoir, un client arrivant à l'instant t pénètre si et seulement si $\tilde{W}(t-) \leq b$. Les clients arrivent selon un processus de Poisson de taux d'arrivée $\tilde{\lambda}(t)$ en fonction de $\tilde{N}(t)$ comme suit :

$$\tilde{\lambda}(t) = \begin{cases} \gamma & \text{si } \tilde{N}(t) = 0 \\ \lambda & \text{sinon} \end{cases}$$

avec $\gamma = 1/E(I)$. Désignons la fonction de distribution cumulative (cdf) limitant du processus de W et le processus \tilde{W} comme suit.

$$F(x) = \lim_{t \rightarrow +\infty} Pr\{W(t) \leq x\}, x \geq 0;$$

$$\tilde{F}(x) = \lim_{t \rightarrow +\infty} \mathbf{P}\{\tilde{W}(t) \leq x\}, x \geq 0$$

soit $F(0) = c$ et $\tilde{F}(0) = \tilde{c}$

et soit $f(x) = \frac{dF(x)}{dx}$; $\tilde{f}(x) = \frac{d\tilde{F}(x)}{dx}$; $x > 0$ la pdf.

Notons que pour un système à un seul serveur le temps d'attente est égal au temps d'occupation de ce système [48].

Nous appliquons la même méthode utilisé dans Liu et Kulkarni [33] et les résultats dans le Théorème 3.2.2 qui donne les équations d'équilibre et de normalisation satisfaite par $f(x)$ et le Théorème 3.2.3 qui donne l'expression de $\tilde{f}(x)$ de faon explicite en résolvant les équations.

Théorème 3.2.2. [33] *La fonction de distribution des probabilité à l'état d'équilibre de $\tilde{f}(x)$ du processus \tilde{W} satisfait :*

$$\tilde{f}(x) = \lambda \int_0^{x \wedge b} f(u) e^{-s\mu(x-u)} du + \tilde{c}\gamma e^{-s\mu x},$$

$$\int_0^{\infty} \tilde{f}(x) dx + \tilde{c} = 1$$

ou $x \wedge b = \min(x, b)$ on pose $\rho = \frac{\lambda}{s\mu}$ l'intensité du trafic.

Théorème 3.2.3. [33]

La fonction de distribution des probabilité du processus \widetilde{W} est :

$$\tilde{f}(x) = \begin{cases} \tilde{c}\gamma e^{-(s\mu-\lambda)x} & \text{si } 0 < x < b \\ \tilde{c}\gamma e^{\lambda b} e^{-s\mu x} & \text{si } x \geq b. \end{cases} \quad (3.2)$$

Avec

$$\tilde{c} = \begin{cases} \left[\gamma \left(\frac{1}{s\mu-\lambda} - e^{-(s\mu-\lambda)b} \frac{\lambda}{(s\mu-\lambda)s\mu} \right) + 1 \right]^{-1} & \text{si } \rho \neq 1 \\ \frac{\lambda}{\lambda+\gamma+\lambda\gamma b} & \text{si } \rho = 1 \end{cases} \quad (3.3)$$

Conclusion générale

Dans ce mémoire nous nous sommes intéressés aux systèmes de files d'attente avec dérobade.

Le premier chapitre était un rappel des Processus Stochastiques qui sont un outil incontournable dans l'analyse dans différents systèmes dynamiques (réseaux de files d'attente, systèmes de communication et informatiques, biologiques, économiques...etc).

Dans le deuxième chapitre, nous avons introduit les notions de base de la théorie des files d'attente. En suite nous avons traité certains modèles de files d'attente classiques.

Le troisième chapitre a été consacré à l'étude de deux modèles d'attente avec dérobade, le cas de files d'attente $M/M/1$ avec dérobade et le modèle $M/M/s$ avec dérobade.

Bibliographie

- [1] Abou El-Ata, M. O. and Hariri, A. M. A. The $M/M/c/N$ queue with balking and renegeing. *Computers and Operations Research*. 19 (1992), No. 13 713-716.
- [2] Abdel-Karim Aboul-Hassan, Sherif I. Rabia and Ahmed Kadry, Analytical study of a discrete time retrial queue with balking customers and early arrival scheme, *Alexandria Engineering Journal*, 44 (2005), No. 6, 911-917
- [3] Allen, A. O. 1990. *Probability, Statistics, and Queueing Theory with Computer Science Applications*. Second edition, Academic Press, New York (First edition : 1978).
- [4] Anisimov, V. V., Zakusilo, O. K., and Donchenko, V. S. 1987. *Elements of Queueing Theory and Asymptotic System Analysis*. Vishcha Shkola, Kiev (in Russian).
- [5] Ancker, Jr., C. J. and Gafarian, A. V. Some queuing problems with balking and renegeing : I. *Operations Research*. 11 (1963), No. 1, 88-100.
- [6] Ancker, Jr., C. J. and Gafarian, A. V. Some queuing problems with balking and renegeing : II. *Operations Research*. 11 (1963), No. 6, 928-937.
- [7] Arnaud GUYADER , *Processus Markoviens de sauts* , Université de Rennes 2006/2007.
- [8] Baccelli, F., and Bremaud, P. 1990. *Mathematical Theory of Queues*. Springer-Verlag, Berlin.
- [9] Bagchi, T. P., and Templeton, J. P. C. 1972. *Numerical Methods in Markov Chains and Bulk Queues*. Lecture Notes in Economics and Mathematical Systems 72, Springer-Verlag, New York.
- [10] A. A. Bouchentouf, M. Kadi, A. Rabhi 2013 Analysis of Two Heterogeneous Server Queueing Model With Balking, Reneging and Feedback. *Mathematical Sciences And Applications E-Notes Volume 2 No. 2 pp. 10-21 (2013)*

-
- [11] Chaudhry, M. L., and Templeton, J. G. C. 1983. A First Course in Bulk Queues. John Wiley and Sons, New York.
- [12] Chapel HILL , Service Systems With Balking On Queueing Time, page67, 2007.
- [13] Cheprasov, V. P. 1985. Elements of Queueing Theory. Kazan Aviation Institute (in Russian).
- [14] Claudie Chabriac. Processus stochastiques et modélisation. (2012-2013).
- [15] Claudie Hasseforder CHABRIAC , Eléments de Théorie des files d'attente, page05, Janvier 2008.
- [16] Fedosejev, Yu. N. 1982. Methods of Queueing Systems Analysis. Moscow Institute of Engineering and Physics (in Russian).
- [17] Frédéric Sur. Programmation dynamique, chaînes de Markov, files d'attente. école des Mines de Nancy (2013)-(2014).
- [18] O. Garnett, A. Mandelbaum, and M. Reiman. Designing a call center with impatient customers. Manufacturing & Service Operations Management, 4(3) :208-227, June 2002.
- [19] Gelenbe, E.,and Pujolle, G. 1987. Introduction to Queuing Networks. John Wiley and Sons, Chichester, England (French Original : Introduction aux réseaux de files d'attente, Edition Eyrolles ; Paris).
- [20] A. Gomez-Corral and M.F. Ramalhoto. On the waiting time distribution and the busy period of a retrial queue with constant retrial rate. Stochastic Modelling and Applications, 3, 37-47, (2000).
- [21] Gross,D.and C.M,Harris,Fundamentals of Queueing Theory ,Wilcy,New york,1975 1984.
- [22] Haight, F. A., 1957, Queueing with balking, I, Biometrika, 44, 360-369.
- [23] Jean Louis Poss, Probabilité et statistique version 2.1 Mai 2003 p74.
- [24] J. Lubacz and J. Roberts. A new approach to the single-server repeated attempt system with balking. Proceedings of the Third Int. Seminar on Teletraffic Theory. Moscow, 290-293, (1984).
- [25] Kangzhou Wang, Na Li, and Zhibin Jiang. Queueing system with impatient customers : A review. In 2010 IEEE International Conference on Service Operations and Logistics and Informatics (SOLI), pages 82-87, 2010.

- [26] Kashyap, B. R. K., and Chaudhry, M. L. 1988. An Introduction to Queueing Theory. Aarkay, Calcutta, India.
- [27] K.B, GK, Techniques de modélisation : Méthodes analytiques.
- [28] Keilson, J. 1979. Markov Chain Models - Rarity and Exponentiality. Springer-Verlag, New York.
- [29] Khintchine, A. Y 1969. Mathematical Methods in the Theory of Queueing. Second edition, Hafner Publishing Company, New York (First edition : Griffin, London, 1960 ; Russian original : 1955).
859-866.
- [30] Kumar, R. and Sharma, S.K. An $M/M/1/N$ Queueing Model with Retention of reneged customers and Balking, American Journal of Operational Research. 2(2012), No. 1, 1-5.
- [31] LeGall, P. (1962). Les systemes avec ou sans attente et les processus stochastiques, Tome I. Dunod, Paris.
- [32] Libman, L. and Orda, A. (2002). Optimal retrial and timeout strategies for accessing network resources. IEEE/ACM Transactions on Networking, 10(4), 551-564.
- [33] Liqiang Liu. (2007) Service Systems with Balking Based on Queueing Time. Chapel Hill Under the direction of Dr. Vidyadhar G. Kulkarni
- [34] J. Lubacz and J. Roberts. A new approach to the single-server repeated attempt system with balking. Proceedings of the Third Int. Seminar on Teletraffic Theory. Moscow, 290-293, (1984).
- [35] Mahdy S. El-Paoumy and Hossam A. Nabwey, The Poissonian Queue with Balking Function, Reneging and two Heterogeneous Servers. International Journal of Basic & Applied Sciences IJBAS-IJENS. 11 (2011), No. 6, 149-152.
- [36] Moulay Hachemi , Files d'attente et applications ,page 12 , 2014/2015.
- [37] M.Petito , Introduction à la modélisation des réseaux, page21, 26 Octobre 2010.
- [38] M.Youcef TALEB , Principes d'invariace dans les systèmes de files d'attente avec Rappels, page05, 2013.
- [39] Neuts, M. F. 1989. Structured Stochastic Matrices of $M/G/I$ Type and Their Applications. Marcel Dekker, Inc., New York.
- [40] Newell, G. F. 1982. Applications of Queueing Theory. Second edition, Chapman and Hall, London (First edition : 1971).

-
- [41] Niles BERGLUND , Processus aléatoires et applications, page 78, Janvier 2014.
- [42] Olivier Scaillet , Processus de naissance et de Poisson , page 27 , University of Geneva.
- [43] G. Rubino , Processus Stochastiques , Février 2006.
- [44] Saaty, T. L. (1961). Elements of queueing theory with applications. New York : McGraw Hill. Thomopoulos, N. T. (1990). Strategic inventory management and planning. Carol Stream : Hitchcock Publishing Co.
- [45] Statistiques des processus 3A, 2 novembre 2015.
- [46] Tkachenko Andrey. Multichannel queueing systems with balking and regenerative input flow, Basic research program, serie : Science, Technologie and innovation, WP BRP 14/STI/2013.
- [47] van Doorn, E. 1981. Stochastic Monotonicity and Queueing Applications of Birth-Death Processes. Lecture Notes in Statistics 4, Springer-Verlag, New York,
- [48] Wang, K-H. and Chang, Y-C. Cost analysis of a finite $M/M/R$ queueing system with balking, reneging and server breakdowns. Mathematical Methods of Operations Research 56 (2) (2002) 169-180.
- [49] Y. Velenik, Probabilités et Statistique, université de Genève, Version du (24 mai 2012), pp 195-202
- [50] Yves Caumel. Probabilités et processus stochastiques. Springer. Verlag France, (2011).
- [51] Zakhar Kabluchko, Stochastic Processes (Stochastik II), University of Ulm Institute of Stochastics, (2013-2014).