Ministère de l'Enseignement Supèrieur et de la Recherche Scientifique

Université Dr Tahar Moulay Saida

Faculté des Sciences

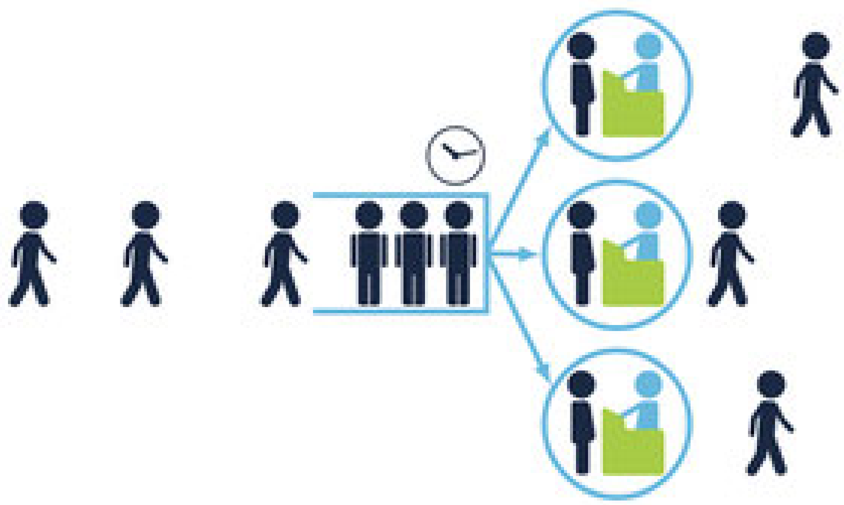
Département des Mathématiques

POLYCOPIE SUR

RÉSEAUX DE FILES D'ATTENTE STOCHASTIQUES

cours et exercices corrigés

M2 Probabilités et Applications



Présenté par

Dr **M**okhtar **K**adi

kadi1969@yahoo.fr

Table des matières

Préambule

Dans ce polycopie nous s'intéressons aux réseaux de files d'attente stochastiques, il est destiné aux étudiants de Master 2 spécialité probabilités et applications. Ce cour a pour but d'initier les étudiants à un domaine d'application des probabilités qui est les réseaux de files d'attente.

L'idé des réseaux a commencé vers les années 1900 sur des réseaux de type téléphonique 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 et d'autre ingénieurs Engset, Palm, . . . , qui ont contribuent principalement à ce domaine de recherche.

Les réseaux de files d'attente ont été intégrés dans la modélisation de divers domaines d'activité dans les années 1960. C'est l'échange de données entre ordinateurs qui a poussé les recherches dans le domaine de la théorie des réseaux de files d'attente, notamment Les problèmes concernent les temps de traitements de requêtes sur un système centralisé ou encore les délais, le taux d'occupation des différents nœuds du réseau.

L'aparution de l'Internet et son dévelopement entre les années 70-90 ont poussé les mathématisiens à essayer de modéliser ces réseaux.

Des techniques de renormalisation ont été introduites au début des années 1990, dans le cadre d'étudier les probabilités invariantes des grands réseaux avec perte. Les méthodes de convergence de processus, de calcul stochastique ont fait ainsi progressivement leur entrée dans l'étude de ces réseaux.

Par la suite, on assista alors à une évolution rapide de la modélisation des réseaux qu'on appliqua à l'évaluation des performances des systèmes informatiques et aux réseaux de communication.

Le polycopie est partagé comme suit :

Dans le premier chapitre nous abordons les processus à la base de l'étude de tels systèmes d'attente qui sont les processus stochastiques. Nous présentons les concepts de base des processus stochastiques en temps discret et en temps continu. Nous présentons également les propriétés fondamentales des chaînes de Markov (l'homogénéité, la périodicité, irréductibilité, ...). Ensuite nous donnons certaines relations entre les chaînes de Markov des états finis et la théorie des matrices. Enfin nous classons les différents états des chaînes de Markov (états communiquants, états périodiques, ...

Ensuite dans le deuxième chapitre, nous abordons la terminologie de la théorie des files d'attente. Certaines définitions et notations qui sont nécessaires dans l'étude des systèmes de files d'attente (la notation de KANDELL, la formule de LITTLE ...) sont nottamment données. Et nous étudions quelque modèles de files d'attente markoviennes ($M/M/1$, $M/M/1/K$, $M/M/c$, $M/M/\infty$,) et non markovienne ($M/G/1$) et l'évaluation de leurs paramètres de performance. Quelques exercices sur les files d'attente et leurs applications.

Enfin, dans le troisième chapitre, qui a été inspiré par le document de Philippe Robert, nous donnons quelques exemples sur les réseaux de files d'attente (réseaux de Jackson, le Réseau de Rybko-Stolyar/Lu-Kumar, le Réseau de Bramson et les réseaux Gordon-Newel). Et nous terminons ce chapitre par introduire les méthodes de renormalisation et les limites fluides. En suite les principales définitions et notions relatives à la renormalisation d'un processus sont traitées. Enfin les limites fluides notamment seront données, qui sont les limites des processus renormalisés. et on termine le chapitre par des exercices illustrent ces définitions.

Chapitre 1

Processus stochastiques et chaînes de Markov

Dans ce chapitre, nous allons examiné certaines variables aléatoires courantes utilisées pour l'étude des phénomènes liés aux temps, des phénomènes que l'on rencontre quotidiennement dans de très nombreux domaines et sous diverses formes. Le temps entre les messages dans un système informatiques, le temps passé dans un centre de péage dans une autoroute, le temps séparent le passage de clients dans un guichet de poste, gestion d'un stock de production, sont des exemples sur ces phénomènes. Tous ces phénomènes et beaucoup d'autres peuvent être traités par des processus connus qui sont les processus stochastiques afin de facilité l'étude et d'optimiser la rentabilité du système.

1.1 Processus stochastiques

Une famille de variables aléatoires $\{X(t), t \in T\}$ définies sur le même espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$ et à valeurs dans un espace d'états E, est appelée processus stochastique. Pour chaque

$t \in T$, où T est l'ensemble d'indices du processus, $X(t)$ est une variable aléatoire. Cette variable représente l'état du processus au temps t. Un élément de T est généralement un paramètre de temps. L'espace d'états du processus est l'ensemble de toutes les valeurs possibles que les variables aléatoires $X(t)$ peuvent prendre. Chacune de ces valeurs est appelée un état du processus.

1.1.1 Définitions et propriétés de base

Définition 1.1.1. *[22] Un processus stochastique est une famille de variables aléatoires $X(t), t \in T$ où chaque variable aléatoire $X(t) \in E$ 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.

- *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 (suite stochastique) notée $(X_n)_{n\in\mathbb{N}}$*

  .

- *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'états, 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.

En Générale les valeurs de l'espace d'états du processus et l'ensemble d'indices du processus, donnent quatre cas possibles :

- processus à temps discret et à espace d'états discret.
- processus à temps continu et à espace d'états discret.

- processus à temps discret et à espace d'états continu.
- processus à temps continu et à espace d'états continu.

1.1.2 Matrices stochastiques

Définition 1.1.2. *[23] Une matrice stochastique est une matrice* (a_{ij}) *avec* $i, j \in \{1, 2, \cdots r\}$ *telle que :*

$$\begin{aligned} a_{ij} &\geq 0 \\ \sum_{j=1}^{r} a_{ij} &= 1 \quad \forall i \in \{1, 2, \cdots, r\} \end{aligned}$$

Théorème 1.1.2.1. *[23] Le produit de deux matrices stochastiques est une matrice stochastique*

Démonstration. Supposons que les matrices A et B sont stochastiques et posons $C = AB$. On a

$$\forall i, j \in \{1, 2, \cdots, r\}, \qquad c_{ij} = \sum_{k=1}^{r} a_{ik} b_{kj} \geq 0,$$

et

$$\begin{aligned} \forall i, j \in \{1, 2, \cdots, r\}, \qquad \sum_{k=1}^{r} c_{ij} &= \sum_{j=1}^{r} \sum_{k=1}^{r} a_{ik} b_{kj} \\ &= \sum_{k=1}^{r} a_{ik} \underbrace{\sum_{j=1}^{r} b_{kj}}_{=1, \forall k} \\ &= \sum_{k=1}^{r} a_{ik} = 1. \end{aligned}$$

En général la matrice A^n de transition en n étapes est stochastique.

1.1.3 Processus stochastiques en temps continu

Les arrivées des clients à un système de file d'attente sont des phénomènes qui peuvent être caractérisées par l'ensemble des instants $(A_n)_{n\in\mathbb{N}^*}$ ou dates d'arrivée de chaque client qui sont des variables aléatoires. Mais on peut aussi le faire à partir du processus de comptage (qui va être définit dans la suite) $(N_t)_{t\in\mathbb{R}_+}$, ou par la famille $(T_n)_{n\in\mathbb{N}^*}$. des intervalles de temps entre deux arrivées. La collection de ces dates d'arrivée peut être modéliser par le processus de Poisson.

Définitions et propriétés élémentaires

Il y a trois façons naturelles de décrire un tel processus

- On peut, tout d'abord, encoder une réalisation d'un tel processus par une collection $0 < T_1(\omega) < T_2(\omega) < \ldots$ de nombres réels positifs, correspondant à la position des points sur $\mathbb{R}_+$. Il est pratique de poser également $T_0 = 0$.
- Une seconde façon de coder une réalisation revient à donner, pour chaque intervalle de la forme $I = (t, t+s]$, le nombre de points $N_I(\omega)$ contenus dans l'intervalle. Si l'on utilise la notation simplifiée $N_t = N_{(0;t]}$, on aura alors $N_{(t;t+s]} = N_{t+s} - N_t$. La relation entre les variables aléatoires T_n et N_t est donc simplement

$$N_t(\omega) = sup\{n \geq 0 : T_n(\omega) \leq t\}, \quad T_n(\omega) = inf\{t \geq 0 : N_t(\omega) \leq n\}.$$

- Une troisième façon naturelle d'encoder cette information est de considérer la suite $X_1(\omega); X_2(\omega); X_3(\omega), \ldots$ de nombres réels positifs correspondant aux distances successives entre deux points. La relation entre ces variables et les T_n est donnée par

$$X_k = T_k - T_{k-1}; \quad T_k = \sum_{i=1}^{k} X_i.$$

1.1.4 Processus de comptage

Définition 1.1.4.1. *[38] Un processus de comptage, est un processus stochastique à valeurs dans* $\mathbb{N}$*, l'espace des entiers naturels. Le processus de comptage* $(N(t))_{t\in\mathbb{R}_+}$ *représente le nombre total d'événements qui se sont produits entre* 0 *et* t *vérifiant :*

(i) $N(t) \in \mathbb{N}, \quad \forall t \in \mathbb{R}_+,$

(ii) $\forall t > s \in T, N(t) \geq N(s),$

(iii) $N(t)-N(s)$*, représente le nombre d'évènements se produisant dans l'intervalle* $[s,t], \quad \forall s < t \in T$.

Soient $X_1, X_2, \ldots$ une suite de variables aléatoires satisfaisant $\mathbb{P}(X_k > 0) = 1$ pour tout $k \geq 1$. Soit $T_0 = 0$ et $T_n = \sum_{i=1}^{n} X_i$. Finalement, posons $N_t = sup\{n \geq 0 : T_n \leq t\}$. Le processus $(N_t)_{t\geq 0}$ est appelé processus de comptage.

Remarque 1.1.4.1. *On supposera toujours par la suite qu'un processus de comptage satisfait presque sûrement* $T_n \to \infty$ *lorsque* $n \to \infty$.

On appelle souvent les variables aléatoires X_n les temps d'attente ou durées de vie du processus $(N_t)_{t\geq 1}$.

Le cas le plus simple, mais très important, est celui où les temps d'attente forment un processus i.i.d.

Définition 1.1.4.2. *[38] Un processus de comptage pour lequel les temps d'attente sont i.i.d. est appelé processus de renouvellement.*

Le processus de Poisson est l'exemple le plus important de processus de renouvellement.

1.1.5 Processus de Poisson

Définition 1.1.5.1. *[12] Un processus de Poisson d'intensité* λ *est un processus de renouvellement dont les durées de vie suivent une loi* $exp(\lambda)$*. On note* $\mathbb{P}\lambda$ *la loi du processus de Poisson* *d'intensité λ.*

Théorème 1.1.5.1. *[12] Soit $(N_t)_{t\geq 1}$ un processus de Poisson d'intensité λ. Alors, pour tout $t, s \geq 0$,*

1. *$N_{t+s} - N_t$ suit la même loi que N_s.*
2. *$\forall 0 < t_1 < t_2 < \ldots < t_n$, les variables aléatoires $(N_{t_{i+1}} - N_{t_i})_{i=1,\ldots,n-1}$ sont indépendantes.*

Démonstration. Soit $t > 0$ fixé. Notons $X_1^t = T_{N_{t+1}} - t$ le temps restant après t jusqu'au point suivant du processus, $X_k^t = X_{N_{t+k}}, k \geq 2$, et $T_k^t = X_1^t + \ldots + X_k^t, k \geq 1$. Évidemment,

$$N_{t+s} - N_t = n \Leftrightarrow T_n^t \leq s < T_{n+1}^t.$$

Observons à présent que l'indépendance de X_{n+1} et de T_n implique que, pour tout $x > 0$,

$$\begin{aligned}
\mathbb{P}_\lambda\left(X_1^t > x | N_t = n\right)\mathbb{P}_\lambda\left(N_t = n\right) &= \mathbb{P}_\lambda\left(X_1^t > x, T_n \leq t < T_{n+1}\right) \\
&= \mathbb{P}_\lambda\left(T_n \leq t, X_{n+1} > t + x - T_n\right) \\
&= \int_0^t dy \int_{t+x-y}^{\infty} dz f_{(T_n, X_{n+1})}(y, z) \\
&= \int_0^t dy f_{T_n}(y) \int_{t+x-y}^{\infty} dz f_{X_{n+1}}(z) \\
&= \int_0^t \mathbb{P}_\lambda\left(X_{n+1} > t + x - y\right) f_{T_n}(y) dy \\
&= e^{-\lambda x} \int_0^t \mathbb{P}_\lambda\left(X_{n+1} > t - y\right) f_{T_n}(y) dy \\
&= e^{-\lambda x}\mathbb{P}_\lambda\left(T_n \leq t, X_{n+1} > t - T_n\right) \\
&= e^{-\lambda x}\mathbb{P}_\lambda\left(N_t = n\right). \qquad (1.1)
\end{aligned}$$

En procédant de la même façon, on voit que

$$\begin{aligned}
\mathbb{P}_\lambda\left(X_1^t > x_2, \dots, X_k^t > x_k | N_t = n\right) &= \mathbb{P}_\lambda\left(T_n \leq t, X_{n+1} > t + x_1 - T_n, X_{n+2} > x_2, \dots, X_{n+k} > x_k\right) / \\
\mathbb{P}_\lambda\left(N_t = n\right) & \\
&= \prod_{l=2}^{k} \mathbb{P}_\lambda\left(T_n \leq t, X_{n+1} > t + x_1 - T_n\right) / \mathbb{P}_\lambda\left(N_t = n\right) \\
&= e^{-\lambda(x_1 + \dots + x_k)}.
\end{aligned}$$

La seconde identité suivant de l'indépendance de T_n, $X_{n+1}, \dots, X_{n+k}$, et la dernière identité de (1.1). On en déduit que, conditionnellement à $N_t = n$, les variables aléatoires $X_t^k, k \geq 1$, sont des variables aléatoires i.i.d. de loi $exp(\lambda)$. Par conséquent, la loi conjointe des variables aléatoires $T_t^k, k \geq 1$, sous $\mathbb{P}_\lambda(.|Nt = n)$ coïncide avec celle des variables aléatoires $T_k, k \geq 1$, sous $\mathbb{P}_\lambda$. On a donc

$$\begin{aligned}
\mathbb{P}_\lambda\left(N_{t+s} - N_t = n\right) &= \sum_{n \geq 0} \mathbb{P}_\lambda\left(N_{t+s} = n + k | N_t = n\right) \mathbb{P}_\lambda\left(N_t = n\right) \\
&= \sum_{n \geq 0} \mathbb{P}_\lambda\left(T_k^t \leq s < T_{k+1}^t | N_t = n\right) \mathbb{P}_\lambda\left(N_t = n\right) \\
&= \sum_{n \geq 0} \mathbb{P}_\lambda\left(T_k \leq s < T_{k+1}\right) \mathbb{P}_\lambda\left(N_t = n\right) \\
&= \mathbb{P}_\lambda\left(T_k \leq s < T_{k+1}\right) \\
&= \mathbb{P}_\lambda\left(N_s = k\right).
\end{aligned}$$

Passons à la seconde affirmation. Les arguments ci-dessus montrent que la loi conjointe des variables aléatoires $N_s - N_t = \max\left\{n \geq 0 : T_n^t \leq s - t\right\}$

sous $\mathbb{P}_\lambda\left(.|N_t = l\right)$ coïncide avec celle des variables aléatoires $N_{s-t} = \max\left\{n \geq 0 : T_n \leq s - t\right\}$ sous $\mathbb{P}_\lambda$. Posons $m_i = k_1 + \cdots + k_i, i \geq 1$. On a alors

$$\begin{aligned}
\mathbb{P}_\lambda\left(N_{t_{i+1}} - N_{t_i} = k_i, i = 1, \ldots, n-1\right) &= \sum_{l \geq 0} \mathbb{P}_\lambda\left(N_{t_{i+1}} - N_{t_i} = k_i, i = 1, \ldots, n-1 | N_t = l\right) \mathbb{P}_\lambda\left(N_{t_1} = l\right) \\
&= \sum_{l \geq 0} \mathbb{P}_\lambda\left(N_{t_{i+1}} - N_{t_i} = m_i, i = 1, \ldots, n-1 | N_t = l\right) \mathbb{P}_\lambda\left(N_{t_1} = l\right) \\
&= \sum_{l \geq 0} \mathbb{P}_\lambda\left(N_{t_{i+1}-t_1} = m_i, i = 1, \ldots, n-1\right) \mathbb{P}_\lambda\left(N_{t_1} = l\right) \\
&= \mathbb{P}_\lambda\left(N_{t_{i+1}-t_1} = m_i, i = 1, \ldots, n-1\right) \\
&= \mathbb{P}_\lambda\left(N_{t_2-t_1} = k_1, N_{t_{i+1}-t_1} - N_{t_2-t_1} = m_i - m_1, i = 2, \ldots, n-1\right).
\end{aligned}$$

De la même façon, puisque $t_{i+1} - t_1 - (t_2 - t_1) = t_{i+1} - t_2$,

$$\mathbb{P}_\lambda\left(N_{t_2-t_1} = k_1, N_{t_{i+1}-t_1} - N_{t_2-t_1} = m_i - m_1, i = 2, \ldots, n-1\right)$$

$$\begin{aligned}
&= \mathbb{P}_\lambda\left(N_{t_{i+1}} - N_{t_1} = m_i - m_1, i = 2, \ldots, n-1 | N_{t_2-t_1} = k_1\right) \times \mathbb{P}_\lambda\left(N_{t_2-t_1} = k_1\right). \\
&= \mathbb{P}_\lambda\left(N_{t_{i+1}-t_2} = m_i - m_1, i = 2, \ldots, n-1\right) \mathbb{P}_\lambda\left(N_{t_2-t_1} = k_1\right). \qquad (1.2)
\end{aligned}$$

Mais

$$\mathbb{P}_\lambda\left(N_{t_{i+1}-t_2} = m_i - m_1, i = 2, \ldots, n-1\right)$$
$$= \mathbb{P}_\lambda\left(N_{t_3-t_2} = k_2, N_{t_{i+1}-t_2} - N_{t_3-t_2} = m_i - m_2, i = 2, \ldots, n-1\right),$$

et l'on peut donc répéter la procédure (1.2), pour obtenir finalement

$$\begin{aligned}
\mathbb{P}_\lambda\left(N_{t_{i+1}} - N_{t_i} = k_i, i = 1, \ldots, n-1\right) &= \prod_{i=1}^{n-1} \mathbb{P}_\lambda\left(N_{t_{i+1}-t_i} = k_i\right) \\
&= \prod_{i=1}^{n-1} \mathbb{P}_\lambda\left(N_{t_{i+1}} - N_{t_i} = k_i\right),
\end{aligned}$$

la dernière identité résulte de la première partie du Théorème.

Proposition 1.1.5.1. *[12] Soit $(N_t)_{t \geq 0}$ un processus de Poisson d'intensité λ. Alors, T_n suit une loi gamma (λ, n),*

$$f_{T_n}(x) = \frac{1}{(n-1)!} \lambda^n x^{n-1} e^{-\lambda x} \mathbb{1}[0; \infty)(x).$$

Démonstration. T_n est une somme de n variables aléatoires i.i.d. de loi $exp(\lambda)$. Manifestement, T_1 suit une loi $exp(\lambda)$, et celle-ci coïncide avec la loi $gamma(\lambda, 1)$. On procède par récurrence. Supposons l'énoncé vrai pour T_n. On a alors,

$$\begin{aligned}
f_{T_{n+1}}(x) &= f_{T_n+X_{n+1}}(x) \\
&= \int_{-\infty}^{+\infty} f_{T_n}(u) f_{X_{n+1}}(x-u) du \\
&= \int_0^x \frac{1}{(n-1)!} \lambda^n u^{n-1} e^{-\lambda u} \lambda e^{-\lambda(x-u)} du \\
&= \frac{\lambda^{n+1}}{(n-1)!} e^{-\lambda x} \int_0^x u^{n-1} du \\
&= \frac{\lambda^{n+1}}{n!} e^{-\lambda x} x^n.
\end{aligned}$$

Théorème 1.1.5.2. *[12] Soit $(N_t)_{t\geq 0}$ un processus de Poisson d'intensité λ. Alors, pour tout $t \geq s \geq 0$, $N_t - N_s$ suit une loi poisson $(\lambda(t-s))$.*

Démonstration. Il suit du Théorème 1.1.5.1 qu'il suffit de considérer le cas $s = 0$. Puisque $N_t = n \Leftrightarrow T_n \leq t < T_{n+1}$, on a immédiatement

$$\begin{aligned}
\mathbb{P}_\lambda (N_t = n) &= \mathbb{P}_\lambda (T_n \leq t < T_{n+1}) \\
&= \mathbb{P}_\lambda (T_{n+1} > t) - (T_n > t) \\
&= \frac{\lambda^n}{n!} \int_t^{+\infty} \left(x^n \lambda e^{-\lambda} - n x^{n-1} e^{-\lambda x} \right) dx \\
&= \frac{\lambda^n}{n!} \int_t^{+\infty} \frac{d}{dx} \left(-x^n e^{-\lambda x} \right) dx \\
&= \frac{\lambda^n}{n!} t^n e^{-\lambda x}.
\end{aligned}$$

Les Théorèmes 1.1.5.1 et 1.1.5.2 montrent que les accroissements $N_{t+s} - N_t$ d'un processus de Poisson d'intensité λ sont stationnaires, indépendants et suivent une loi de Poisson de paramètre λs. Nous allons montrer que ces propriétés caractérisent ce processus. Ceci fournit donc une définition alternative du processus de Poisson.

Théorème 1.1.5.3. *[12] Un processus de comptage $(N_t)_{t\geq 0}$ est un processus de Poisson d'intensité λ si et seulement si ses accroissements $N_{t+s} - N_t$ sont stationnaires et indépendants, et suivent une loi poisson (λs).*

Remarque 1.1.5.1. *[12] En fait, on peut montrer assez facilement qu'un processus de comptage $(N_t)_{t\geq 0}$ est un processus de Poisson (d'intensité non spécifiée) si et seulement si ses accroissements $N_{t+s} - N_t$ sont stationnaires et indépendants. Cela montre que ce processus va correctement modéliser toutes les situations où ces deux hypothèses sont approximativement vérifiées.*

Démonstration. On a déjà montré que le processus de Poisson possède les propriétés énoncées. Montrons donc que ces propriétés caractérisent ce processus. Fixons $0 = t_0 \leq s_1 \leq t_1 \leq s_2 \leq t_2 \leq \cdots \leq s_n \leq t_n$. En observant que $T_1 \in (s_1, t_1], \ldots, T_n \in (s_n, t_n]$ si et seulement si

- $N_{s_i} - N_{t_{i-1}} = 0, 1 \leq i \leq n$, (avec $t_0 = 0$),
- $N_{t_i} - N_{s_i} = 1, 1 \leq i < n$,
- $N_{t_n} - N_{s_n} \geq 1$,

et en utilisant les hypothèses sur les accroissements, on obtient

$$\begin{aligned}
\mathbb{P}\left(T_1 \in (s_1, t_1], \ldots, T_n \in (s_n, t_n]\right) &= \prod_{i=1}^{n} \mathbb{P}\left(N_{s_i} - N_{t_{i-1}} = 0\right) \prod_{i=1}^{n-1} \mathbb{P}\left(N_{t_i} - N_{s_i} = 1\right) \mathbb{P}\left(N_{t_n} - N_{s_n} \geq 1\right) \\
&= \prod_{i=1}^{n} e^{-\lambda}(s_i - t_{i-1}) \prod_{i=1}^{n-1} \lambda(t_i - s_i) e^{-\lambda(t_i - s_i)} \left(1 - e^{-\lambda(t_n - s_n)}\right) \\
&= \lambda^{n-1}\left(e^{-\lambda s_n} - e^{-\lambda t_n}\right) \prod_{i=1}^{n-1} (t_i - s_i) \\
&= \int_{s_1}^{t_1} \cdots \int_{s_n}^{t_n} \lambda^n e^{-\lambda u_n} du_n \ldots du_1.
\end{aligned}$$

La loi conjointe de $(T_1, \ldots, T_n)$ possède donc la densité

$$f_{(T_1,\ldots,T_n)}(u_1, \ldots, u_n) = \begin{cases} \lambda^n e^{-\lambda u_n} & \text{si} \quad 0 < u_1 < \cdots < u_n, \\ 0 & \text{sinon.} \end{cases}$$

Déterminons à présent la densité de la loi conjointe de $(X_1, \ldots, X_n)$. La fonction de répartition conjointe est donnée par

$$\begin{aligned} \mathbb{P}\left(X_1 \leq x_1, \ldots, X_n \leq x_n\right) &= \mathbb{P}\left(T_1 \leq x_1, T_2 - T_1 \leq x_2 \ldots, T_n - T_{n-1} \leq x_n\right) \\ &= \int_0^{x_1} \int_{u_1}^{u_1+x_2} \cdots \int_{u_{n-1}}^{u_{n-1}+x_n} f_{(T_1,\ldots,T_n)}(u_1, \ldots, u_n) du_n \ldots du_1, \end{aligned}$$

et la densité conjointe est donc donnée par

$$\begin{aligned} \frac{\partial^n}{\partial x_1 \cdots \partial x_n} \mathbb{P}\left(X_1 \leq x_1, \ldots, X_n \leq x_n\right) &= f_{(T_1,\ldots,T_n)}(x_1, x_1 + x_2, \ldots, x_1 + \cdots + x_n) \\ &= \lambda^n e^{-\lambda(x_1+\cdots+x_n)}. \end{aligned}$$

On reconnait la densité conjointe de n variables aléatoires i.i.d. de loi $exp(\lambda)$.

1.1.6 Processus non homogènes

Le taux λ dans un processus de Poisson $X(t)$ est la constante de proportionnalité de la probabilité qu'un évènement se produise au cours d'un petit intervalle arbitraire, plus précisément,

$$\begin{aligned} P\{X(t+h)-X(t)=1\} &= \frac{(\lambda h)e^{-\lambda h}}{1!} \\ &= (\lambda h)\left(1-\lambda h+\frac{1}{2}\lambda^2h^2-\cdots\right) \\ &= \lambda h + o(h) \end{aligned}$$

où $o(h)$ désigne un terme général et sans précision d'ordre inférieur à h.

1.2 Processus de Markov à temps continus

Dans cette section, nous introduisons la définition et les principaux Théorèmes pour les processus de Markov continus.

Un processus de Markov continu est un processus aléatoire $(X_t)_{t\in T}$ indexé par un un espace continu. Dans notre cas, il sera indexé par $\mathbb{R}^+ (T=\mathbb{R}^+)$

Définition 1.2.1. *[12] Un processus stochastique $(X_t)_{t\in\mathbb{R}^+}$ est un processus de Markov si est seulement si, pour tout $(t_1,t_2,...,t_n)\in(\mathbb{R}^+)^n$ avec $(t_1<t_2<t_3\ldots<t_{n-1}<t_n)$ et pour tout $(i_1,i_2,...,i_n)\in E^n$,*

$$\mathbb{P}\left(X_{t_n}=i_n\left|X_{t_{n-1}}=i_{n-1},X_{t_{n-2}}=i_{n-2},...,X_{t_1}=i_1\right.\right)=\mathbb{P}\left(X_{t_n}=i_n\right|X_{t_{n-1}}=i_{n-1}\left.\right)$$

Un processus de Markov en temps continu est homogène si ne dépend pas de t. Dans ce qui suit, nous allons considérer que les processus de Markov homogènes.

Notations : On pose $p_{xy}(t)=\mathbb{P}\left([X_{t+s}=y]\,|\,[X_s=x]\right)=\mathbb{P}\left([X_t=y]\,|\,[X_0=x]\right)$ et $P(t)=\left(p_{xy}(t)\right)_{x,y\in E}$ On pose aussi $\vec{\pi}(t)=P_{X_t}$(vecteur ligne de composantes $\pi_x(t)=\mathbb{P}\left([X_t=x]\right)$).

Propriété 1.2.1. *[12]*

1. *$P(t)$ est une matrice stochastique (c.à.d $p_{xy}(t) \geq 0$ et $\sum_y p_{xy}(t) = 1$ pour tout x)*

2. *Pour tout s et pour tout t, $P(s+t) = P(s)P(t)$.*

3. *Pour tout s et pour tout t, $\vec{\pi}(s+t) = \vec{\pi}(s)P(t)$.*

Démonstration.

1. $p_{xy}(t) \in [0,1]$ car c'est une probabilité. De plus, la ligne x correspond à la loi $\mathbb{P}_{X_t}^{[X_0=x]}$ et on a bien

$$\sum_y p_{xy}(t) = \sum_y \mathbb{P}^{[X_0=x]}\left([X_t = y]\right) = 1$$

2. Pour calculer $p_{xy}(t+s) = \mathbb{P}^{[X_0=x]}\left([X_{t+s} = y]\right)$ on va faire intervenir les différents états pouvant être occupés à l'instant t. Ainsi, on a

$$p_{xy}(t+s) = \sum_{z\in E} p_{x,z}(t)p_{z,y}(s)$$

qui est le coefficient (x,y) de la matrice produit $P(t)P(s)$.

3.

$$\pi_y(t+s) = \mathbb{P}\left([X_{t+s} = y]\right) = \sum_{x\in E} \mathbb{P}\left([X_{t+s} = y] \,/\, [X_t = x]\right)\mathbb{P}\left([X_t = x]\right)$$

$$\pi_y(t+s) = \sum_{x\in E} \pi_x(t)p_{xy}(s)$$

Définition 1.2.2. *[12] Le vecteur (ligne) V_n de X_n des probabilités d'état du système est défini par*

$$V_n = (\mathbb{P}(X_n = 0), \mathbb{P}(X_n = 1), \mathbb{P}(X_n = 2), ..)$$

Définition 1.2.3. *[] Si X_n a le même vecteur de distribution X_k pour tous $(k; n) \in \mathbb{N}^2$, alors la chaîne de Markov est stationnaire.*

Théorème 1.2.1. *[12] Soit $(X_n)_{n\in\mathbb{N}}$ une chaîne de Markov. Nous supposons que X_n est irréductible, apériodique et homogène. La chaîne de Markov peut posséder **une distribution d'équilibre** également appelée **distribution stationnaire**, soit un vecteur de distribution π (avec $\pi = (\pi_0, \pi_1, \cdots)$) satisfaisant*

$$\pi P = \pi, \tag{1.3}$$

et

$$\sum_{k\in E} \pi_k = 1. \tag{1.4}$$

Si nous pouvons trouver un vecteur π équations satisfaisant (1.3) et (1.4), la distribution est unique et

$$\lim_{n\to+\infty} \mathbb{P}(X_n = i | X_0 = j) = \pi_i$$

de sorte que π est la distribution limite de la chaîne de Markov.

1.3 Processus de naissance et de mort

Les processus Markoviens de saut, nous serviront à modéliser l'évolution aléatoire au cours du temps d'un nombre d'individus, dans une population ou un système d'attente. l'ensemble d'arrivé des valeurs de ce processus est l'ensemble des entiers $\mathbb{N}$. Pendant une certaine durée exponentielle, le nombre d'individus reste constant puis il y'aura un saut vers une autre valeur. Nous ne considérerons que des sauts vers les deux valeurs voisines : étant donné que la taille de la population est n, soit elle augmente de 1 (naissance ou arrivée) et donc la taille de la population devient $n+1$ soit elle diminue de 1 (mort ou départ)et elle devient $n-1$.

Définition 1.3.1. (Processus de naissance et de mort)[38]

Soient $(\lambda_n)_{n\in\mathbb{N}}$ et $(\mu_n)_{n\in\mathbb{N}^*}$ deux suites de réels strictement positifs. On appelle processus de naissance et de mort de taux de naissance (λ_n) et taux de mort (μ_n) respèctivement toute chaîne de Markov en temps continu a valeurs dans $\mathbb{N}$.

Dans de tels processus les seules transitions possibles à partir de n sont, soient vers $n+1$ en cas de naissance ou vers $n-1$ en cas de décès.

Remarque 1.3.1. *comme le processus est à valeurs dans $\mathbb{N}$, on a forcément $\mu_0 = 0$.*

Autrement dit, le générateur d'un processus de naissance et mort de taux de naissance $\lambda_n, n \in \mathbb{N}$ et de taux de mort $\mu_n, n \in \mathbb{N}^*$ est donné par une matrice tridiagonale :

$$A = \begin{pmatrix} -\lambda_0 & \lambda_0 & & & (0) \\ \mu_1 & -(\lambda_1+\mu_1) & \lambda_1 & & \\ & \mu_2 & -(\lambda_2+\mu_2) & \lambda_2 & \\ & & \mu_3 & \ddots & \ddots \\ (0) & & \ddots & \ddots & \ddots \end{pmatrix}$$

avec comme graphe :

Ce graphe représente les transitions d'un état à un autre. La transition vers la droite représente une naissance et celle vers la gauche représente une mort.

- Si tous les λ_n sont nuls, on parle de processus de mort.

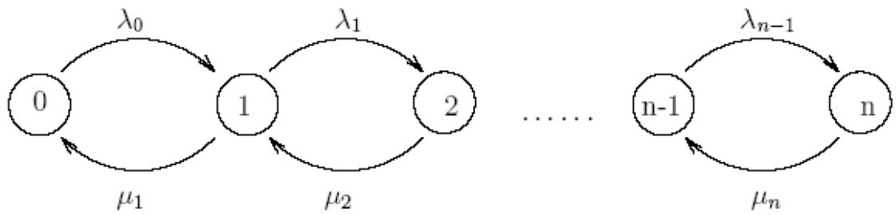


FIGURE 1.1 – Graphe de transition d'un processus de naissance et de mort.

- Si tous les μ_n sont nuls, on parle de processus de naissance.

Comme les λ_n et les μ_n sont tous strictement positifs, tous les états communiquent entre eux et la chaîne est irréductible, d'où le processus admet une distribution stationnaire π. Pour la trouver, on résout le système d'équation $\pi A = 0$ qui donne :

$$\begin{cases} \lambda_0\pi_0 - \mu_1\pi_1 = 0 \\ \lambda_{i-1}\pi_{i-1} - (\lambda_i\mu_i)\pi_i + \mu_{i+1}\pi_{i+1} = 0 \quad \forall i \geq 1. \end{cases}$$

D'où il vient :

$$\pi_n = \frac{\lambda_0\lambda_1\cdots\lambda_{n-1}}{\mu_1\mu_2\cdots\mu_n}\pi_0,$$

or

$$\sum_{i=0}^{+\infty}\pi_i = 1,$$

l'existence d'une loi stationnaire dépend de :

$$\begin{aligned} \pi(0) &= \left[1 + \sum_{i=0}^{+\infty}\prod_{j=1}^{i}\frac{\lambda_{j-1}}{\mu_j}\right]^{-1} \\ &= \left[\sum_{i=0}^{+\infty}\prod_{j=1}^{i}\frac{\lambda_{j-1}}{\mu_j}\right]^{-1}, \end{aligned}$$

avec la convention usuelle qu'un produit vide

$$\prod_{j=1}^{0} = 1$$

1.3.1 Le raisonnement différentiel

notons la probabilité d'avoir n client dant le système à l'instant t $\mathbb{P}(N_t = n)$ par $p_n(t)$ on peut obtenir l'équation différentielle suivante :

$$\frac{dp_n(t)}{dt} = -(\lambda_n + \mu_n)p_n(t) + \lambda_{n-1}p_{n-1}(t) + \mu_{n+1}p_{n+1}(t)$$

Cette équation exprime simplement le fait que pour que la population ait un effectif de n en $t + dt$, il faut que, en t :

- ou bien on ait un effectif de n et ni entrée ni sortie pendant la durée dt ;
- ou bien on ait un effectif de $n - 1$ et une entrée pendant la durée dt ;
- ou bien on ait un effectif de $n + 1$ et une sortie pendant la durée dt.

Cette approche permet d'avoir des informations sur la distribution transitoire deN_t. La distribution stationnaire s'obtient en remarquant que si $p_n(t)$ ne dépend plus de t, alors :

$$0 = -(\lambda_n + \mu_n)\pi_n + \lambda_{n-1}\pi_{n-1} + \mu_{n+1}\pi_{n+1}$$

. Les files d'attente de type Markovien (M/M) sont des cas particuliers très importants de processus de naissance et de mort. Leur étude complète sera effectuée dans le chapitre suivant.

1.4 Chaînes de Markov

Les chaînes de Markov fournissent des moyens très puissants et efficaces pour la description et l'analyse des propriétés des systèmes dynamiques (files d'attente, réseaux informatiques et téléphoniques, physique, biologiques ou économiques *ldots* etc). Leurs utilisations pour la modélisation des phénomènes aléatoires évoluant dans le temps, donnent aux chaînes de Markov une place importante. La structure simple des chaînes de Markov permet de dire beaucoup de choses sur le comportement et l'évolution de ses phénomènes. En effet, l'ensemble des études mathématiques des processus aléatoires peut être considéré comme une généralisation d'une manière ou d'une autre de la théorie des chaînes de Markov.

Le processus de Markov fournit un outil simple de modélisation des systèmes de files d'attente. Dans ce chapitre, nous donnons d'abord une introduction aux concepts de base de la théorie des chaînes de Markov en temps discret et en temps continu. Nous présentons également les propriétés fondamentales des chaînes de Markov. Ensuite nous donnons certaines relations entre les chaînes de Markov des états finis et la théorie des matrices.

1.5 Chaînes de Markov à temps discret

Les Chaînes de Markov à temps discret jouent un rôle central dans la théorie des processus aléatoires. Leur vastes domaines d'applications concernent la modélisation dans les secteurs économiques, le traitement des signaux et des systèmes informatiques, phénomènes météorologiques ou sismiques, $\ldots$. elles sont définies par des familles discrètes de variables ou de vecteurs aléatoires $(X_n)_{n\in T}$, où T est l'ensemble des temps des observations des états du processus.

Pour cette section nous donnons les définitions de base sur les chaînes de Markov homogènes ainsi que certaines propriétés.

Définition 1.5.1. *(***Chaînes de Markov.***)[23] Soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires prenant leurs valeurs dans un espace dénombrable E (dans notre cas, nous prendrons $\mathbb{N}$ ou un sous-ensemble de $\mathbb{N}$). $(X_n)_{n\in\mathbb{N}}$ est une chaîne de Markov si et seulement si*

$$\mathbb{P}\left(X_{n+1}=j|X_0=i_0,\cdots,X_{n-1}=i_{n-1},X_n=i\right)=\mathbb{P}\left(X_{n+1}=j|X_n=i\right).$$

Ainsi, pour une chaîne de Markov, la probabilité d'un comportement futur particulier du processus, lorsque son présent actuel est bien connu, n'est pas modifié par toute connaissance supplémentaire du passé. Le processus est dit sans mémoire ou sans usure.

Soit la notation :

$$p_{ij}=\mathbb{P}\left(X_n=j|X_{n-1}=i\right)$$

p_{ij} est appelé la probabilité de transition de l'état i vers l'état j. Lorsque cette probabilité ne dépend pas de n, nous disons que la chaîne de Markov est homogène.

Remarque 1.5.1. *En général les probabilités de transition sont des fonctions non seulement des états initiaux et finaux, mais aussi du temps de transition. Lorsque les probabilités de transition en une seule étape sont indépendantes de la variable temps n, nous disons que la chaîne de Markov est stationnaire.*

1.5.1 États périodiques

Définition 1.5.1.1. *[38]* **La périodicité.** *La période $d(i)$ d'un état i est définie par :*

$$d(i) = P.G.C.D.\{n \in \mathbb{N}^*; p_{ii}(n) > 0\}$$

(avec $d(i) = 0$ si pour tout $n \in \mathbb{N}^*, p_{ii}(n) = 0$).

Si $d(i) = 1$, i est dit apériodique.

Théorème 1.5.1.1. *[27] La période est une propriété de classe. Si les états i et j communiquent entre eux alors ils ont la même période.*

Démonstration. Comme i et j communiquent, alors il existe des entiers N et M tel que $p_{ij}(M) > 0$ et $p_{ji}(N) > 0$. Pour tout $k > 1$.

$$p_{ii}(M + nk + N) \geq p_{ij}(M)(p_{jj}(k))^n p_{ji}(N)$$

(En effet, le chemin $X_0 = i; X_M = j; X_{M+k} = j; X_{M+nk} = j; X_{M+nk+N} = i$ est une possibilité d'aller de i à i dans $M + nk + N$ étapes).

Par conséquent, pour tout $k > 1$ tel que $p_{jj}(k) > 0$; on a $(M + nk + N) > 0$ pour tous $n > 1$. Et par conséquent, d_i divise $M + nk + N$ pour tout $n > 1$, et en particulier, d_i divise k. Nous avons donc démontré que d_i divise tous les k de telle sorte que $p_{jj}(k) > 0$, et en particulier, d_i divise d_j. Par symétrie, d_j divise d_i, et donc, finalement, $d_i = d_j$

Théorème 1.5.1.2. *[33] Soit P une matrice stochastique irréductible avec la période d, alors pour tous les états i, j il existent $m > 0$ et $n_0 > 0$ (m et n peuvent être dépendant de i, j) de* *telle sorte que*

$$p_{ij}(m+nd) > 0 \quad \forall n \geq 0 \tag{1.5}$$

Démonstration. Il suffit de démontrer pour $i = j$. En effet, il existe m tel que $p_{ij}(m) > 0$, parce que j est accessible à partir de i, la chaîne étant irréductible, et donc, si pour une $n_0 > 0$ nous avons $p_{jj}(nd) > 0$ pour tout $n > n_o$, alors $p_{ij}(m+nd) > p_{ij}(m)p_{jj}(nd) > 0$ pour tout $n > n_0$. Le PGCD de l'ensemble $A = k \geq 1; p_{jj}(k) > 0$ est d et A est fermé pour addition. L'ensemble A contient donc tout sauf un nombre fini des multiples positifs de d. En d'autres termes, il existe n_0 tel que $n > n_0$ implique $p_{jj}(nd) > 0$

Propriété 1.5.1.1. *[12] Tous les états d'une même classe de communication ont la même période.*

Démonstration. Si x et $x^{'}$ communiquent, il existe k et l tels que $p_{xx^{'}}(k) > 0$ et $p_{x^{'}x}(l) > 0$. Alors $p_{xx}(k+l) > 0$. Soit m tel que $p_{x^{'}x^{'}}(m) > 0$. Alors $p_{xx}(k+m+l) > 0$. Donc $d(x)$ divise $k+l$ et $d(x)$ divise $k+m+l$; donc $d(x)$ divise m et ceci, pour tout m tel que $p_{x^{'}x^{'}}(m) > 0$, donc $d(x)$ divise $d(x)$. Comme x et $x^{'}$ jouent le même rôle, alors $d(x) = d(x^{'})$

Remarque 1.5.1.1. *Une classe dont un élément admet une boucle, (c'est-à-dire $p_{xx} > 0$), est obligatoirement apériodique (c'est-à-dire de période 1), mais ce n'est pas une condition nécessaire.*

Théorème 1.5.1.3. *[12] Si x et y sont dans une même classe de communication de période d, si $p_{xy}(n) > 0$ et si $p_{xy}(m) > 0$, alors d divise $m - n$.*

Démonstration. Comme y mène à x, il existe k tel que $p_{yx}(k) > 0$. On a donc $p_{xx}(m+k) > 0$ et $p_{xx}(n+k) > 0$. Donc d divise $m+k$ et $n+k$, il divise donc la différence.

Ainsi, on peut partitionner les classes périodiques de la façon suivante.

Définition 1.5.1.2. *[38] Soit C une classe de communication de période d, et soit x_0 fixé dans C. On définit, pour $k \in 0, 1, \cdots, d-1$,*

$$C_k^{'} = y \in E; \quad si \quad p_{x_0,y}(n) > 0 \quad alors \quad n \equiv k[d].$$

Les $C_k^{'}$ sont appelées sous-classes cycliques de C.

Théorème 1.5.1.4. *[27] Soit X une chaîne de Markov homogène dans un espace d'états E, et soit $E^{'} \subseteq E$ une classe irréductible non vide. Alors pour tout $i, j \in E^{'}$, les périodes de i et j sont les même, i.e., $d(i) = d(j)$.*

Démonstration. Soient $i, j \in E^{'}$, $i \neq j$ deux états. Comme $E^{'}$ est une classe irréductible, alors ils existent deux entiers $t, s \geq 1$ tel que $p_{ij}(t) > 0$ et $p_{ij}(s) > 0$. L'équation de Chapman Kolmogorov donnent :

$$p_{ii}(t+s) \geq p_{ij}(t)p_{ji}(s) > 0 \quad \text{et} \quad p_{jj}(t+s) \geq p_{ji}(s)p_{ij}(t) > 0;$$

Par conséquent, les nombres $d(i)$ et $d(j)$ sont différents de 0. En choisissant arbitrairement un entier $m \geq 1$,

tel que $p_{ii}(m) > 0$, appliquons plusieurs fois l'équation de Chapman Kolmogorov, nous aurons pour tout $k \geq 1$

$$p_{jj}(t+s+km) \geq p_{ji}(s)p_{ii}(km)p_{ij}(t) \geq p_{ji}(t)\left(p_{ii}(m)\right)^k p_{ij}(s) > 0.$$

Ainsi, par la définition de la période de l'état j, $d(j)$ est un diviseur à la fois de $(t+s+m)$ et de $(t+s+2m)$, et donc c'est aussi un diviseur de leur différence $(t+s+2m)-(t+s+m)=m$. Il en découle immédiatement que $d(j)$ est un diviseur de tout m pour lequel $p_{ii}(t+s)>0$, et donc un diviseur de $d(i)$; Donc, $d(j)\leq d(i)$. En changeant le rôle de i et j on obtient l'inégalité inverse $d(j)\geq d(i)$, et par conséquent $d(j)=d(i)$.

1.5.2 Etats récurrents et états transients

Définition 1.5.2.1. Les probabilités de transition.*[38] Pour une chaîne de Markov homogène,*

$$p_{ji}=\mathbb{P}\left(X_n=i|X_{n-1}=j\right)$$

est appelé la probabilité de transition *de l'état j à l'état i.*

Définition 1.5.2.2. *[38] Si au temps 0, la chaîne est en l'état e_i, la variable aléatoire $T_i = Min\{n\geq 1; X_n=e_i\}$ définit le temps de premier retour en e_i.*

Définition 1.5.2.3. *[38] Un état e_i est récurrent si partant de e_i on y revient sûrement, ce qui s'exprime par :*

$$\mathbb{P}((\quad \textit{il existe} \quad n\geq 1 \quad \textit{tel que} \quad X_n=e_i)|X_0=e_i)=1,$$

soit encore :

$$\mathbb{P}(T_i<+\infty|X_0=e_i)=1.$$

Un état récurrent est donc visité un nombre infini de fois.

Définition 1.5.2.4. *[38] Un état e_i est dit transient s'il n'est pas récurrent :*

$$\mathbb{P}(T_i < +\infty | X_0 = e_i) < 1.$$

Un état est transient s'il n'est visité qu'un nombre fini de fois.

Par définition, une chaîne irréductible ne contient aucun état transient ou absorbant.

Définition 1.5.2.5. *[38] Une chaîne dont tous les états sont récurrents est dite récurrente, une chaîne dont tous les états sont transients est dite transiente.*

Définition 1.5.2.6. *[38] La probabilité de première transition en n unités de temps, de l'état e_i à l'état e_j, notée $f_{i,j}^{(n)}$ est définie par :*

$$f_{i,j}^{(n)} = \mathbb{P}\left(T_j = n | X_0 = e_i\right) = \mathbb{P}\left(X_n = e_j, X_{n-1} \neq e_j, \cdots, X_1 \neq e_j | X_0 = e_i\right),$$

$f_{i,j}^{(n)}$ est la probabilité de premier retour à l'état e_i en n unités de temps.

Théorème 1.5.2.1. *[38]*

$$e_i \text{ est récurrent} \Leftrightarrow \sum_{n=1}^{+\infty} f_{i,i}^{(n)} = 1,$$

$$e_i \text{ est transitoire} \Leftrightarrow \sum_{n=1}^{+\infty} f_{i,i}^{(n)} < 1.$$

Définition 1.5.2.7. *[38] On appelle temps moyen de retour en e_j :*

$$\mu_j = \mathbb{E}\left[T_j = n | X_0 = e_i\right] = \sum_n n P_{j,j}^{(n)}.$$

Il existe deux sortes d'états récurrents

- *Les états récurrents positifs*
- *Les états récurrents nuls.*

Définition 1.5.2.8. *[38] Un état e_j est récurrent positif (ou non nul) s'il est récurrent et si le temps moyen μ_j de retour est fini. Dans le cas où le temps moyen de retour est infini, l'état est dit récurrent nul.*

Théorème 1.5.2.2. *[38] (Caractérisation des états)*

$$(e_i \quad \textit{transient} \quad) \Leftrightarrow \sum_{n\geq 1} P_{i,i}^{(n)} < +\infty,$$

$$(e_i \quad \textit{récurrent nul} \quad) \Leftrightarrow \sum_{n\geq 1} P_{i,i}^{(n)} = +\infty \quad \textit{et} \quad \lim_{n\to+\infty} P_{i,i}^{(n)} = 0,$$

$$(e_i \quad \textit{récurrent positif} \quad) \Leftrightarrow \sum_{n\geq 1} P_{i,i}^{(n)} = +\infty \quad \textit{et} \quad \lim_{n\to+\infty} P_{i,i}^{(n)} > 0.$$

Théorème 1.5.2.3. *[38] (Propriétés des chaînes finies)*

1. *Aucun état n'est récurrent nul.*
2. *Aucune chaîne finie n'est transiente ; en revanche, certains de leurs états peuvent être transients.*
3. *Il existe au moins un état récurrent positif.*
4. *Si en outre la chaîne est irréductible, elle est récurrente positive.*

Démonstration. (1) Supposons qu'il existe un état récurrent nul e_j, et C_j sa classe de communication, alors

$$\forall n \in \sum_{k\in C_j} P_{i,i}^{(k)} = 1 \tag{1.6}$$

Mais $\lim\limits_{n\to+\infty} p_{j,k}^{(n)} = 0$, qui est en contradiction avec (1.6).

(2) Supposons tous les états transients, on montre alors que $\forall e_i, e_j$ transients, la série $\sum p_{j,k}^{(n)}$ converge, donc

$$\lim_{n\to+\infty} p_{j,k}^{(n)} = 0. \tag{1.7}$$

Or, $\sum\limits_{j=1}^{card(E)} p_{ij}^{(n)} = 1 (\forall n \in \mathbb{N})$, donc, par passage à la limite quand $n \to +\infty$, on arrive à une contradiction de (1.7).

(3)Est conséquence de (1) et (2).

(4) La chaîne ayant au moins un état récurrent d'après (c) et une seule classe est récurrente positive.

Définition 1.5.2.9. *[38] Un état récurrent positif et apériodique est dit ergodique. Une chaîne irréductible, apériodique et récurrente positive est dite ergodique.*

Théorème 1.5.2.4. *[38] Les états d'une classe de communication qui contient au moins un état ergodique sont ergodiques.*

Preuve On sait que les propriétés de récurrence et d'apériodicité sont des propriétés de classe : l'ergodicité l'est donc aussi. Il suffit donc qu'un état de la classe soit ergodique pour que la classe qui le contient le soit.

Théorème 1.5.2.5. *[34] (**théorème ergodique**) Soit $(X_n)_{n\in\mathbb{N}}$ une chaîne de Markov irréductible récurrente positive, de probabilité invariante π Soit $f : E \to \mathbb{R}$ une fonction bornée. Alors*

$$\frac{1}{n}\sum_{k=1}^{n} f(X_k) \xrightarrow[n\to+\infty]{p.s.} \sum_{k=1}^{n} \pi_k f(x). \tag{1.8}$$

Preuve Soit $x \in E$. On note

$$N_x(n) = \sum_{1 \leq k \leq n} \mathbb{1}_{X_k = x},$$

le nombre de retour à x jusqu'à l'instant n. Si on note $S_x^0, S_x^1, \ldots$ les longueurs des excursions $\epsilon_0, \epsilon_1, \ldots$, on obtient :

$$S_x^0 + S_x^1 + \ldots + S_{N_x(n)-1}^0 \leq n < S_x^0 + S_x^1 + \ldots + S_{N_x(n)}^0.$$

Donc

$$\frac{S_x^0 + S_x^1 + \ldots + S_{N_x(n)-1}^0}{N_x(n)} \leq \frac{n}{N_x(n)} < \frac{S_x^0 + S_x^1 + \ldots + S_{N_x(n)}^0}{N_x(n)}.$$

On sait que sous $\mathbb{P}_x, \epsilon_0, \epsilon_1, \ldots$ est une suite i.i.d., il en est donc de même pour la suite $S_x^0, S_x^1, \ldots$ En appliquant la loi des grands nombres :

$$\frac{S_x^0 + S_x^1 + \ldots + S_{N_x(n)}^0}{N_x(n)} \xrightarrow[n \to +\infty]{\mathbb{P}_x - p.s.} \mathbb{E}_x T_x = m_x.$$

Donc par l'encadrement précédent, on obtient

$$\frac{n}{N_x(n)} \xrightarrow[n \to +\infty]{\mathbb{P}_x - p.s.} m_x,$$

ou bien

$$\frac{N_x(n)}{n} \xrightarrow[n \to +\infty]{\mathbb{P}_x - p.s.} \frac{1}{m_x}.$$

D'après la proprieté de Markov forte, la limite de $(X_n)_{n \in \mathbb{N}}$ est la même que $(X_{T_x+n})_{n \in \mathbb{N}}$. Donc cette limite a lieu $\mathbb{P}_\mu - p.s.$ pour toute loi initiale μ. Soit $F \subset E$. On note $\overline{f} = \sum_{x \in E} \pi_x f(x)$, et

$c = \sup |f(x)|$. On a :

$$\begin{aligned}
\left|\frac{1}{n}\sum_{k=1}^{n} f(X_k) - \overline{f}\right| &= \left|\sum_{x\in E}^{n}\left(\frac{N_x(n)}{n} - \pi_x\right) f(x)\right| \\
&\leq c\sum_{x\in F}^{n}\left|\frac{N_x(n)}{n} - \pi_x\right| + c\sum_{x\notin E}^{n}\left(\frac{N_x(n)}{n} + \pi_x\right) \\
&= c\sum_{x\in E}^{n}\left|\frac{N_x(n)}{n} - \pi_x\right| + c\sum_{x\in E}^{n}\left(\pi_x - \frac{N_x(n)}{n}\right) + 2c\sum_{x\notin E}^{n}\pi_x \\
&\leq 2c\sum_{x\in E}^{n}\left|\frac{N_x(n)}{n} - \pi_x\right| + 2c\sum_{x\notin E}^{n}\pi_x.
\end{aligned}$$

Soit $\epsilon > 0$. On choisit F fini tel que $\sum_{x\notin E}^{n} \pi_x \leq \frac{\epsilon}{4c}$ et $N(\omega)$ tel que $\forall n \geq N(\omega)$,

$$\sum_{x\in E}^{n}\left|\frac{N_x(n)}{n} - \pi_x\right| \leq \frac{\epsilon}{4c}$$

On obtient le résultat. ■

1.6 Matrices des probabilités de transition d'une chaîne de Markov

L'analyse d'une chaîne de Markov concerne principalement le calcul des probabilités des réalisations possibles du processus. Ces calculs concernent les matrices des n-étapes de transitions $P^n = P_{ij}^n$, ici P_{ij}^n désigne la probabilité que le processus passe de l'état i à l'état j en n transitions. Nous disons qu'une matrice $P = (p_{ij} : i, j \in T)$ est stochastique si chaque ligne $(p_{ij} : j \in T)$ est une distribution. Si l'espace d'états E est fini la matrice (1.9) s'écrit :

Définition 1.6.1. *[20] La matrice*

$$P = \begin{pmatrix} P_{00} & P_{01} & P_{02} & P_{03} & \cdots \\ P_{10} & P_{11} & P_{12} & P_{13} & \cdots \\ P_{20} & P_{21} & P_{22} & P_{23} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \\ P_{i0} & P_{i1} & P_{i2} & P_{i3} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ldots \end{pmatrix} \tag{1.9}$$

dont les coefficients sont les probabilités de transition P_{ij} est appelée à la fois matrice de Markov et matrice des probabilités de transitions du processus.

La $i$ème ligne de P, pour $i = 0, 1, \cdots$ est la distribution de probabilité des valeurs X_{n+1} avec la condition que $X_n = i$. Si le nombre d'états est fini, alors P est une matrice carrée finie dont l'ordre (le nombre de lignes) est égal au nombre d'états. Les quantités P_{ij} satisfont la conditions :

$$P_{ij} \geq 0 \quad \text{pour} \quad i, j = 0, 1, 2, \cdots$$

$$\sum_{j=0}^{\infty} P_{ij} = 1 \quad \text{pour tout} \quad i = 0, 1, 2, \cdots$$

les nombres P_{ij} sont des probabilités donc des nombres positifs.

Un processus de Markov est complètement défini une fois que sa probabilité de transition et X_0 son état initial sont donnés.

$$A = \begin{pmatrix} P_{00} & P_{01} & \cdots & P_{0r} \\ P_{10} & P_{11} & \cdots & P_{1r} \\ \vdots & \vdots & \ddots & \vdots \\ P_{r0} & P_{r1} & \cdots & P_{rr} \end{pmatrix}$$

Définition 1.6.2. *[20] Une chaîne de Markov est* **irréductible** *si et seulement si chaque état peut être atteint à partir de tout autre état.*

Définition 1.6.3. *[20] Pour une chaîne de Markov homogène,*

$$p_{ji} = \mathbb{P}\left(X_n = i | X_{n-1} = j\right)$$

est appelé la **probabilité de transition** *de l'état j à l'état i.*

Définition 1.6.4. *[20] Un état i est* **périodique** *si il existe un entier $\delta > 1$ de telle sorte que $P\left(X_{n+k} = i | X_n = i\right) = 0$ moins k est divisible par δ ; autrement l'état est* **apériodique.**. *Si tous les états d'une chaîne de Markov sont périodiques (respectivement apériodique), on dit que la chaîne est périodique (respectivement apériodique).*

1.7 Exercices

Exercice 01 :[19]

Soient $X_1, \ldots, X_n$ des variables aléatoires indépendantes est de distribution exponentielles de moyenne $\mathbb{E}(X_i) = \frac{1}{\mu_i}$, $\quad i = 1, \ldots, n$. On considère les deux variables aléatoires

$$Y_n = \min(X_1, \ldots, X_n) \quad \text{et} \quad Z_n = \max(X_1, \ldots, X_n)$$

1. Déterminer les distributions de Y_n et Z_n.
2. Montrer que la probabilité que X_i soit le plus petit parmi $X_1, \ldots, X_n$ est égale à

$$\frac{\mu_i}{\mu_1 + \ldots + \mu_n} \quad i = 1, \ldots, n$$

<u>**corrigé :**</u>

1. En utilusant

$$\mathbb{P}(Y_n > x) = \prod_{i=1}^{n} \mathbb{P}(X_i > x)$$

et

$$\mathbb{P}(Z_n \leq x) = \prod_{i=1}^{n} \mathbb{P}(X_i \leq x).$$

2. Il s'ensuit que

$$\begin{aligned}
\mathbb{P}(X_i \min = (X_1, \ldots, X_n)) &= \int_{i=0}^{+\infty} \prod_{j \neq i} \mathbb{P}(X_j > x) f_{X_i}(x) dx \\
&= \int_{i=0}^{+\infty} \mu_i e^{-(\mu_1 + \ldots + \mu_n)x} dx \\
&= \frac{\mu_i}{\mu_1 + \ldots + \mu_n} \quad i = 1, \ldots, n
\end{aligned}$$

<u>**Exercice 02 :**</u>

On considère une chaîne de Markov modélisant K machines indépendantes. Chacune d'elles peut tomber en panne avec un taux λ. Un réparateur répare une seule machine à la fois avec le taux μ.

1. Exprimer le taux de passage de l'état (n machines en panne) à l'état ($n+1$ machines en panne).

   On fera une démonstration rigoureuse. Il sera peut-être nécessaire de démontrer au préalable un lemme sur la loi de l'inf d'un certain nombre de variables de loi exponentielle.

2. Calculer la probabilité d'avoir K machines en panne.

3. Application numérique : $K = 7$, $\lambda = 0.001$, $\mu = 1$.

Corrigé

Nous considérons une chaîne de Markov à K états. L'état i est atteint lorsque i machines parmi les K sont en pannes.

1. Lemme sur la loi de l'inf de n variables indépendantes de loi exponentielle Soit n variables aléatoires indépendantes X_i de loi exponentielle de paramètre λ_i.

$$\begin{aligned}
P[\inf(X_i) \leq x] &= 1 - P[Inf(Xi) > x] \quad \text{(Passage au complémentaire)} \\
&= 1 - \prod_{i=1}^{n} P[X_i > x] \\
&= 1 - \prod_{i=1}^{n} e^{-\lambda_i x} \\
&= 1 - e^{\sum_{i=1}^{n} \lambda_i x}
\end{aligned}$$

Lorsque x tend vers 0, nous avons donc :

$$P[\inf(X_i) \leq x] = \sum_{i=1}^{n} \lambda_i x + o(x)$$

Expression du taux de passage de l'état n à l'état $n+1$

Nous supposons donc qu'il y a $K-n$ machines en état de marche. Soit X_i le temps qui sépare l'instant courant de la panne de la machine i. La première machine tombera en panne au bout d'un temps $\inf(X_i)$.

La probabilité de passer de l'état n à l'état $n+1$, c'est à dire $P[X(t+dt) = n+1/X(t+dt) = n]$ est

$$P[X(t+dt) = n+1/X(t) = n] = (K-n)\lambda dt + o(dt)$$

Expression du taux de passage de l'état $n+1$ à l'état n

Le réparateur est exponentiel et l'on passe donc de l'état $n+1$ à l'état n avec le taux μ.

2. On a bien un procesus sans mémoire, qui de plus, évolue uniquement par saut de 1 ou -1. C'est donc un processus de naissance et de mort.

   On a donc :

$$P(n) = \frac{\lambda(0)\lambda(1)\cdots\lambda(n-1)}{\mu(1)\mu(2)\cdots\mu(n)} P(0) = \frac{K!}{(K-n)!}\left(\frac{\lambda}{\mu}\right)^n P(0)$$

   La probabilité d'avoir K machines en pannes est donc :

$$P(K) = \frac{K!\left(\frac{\lambda}{\mu}\right)^K}{\sum_{n=0}^{K} \frac{K!}{(K-n)!}\left(\frac{\lambda}{\mu}\right)^n}$$

   soit :

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

d'où en réindexant $(n = K - n)$:

$$P(K) = \frac{1}{\displaystyle\sum_{n=0}^{K} \frac{1}{(n)!}\left(\frac{\lambda}{\mu}\right)^{-n}}$$

3. On trouve : $P(7) = 4.7\ 10^{-11}$.

Exercice 03 :

Le nombre de clients par heure dans un magasin suit un processus de Poisson de taux 4.

1. Quelle est la probabilité qu'il y ait strictement moins de deux clients la première heure ?
2. Si dans la première heure 6 clients sont arrivés , calculer la probabilité d'avoir pendant la deuxième heure, moins de deux clients.
3. Le résponsable du magasin donne une pause aux serveurs après avoir servi dix clients. Quelle est la durée moyenne séparant deux pauses ?
4. Trois quarts des clients sont des hommes, un quart des femmes, ceci indépendamment de l'heure d'arrivé. Calculer la probabilité qu'en une heure il y ait exactement 2 hommes et 3 femmes à visiter le magasin.

Corrigé

Le nombre de clients par heure suit un processus de Poisson de taux 4.

1. Le nombre de clients au bout d'une heure suit une loi exponentielle de paramètre 4. La probabilité qu'il y ait strictement moins de 2 clients au bout d'une heure est celle qu'il y en ait 0 ou 1 :

$$p = e^{-4}\frac{4^0}{0!} + e^{-4}\frac{4^1}{1!} = 5e^{-4}$$

2. Le fait qu'il y ait eu 6 clients la première heure ne change rien à la probabilité qu'il y en ait strictement moins de deux la deuxième heure : c'est toujours $p = 5e^{-4}$.

3. Le temps nécessaire (en heures) pour avoir 10 clients est T_{10}, somme de 10 exponentielles indépendantes de paramètre 4 : c'est donc une loi $\Gamma(10, 4)$. Sa moyenne est tout simplement :

$$d = \mathbb{E}[T_{10}] = \frac{10}{4} = 2\text{h}30.$$

4. Soit p la probabilité cherchée. Avec des notations évidentes, on peut décomposer p comme suit :

$$p = \mathbb{P}(2H, 3F|N_1 = 5)P(N_1 = 5).$$

Le second terme est la probabilité qu'il y ait exactement 5 clients en 1 heure, c'est donc :

$$P(N_1 = 5) = e^{-4}\frac{4^5}{5!}.$$

Le premier terme est la probabilité que sur 5 clients, 2 soient des hommes et 3 femmes. Or, sur 5 clients, le nombre de clients hommes suit une loi binômiale $B(5, 3/4)$. Ceci donne :

$$P(2H, 3F|N_1 = 5) = C_5^2 \left(\frac{3}{4}\right)^2 \left(\frac{1}{4}\right)^3$$

Au total, la probabilité cherchée est donc :

$$p = \left(\frac{3}{4}\right) e^{-4}$$

Exercice 04 :[5]

Entrées dans un magasin On suppose que les arrivées dans un magasin se font selon un processus de Poisson de taux 40 clients par heure. Ce magasin ouvre à 9h et ferme à 19h.

1. Quelle est la loi du nombre de clients sur la journée ?
2. Quelle est la probabilité qu'aucun client n'arrive entre 9h et 9h15.
3. Calculer la probabilité qu'aucun client n'arrive entre 9h et 9h15 sachant qu'il en est arrivé 3 entre 9h et 9h30.
4. Calculer le nombre moyen de clients qui sont entrés dans le magasin jusqu'à 11h sachant qu'il en est arrivé 3 jusqu'à 9h30.
5. Un cadeau est offert par le magasin tous les 13 clients (c'est-à-dire au 13e client, au 26e, etc.). Donner la loi de la durée espaçant deux cadeaux.

Corrigé

On suppose que les arrivées dans un magasin se font selon un processus de Poisson de taux 40 clients par heure. Ce magasin ouvre à 9h et ferme à 19h.

1. Le nombre N_j de clients sur la journée (de durée 10 heures) est une loi de Poisson de paramètre 400, c'est-à-dire à valeurs dans $\mathbb{N}$, avec :

$$\forall n \in \mathbb{N} \mathbb{P}(N_j = n) = e^{-400} \frac{400^n}{n!}.$$

2. Prenons la minute comme unité de temps, on a donc un processus d'intensité $2/3$. Le nombre de clients arrivés jusqu'à 9h15 suit donc une loi de Poisson de paramètre 10. La probabilité qu'aucun client ne soit arrivé à 9h15 est donc :

$$P(N_{15} = 0) = e^{-10}.$$

3. On cherche cette fois : $P(N_{15} = 0 | N_{30} = 3)$.

Première méthode : on utilise la formule de Bayes, ce qui donne :

$$P(N_{15} = 0|N_{30} = 3) = \frac{P(N_{30} = 3|N_{15} = 0)P(N_{15} = 0)}{P(N_{30} = 3)}.$$

or par les propriétés d'absence de mémoire et d'homogénéité du processus de Poisson :

$$P(N_{30} = 3|N_{15} = 0) = e^{-10}\frac{10^3}{3!}.$$

Ainsi on a :

$$P(N_{15} = 0|N_{30} = 3) = \frac{e^{-10}\frac{10^3}{3!}e^{-10}}{e^{-20}\frac{20^3}{3!}}.$$

ce qui donne finalement :

$$P(N_{15} = 0|N_{30} = 3) = \frac{1}{8}$$

Seconde méthode : sachant que 3 personnes sont arrivées en une demi-heure, les dates d'arrivées sont i.i.d. uniformes sur l'intervalle [0, 30]. La probabilité que les 3 soient arrivées dans le second quart d'heure est donc :

$$P(N_{15} = 0|N_{30} = 3) = \frac{1}{2}\frac{1}{2}\frac{1}{2} = \frac{1}{8}$$

4. On veut maintenant calculer $E[N_{120}|N_{30} = 3]$. Or, entre 9h30 et 11h, on a un processus de Poisson de taux 2/3, donc :

$$E[N_{120}|N_{30} = 3] = 3 + E[N_{90}] = 63$$

5. La loi de la durée T espaçant deux cadeaux est la somme de 13 variables aléatoires i.i.d.

exponentielles de paramètre $2/3$, c'est-à-dire une loi $\Gamma(13, 2/3)$. Sa densité est :

$$f(t) = \frac{\left(\frac{2}{3}t\right)^{12}}{12!} = \frac{2}{3}e^{-\frac{2}{3}t}\mathbb{1}_{[0,+\infty[}(t)$$

avec t exprimé en minute

Chapitre 2

Files d'attente

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. 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'autres a permit 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,$\cdots$) 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).

Depuis les travaux d'Erlang [33] Un grand nombre d'applications dans tous les domaines ont été mis en oeuvre et publiées. En 1953, Kendall [25] 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, Saaty [32], auteur de l'un des premiers livres complets sur la théorie des files d'attente. Ensuite c'est les contributions des mathématiciens Khintchine [26], Palm, Pollaczek et Kolmogorov qui ont vraiment poussés la théorie des files d'attente.

Dans ce chapitre, nous introduisons la terminologie de la théorie des files d'attente et de certaines définitions et notations qui sont nécessaires dans l'étude des systèmes de files d'attente (la notation de KANDELL, la formule de LITTLE...). En suite nous étudions quelque modèles de files d'attente markoviennes ($M/M/1$, $M/M/1/K$, $M/M/c$, $M/M/\infty$,) et non markovienne ($M/G/1$) et l'évaluation des paramètres de performance.

Une file d'attente est associé à l'attente, qui est une partie inévitable de la vie moderne. Ces actions vont de l'attente d'être servis dans un guichet de poste ou banque, dans une station d'essence, un bus, dans une agence, jusqu'à l'attente d'un opérateur sur un appel téléphonique, les feux de circulation dans le trafic routier.... Même les situations de stocks et ateliers de réparation peuvent également être considérées comme des files d'attente.

La caractéristique dominante dans tous ces exemples précédents est que certaines unités appelées clients cherchent à accéder à un système, afin de recevoir un service d'une manière aléatoire dont les durées de service sont encore aléatoires. Mais parfois la demande de ce service par plusieurs clients engendre une file d'attente et les clients se dirigent vers l'espace d'attente du système.

2.1 Description d'une file simple

Une file d'attente est un système caractérisé par un espace d'attente qui contient une ou plusieurs places, et un espace de service composé d'un ou plusieurs serveurs. Les clients arrivent de l'extérieur à des instants aléatoires, ils attendent que l'un des serveurs soit libre pour pouvoir être servi puis quittent le système. Les clients peuvent être des individus, des appels téléphoniques, des signaux électriques, des véhicules, des accidents, . . . etc

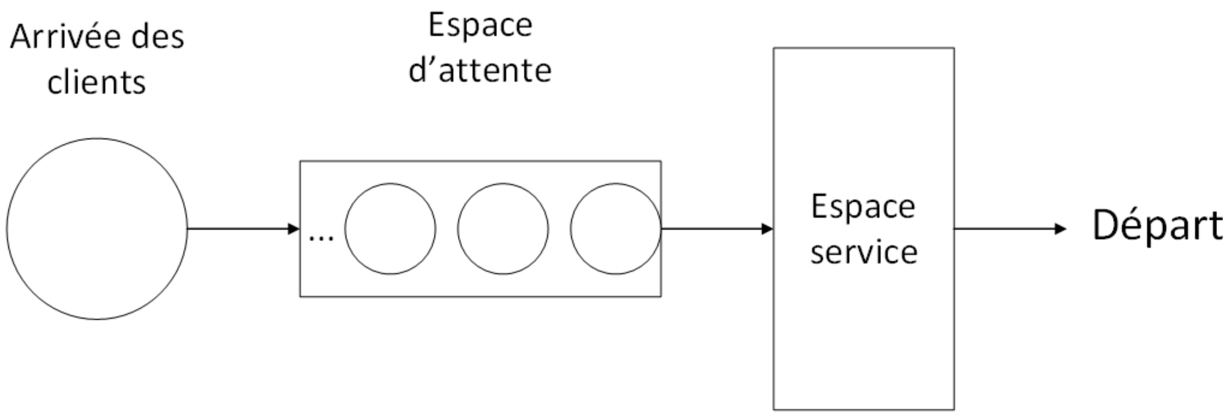


FIGURE 2.1 – Système de file d'attente simple

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

- Le processus d'arrivée : On s'intéresse aux instants d'arrivés des clients dans la file. Ils sont en général aléatoires. Certaines hypothèses sont faites en se basant sur leurs lois. Tout d'abord, il n'arrive qu'un client à la fois. La deuxième hypothèse est l'homogénéité dans le temps. Cela se traduit par le fait que les temps d'interarrivée des clients sont des v.a. de même loi. Ils sont également supposés indépendants. Enfin, la loi des temps d'interarrivée est supposée connue. Le cas le plus courant est celui où cette loi est exponentielle. Dans ce cas, le modèle des temps d'arrivée des clients est un processus de Poisson. Evidemment d'autres cas peuvent se présenter : temps d'interarrivés constants, de loi uniforme, loi normale ou encore gamma.

- Le processus de service : Il comprit le nombre de serveurs et la loi probabiliste décrivant la durée des services.
- Structure et discipline de la file : La discipline de service détermine l'ordre dans lequel les clients sont rangés dans la file et y sont retirés pour recevoir un service.

2.1.0.1 Notation de Kendall

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

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

avec

1. T : indique le processus d'arrivée des clients. Les codes utilisés sont :
   - M (Markov) : Interarrivées des clients sont indépendamment, identiquement distribuées selon une loi exponentielle. Il correspond à un processus de Poisson ponctuel (propriété sans mémoire).
   - D (Répartition déterministe) : les temps interarrivées des clients ou les temps de service sont constants et toujours les mêmes.
   - GI (général indépendant) : Interarrivées des clients ont une distribution générale (il n'y a aucune hypothèse sur la distribution mais les interarrivées sont indépendantes et identiquement distribuées).
   - G (général) : Interarrivées de clients ont une distribution générale et peuvent ï¿½tre dépendantes .

- E_k : Ce symbole désigne un processus oï¿½ 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 ï¿½tre servi. Les codes utilisés sont les suivants :
   - FIFO (first in, first out) ou FCFS (first come first served) : c'est la file standard dans laquelle les clients sont servis dans leur ordre d'arrivée. Notons que les disciplines FIFO et FCFS ne sont pas équivalentes lorsque la file contient plusieurs serveurs. Dans la première, le premier client arrivé sera le premier à quitter la file alors que dans la deuxième, il sera le premier à commencer son service. Rien n'empï¿½che alors qu'un client qui commence son service après lui, dans un autre serveur, termine avant lui.
   - LIFO (last in, first out) ou LCFS (last come, first served). Cela correspond à une pile, dans laquelle le dernier client arrivé (donc posé sur la pile) sera le premier traité (retiré de la pile). A nouveau, les disciplines LIFO et LCFS ne sont équivalentes que pour une file monoserveur.
   - SIRO (Served In Random Order), les clients sont servis aléatoirement.
   - PNPN (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 : 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.

2.2 Modélisation d'une file d'attente

La modélisation mathématique d'un système de files d'attente se fait le plus souvent pour déterminer les paramètres de performance à partir de la description du système. Le paramètre de performance le plus important dans une files d'attente est : le processus aléatoire $(X_t)_{t\geq o}$ enregistrant le nombre de clients dans la file d'attente à l'instant t. Ceci est toujours prise pour inclure à la fois ceux qui sont servis et ceux qui attendent d'être servis. Et à partir de $(X_t)_{t\geq o}$ toutes les valeurs moyennes de la plupart des autres paramètres peuvent être déduites comme suit :

1. Les probabilités d'état $\pi_n(t) = P([X_t = n])$ qui définissent le régime transitoire du processus $(X_t)_{t\geq o}$;
2. Le régime stationnaire du processus, défini par

$$\pi_n = \lim_{t\to\infty} \pi_n(t) = \lim_{t\to\infty} P([X_t = n])$$

A son tour la distribution stationnaire du processus $(X_t)_{t\geq o}$, permet de calculer d'autres paramètres de performances du système telles que :

- Le nombre moyen L de clients dans le système ;

- Le nombre moyen L_q de clients dans la file d'attente ;
- La durée d'attente moyenne W_q d'un client ;
- La durée de séjour moyenne W dans le système (attente + service) ;
- Le taux d'occupation des postes de service ;
- Le pourcentage de clients n'ayant pu être servis ;
- La durée d'une période d'activité, c'est-à-dire de l'intervalle de temps pendant lequel il y a toujours au moins un client dans le système ...

2.2.1 Files d'attente non markoviennes

Le processus est non markovien si :

- L'une des deux quantités stochastiques : le temps des inter-arrivées ou la durée de service n'est pas exponentielle,
- Certains paramètres supplémentaires sur les problèmes de files d'attente, rendent le modèle non markovien.

Ces facteurs rendent l'étude mathématique du modèle très délicate, voire impossible, pour cela on essaye de transformer ce modèle à un processus de Markov à l'aide de l'une des méthodes d'analyse suivantes :

1. **Méthode des étapes d'Erlang :** Elle consiste à approximer la loi de probabilité qui possède une transformation de Laplace rationnelle par une loi de Cox (mélange de lois exponentielles), cette dernière possède la propriété d'absence de mémoire par étape.
2. **Méthode de la chaîne de Markov induite :** Son principe est de choisir une suite d'instants $1, 2, 3, \cdots, n$ tels que la chaîne induite $N_n, n \geq 0$, où $N_n = N(n)$, soit markovienne et homogène.

î

2.2.2 Files d'attente markoviennes

Ce modèle est caractérisé par des interarrivées exponentielles ainsi que les durées de service. Leur notation de Kendall est de la forme $M/M/\cdots$

2.2.3 La propriété PASTA

Avant de spécifier davantage le type de files sur lequel nous allons travailler, nous commençons par un résultat assez général.

Théorème 2.2.3.1. *[16] Supposons que les arrivées se font selon un processus de Poisson. Alors chaque client entrant verra une distribution de clients déjà présents dans le système égale à celle vue en régime permanent. Avec le théorème ergodique, cela signifie qu'en régime permanent un nouveau client verra n clients devant lui avec une probabilité égale à la proportion du temps pendant laquelle le système est occupé par n clients.*

C'est ce qui est appelé la propriété PASTA (Wolff 1982) : Poisson Arrivals See Time Averages. Notons qu'on ne fait aucune hypothèse sur la loi des services.

Démonstration. Soit N_t le nombre de clients dans le système à l'instant $t > 0$, et Δt un intervalle de temps petit. Notons aussi $A_{t-\Delta t,t}$ le nombre d'arrivées entre $t - \Delta t$ et t. Par la formule de Bayès :

$$\mathbb{P}\left(N_{t-\Delta t} = n | A_{t-\Delta t,t} \geq 1\right) = \mathbb{P}\left(A_{t-\Delta t,t} \geq 1 | N_{t-\Delta t} = n\right) \mathbb{P}\left(N_{t-\Delta t} = n\right) / \mathbb{P}\left(A_{\Delta t,t} \geq 1\right)$$

Comme un Processus de Poisson est sans mémoire (indépendance des nombres d'arrivées sur des intervalles disjoints), $\mathbb{P}\left(A_{t-\Delta t,t} \geq 1 | N_{t-\Delta t} = n\right) = \mathbb{P}\left(A_{t-\Delta t,t} \geq 1\right)$.

Maintenant lorsque Δt tend vers 0, $\mathbb{P}\left(N_{t-\Delta t} = n | A_{t-\Delta t,t}\right)$ est la probabilité que le n clients soient dans le système lorsqu'un nouveau client arrive (on ne compte pas ce nouveau client).

D'après ce qui précède, cette probabilité est égale à

$$\lim_{t \to +\infty} \mathbb{P}\left(N_{t-\Delta t} = n\right) = \mathbb{P}\left(N_t = n\right)$$

Cela signifie que la probabilité pour qu'il y ait n clients dans la file à l'instant t est indépendante du fait que t soit un moment d'arrivée. En passant à la limite ($t \to +\infty$), en régime permanent, cela signifie qu'un nouveau client voit n clients devant lui avec une probabilité égale à la proportion de temps pendant laquelle le système est occupé par n clients. En effet, la première probabilité est :

$$\lim_{t \to +\infty} \lim_{\Delta t \to 0} \mathbb{P}\left(N_{t-\Delta t} = n | A_{t-\Delta t, t} \geq 1\right)$$

et la seconde :

$$\lim_{t \to +\infty} \mathbb{P}\left(N_t = n\right)$$

2.2.4 Caractéristiques d'un système de files d'attente markoviennes

La plupart du temps, l'arrivée des clients à une file simple est supposée décrite par un processus de renouvellement. Le processus d'arrivée le plus simple et le plus couramment employé est le processus de Poisson. C'est un processus de renouvellement, tel que les interarrivées sont distribuées selon une loi exponentielle. On note λ le taux des arrivées : $\frac{1}{\lambda}$ est l'intervalle moyen entre deux arrivées consécutives.

Nous prenons quelques exemples sur les files d'attente markoviennes.

2.2.5 Modèle d'attente $M/M/1$

2.2.5.1 Description du modèle

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 exponentiel 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 interarrivées sont indépendants. Ces clients reçoivent le service selon une distribution exponentielle de paramètre μ. Ces temps de service sont également supposés indépendants. En outre, toutes les variables aléatoires concernés sont censés être indépendantes les unes des autres. Les clients sont servis dans leur ordre d'arrivée. Cette file est un cas particulier du processus de naissance et de mort, dans lequel les probabilités de naissance et de mort sont constantes, voir figure (2.2).

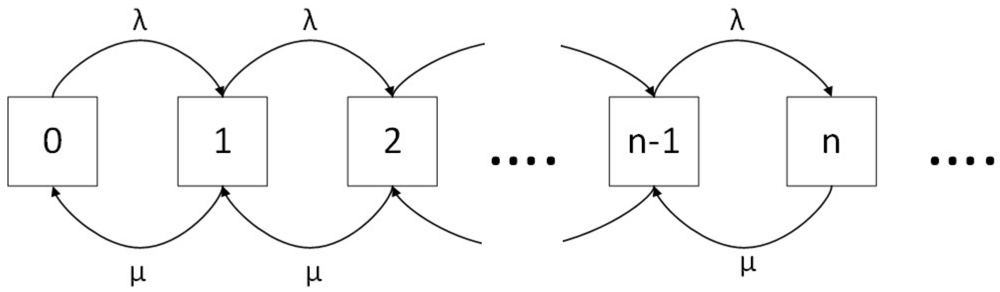


FIGURE 2.2 – Graphe de transition d'une file $M/M/1$

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

Le processus $(N(t); t \geq 0)$ est un processus de naissance et de mort avec un taux de naissance $\lambda_i = \lambda$, pour tout $i \geq 0$ et le taux de mortalité $\mu_i = \mu$ pour tout $i \geq 1$.

En raison de la distribution exponentielle des temps interarrivée et des temps de service, le processus $(N(t); t \geq 0)$ est un processus de Markov. D'autre part, étant donné que la probabilité d'avoir deux évènements (départs, arrivées) dans l'intervalle de temps $(t; t+h)$ est $o(h)$, nous avons

$$\begin{aligned}
P(N(t+h)=i+1|N(t)=i) &= \lambda h+o(h) \quad \forall i\geq 0, \\
P(N(t+h)=i-1|N(t)=i) &= \mu h+o(h) \quad \forall i\geq 1, \\
P(N(t+h)=i|N(t)=i) &= 1-(\lambda+\mu)h+o(h) \quad \forall i\geq 1, \\
P(N(t+h)=i|N(t)=i) &= 1-\lambda h+o(h) \quad \text{for} \quad i=0, \\
P(N(t+h)=j|N(t)=i) &= o(h) \quad \text{si} \quad |j-i|\geq 2.
\end{aligned}$$

Cela montre que $(N(t); t \geq 0)$ est un processus de naissance et de la mort.

Soit $\pi(i), i \geq 0$, la fonction de distribution du nombre de clients dans le système à l'état d'équilibre.

Les équations d'équilibre pour ce processus sont :

$$\begin{aligned}
\lambda\pi(0) &= \mu\pi(1) \\
(\lambda+\mu)\pi(i) &= \lambda\pi(i-1)+\mu\pi(i+1) \quad \forall i\geq 1
\end{aligned}$$

Nous supposons dans tout le reste sauf mention contraire que,

$$\rho = \frac{\lambda}{\mu} \tag{2.1}$$

La quantité ρ mentionne l'intensité du trafic.

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

$$\begin{cases}
\pi_i = \pi_0\rho^i, \\
\pi_0 = \dfrac{1}{\displaystyle\sum_{n=0}^{+\infty}\rho^i} = 1-\rho.
\end{cases}$$

Donc

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

Tous les paramètres de performance sont calculés dans le cas oï¿½ la file d'attente est stable ($\rho < 1$).

Débit d

Le débit du système (noté d) représente la probabilité pour que la file ne soit pas vide donc

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

Ici $d = \lambda$ car $\lambda_i = \lambda$ pour tout $i \geq 0$. Une autre façon de voir les choses est de remarquer que le service s'effectue avec un taux μ dans chaque état où le système contient au moins un client. On retrouve bien que si la file est stable, le débit moyen de sortie est égal au débit moyen d'entrée.

Taux d'utilisation du serveur U

Le taux d'utilisation du serveur (noté U) est la probabilité pour que le serveur de la file soit occupée

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

Nombre moyen de clients L

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

$$
\begin{aligned}
L &= \sum_{i=1}^{+\infty} i\pi_i, \\
&= \sum_{i=1}^{+\infty} i(1-\rho)\rho^i, \\
&= \rho(1-\rho)\frac{d}{d\rho}\left(\frac{1}{1-\rho} - 1\right).
\end{aligned}
$$

Soit

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

Temps moyen de séjour W

Le temps moyen de séjour W se calcule en utilisant la loi de Little :

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

qui peut se décomposer en :

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

On en déduit le 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 :

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

2.2.6 Modèle d'attente $M/M/1/K$

2.2.6.1 Description du modèle

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

Ce modèle de files d'attente est identique à la file $M/M/1$ sauf que la capacité de la file est k, est finie. Soient, comme dans le cas $M/M/1$, λ le taux du processus de Poisson pour les arrivées et μ le paramètre de la distribution exponentielle pour les temps de service. L'espace d'états E est maintenant fini : $E = \{0, 1, 2, \cdots, K\}$.

Donc ce processus est considéré comme un processus de naissance et de mort avec :

un taux de naissance $\lambda_i = \lambda$, pour tout $i < K$ et le taux de mortalité $\mu_i = \mu$ pour tout $i \neq 0$.

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

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

Le calcule de $\pi_{(i)}$ se fait de la manière suivante :

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

Cas particulier $\rho = 1$

Dans ce cas

$$\begin{aligned}\pi_i &= \frac{1}{K+1} \quad \text{pour} \quad i \leq K, \\ \pi_{(i)} &= 0 \quad \text{pour} \quad i > K.\end{aligned}$$

<u>Débit d.</u> On peut calculer le débit du système de deux manières équivalentes : soit par le calcul du taux de départ des clients en sortie du serveur, d_s, soit par le calcul du taux d'arrivée effectif des clients acceptés dans le système d_e.

1. Le débit en sortie du système est égal à μ (le débit du sortie du service) donc :

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

2. Le débit d'entrée dans la file est égal au taux d'arrivée à la station λ :

$$\begin{aligned}d_e &= \mathbb{P}([\text{file non pleine aux instants d'arrivée}])\lambda \\ &= \sum_{i=0}^{K-1} \pi_i \lambda = (1-\pi_K)\lambda = \frac{1-\rho^K}{1-\rho^{K+1}}\lambda\end{aligned}$$

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

<u>Taux d'utilisation du serveur $U(K)$</u> :

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

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

Remarque :

$$\lim_{K\to+\infty} U(K) = \begin{cases} \rho & \text{pour} \quad \rho < 1 \\ 1 & \text{pour} \quad \rho > 1 \end{cases}$$

qui représente le taux d'utilisation du serveur dans le cas de la file $M/M/1$

Nombre moyen de clients L :

$$\begin{aligned} L &= \sum_{i=0}^{K} i\pi_i = \frac{1-\rho}{1-\rho^{K+1}} \sum_{i=0}^{K} i\rho^i \\ &= \frac{\rho(1-\rho)}{1-\rho^{K+1}} \sum_{i=1}^{K} i\rho^{i-1} \\ &= \frac{\rho(1-\rho)}{1-\rho^{K+1}} \frac{d}{dp} \left(\frac{1-\rho^{K+1}}{1-\rho} - 1 \right) \\ &= \frac{\rho(1-\rho)}{1-\rho^{K+1}} \frac{1-(K+1)\rho^K + K\rho^{K+1}}{(1-\rho)^2} \\ &= \frac{\rho}{1-\rho} \frac{1-(K+1)\rho^K + K\rho^{K+1}}{1-\rho^{K+1}}. \end{aligned}$$

Pareil que le taux d'utilisation du serveur, lorsque $K \to +\infty$ 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 dans la file se fait en appliquant la formule de Little qui relie Nombre moyen de clients L et le débit d de la file :

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

2.2.7 Modèle d'attente $M/M/s$

2.2.7.1 Description du modèle

Pour ce modèle de file d'attente, il existe s serveurs identiques et indépendants les uns des autres, et une salle d'attente de capacité infinie. 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 paramètre μ.

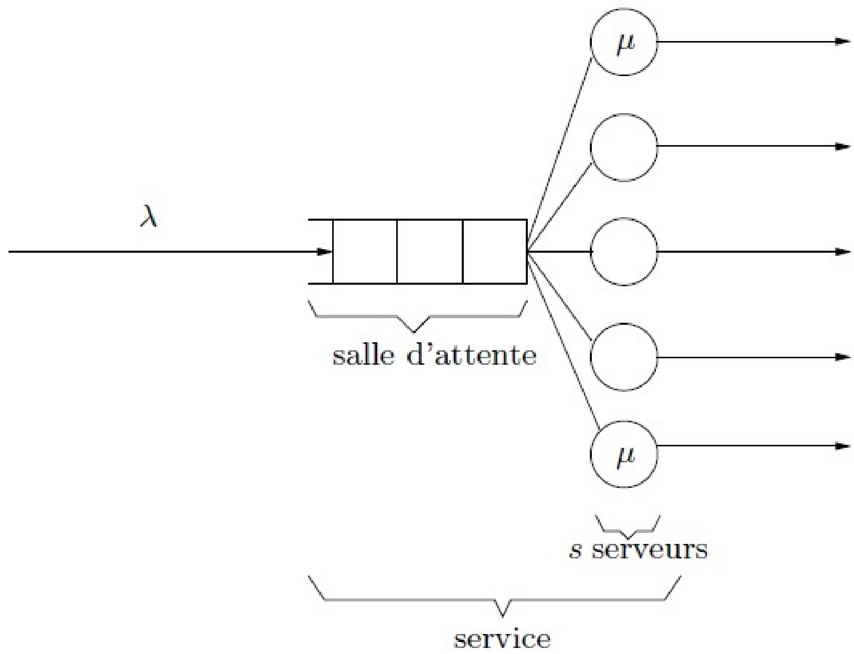


FIGURE 2.3 – Représentation schématique d'une file $M/M/s$

Dans ce cas aussi, le processus modélisant le nombre de clients dans le système est un processus de naissance et de la mort avec :

$$\lambda_i = \lambda$$

$$\mu_i = \begin{cases} i\mu \quad \text{pour} \quad i = 1, 2, \cdots, s-1 \\ \\ s\mu \quad \text{pour} \quad i \geq s \end{cases}$$

Pour calculer $\pi_{(i)}$, le système doit être stabile ($\lambda < s\mu$ 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. Alors

$$\pi_i = \begin{cases} \pi_0 \frac{\rho^i}{i!} \quad \text{pour} \quad i = 1, 2, \cdots, s-1 \\ \\ \pi_0 \rho^i \frac{s^{s-i}}{s!} \quad \text{pour} \quad i \geq s \end{cases}$$

avec

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

Remarque :

pour la valeur de $s = 1$ on obtient :

$$\pi_i = (1 - \rho)\rho^i$$

qui représente le cas de la file $M/M/1$

Débit d Quand le système contient moins de s clients, ces derniers sont pris en charge par les s serveurs à un taux $i\mu$ et avec $s\mu$ un taux dans chaque état où le système contient plus de s clients :

$$d = \sum_{n=1}^{s-1} \pi_n n\mu + \sum_{n=s}^{+\infty} \pi_n s\mu.$$

En remplaçant les expressions obtenues pour les probabilités π_i et π_0, on trouve bien que la file est stable :

$$d = \lambda.$$

Pour des raison de calcul pour la file $M/M/s$, il est plus simple de calculer d'abord le temps moyen de séjour et en suite déduire le nombre moyen de clients.

Temps moyen de séjour W : Le temps moyen de séjour d'un client est composé du temps moyen dans la file d'attente, et le temps moyen de service. Il suffit alors d'appliquer la loi de Little.

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

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

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

On en déduit l'expression du temps moyen de séjour

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

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

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

2.2.8 Modèle d'attente $M/M/\infty$

2.2.8.1 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 ou il n'y a pas d'attente). Dans cette file les clients arrivent à des instants $0 < t_1 < t_2 < \cdots$ formant un processus de poisson de taux λ et les temps de service sont exponentiels de taux μ (pour tous les serveurs). 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. 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 :

$$\lambda_k = \lambda \quad \text{et} \quad \mu_k = k\mu \quad \text{et} \quad k = 0, 1, 2, \cdots$$

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

$$\pi_{n-1} = \pi_n n\mu \quad \text{et} \quad n = 1, 2, \cdots$$

avec

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

et on obtient finalement :

$$\pi_n = \frac{1}{n!}\left(\frac{\lambda}{\mu}\right)^n e^{-\frac{\lambda}{\mu}} \quad \text{pour} \quad n = 1, 2, \cdots$$

par le calcul de π_n on trouve les différents paramètres de performances comme suit :

<u>Débit d</u> le service s'effectue avec un taux $n\mu$ dans chaque état où le système contient n clients :

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

Nombre moyen de clients L :

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

Temps moyen de séjour W :

La loi de Little permet d'écrire :

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

2.2.9 Modèle d'attente $M/G/1$

2.2.9.1 Description du modèle

Pour ce modèle, on étudie une file d'attente composée d'un seul serveur. Le processus d'arrivée de clients est un processus de Poisson d'intensité λ, mais le temps de service indépendants identiquement distribués de loi générale. On notera Y la fonction de répartition de cette loi et, lorsqu'elle est bien définie, y sa fonction de densité. On supposera que la discipline de service est FIFO.

Nous allons commencer par établir l'expression de la probabilité p_0 pour que le serveur soit inoccupé. Le nombre moyen d'arrivées pendant la durée d'un service doit être strictement inférieur à 1, i.e. $\lambda\mathbb{E}(Y) < 1$.

Proposition 2.2.9.1.1. *[16] Dans une file $M/G/1$ avec arrivée Poissonnienne de taux λ et services indépendants de durée moyenne $\mathbb{E}(Y)$, si p_0 est la probabilité pour que le serveur soit inoccupé, alors :*

$$1 - p_0 = \lambda\mathbb{E}(Y)$$

Démonstration. Notons X_n le nombre de clients présents dans le système après le $n$ème départ, A_n le nombre d'arrivées pendant le service du $n$ème client.

Alors :

$$X_{n+1} = X_n + A_n + u(X_n)$$

où $u(X_n) = 1$ si $X_n > 1$ (il y a un départ car dans ce cas le serveur est effectivement occupé) et $u(X_n) = 0$ si $X_n = 0$ (serveur inactif)

Or d'après la propriété PASTA le nombre moyen de clients vu par un nouvel arrivant ne dépend pas de l'instant d'arrivée, donc $\mathbb{E}(X_n) = \mathbb{E}(X_{n+1})$.

Par conséquent : $\mathbb{E}(A_n) = \mathbb{E}(u(X_n))$.

Or $\mathbb{E}(u(X_n)) = 0.\mathbb{P}(X_n = 0) + 1.\mathbb{P}(X_n \geq 1) = 1 - p_0$, et $\mathbb{E}(A_n)$ est le nombre moyen d'arrivées pendant un service, que l'on calcule par :

$$\begin{aligned}
\mathbb{E}(A_n) &= \sum_{k=0}^{n} k \int_0^{+\infty} \mathbb{P}(k \text{ arrivées pendant } t) y(t)dt \\
&= \sum_{k=0}^{n} k \int_0^{+\infty} e^{-\lambda t} \frac{(\lambda t)^k}{k!} y(t)dt \\
&= \int_0^{+\infty} e^{-\lambda t} y(t) \lambda t \left(\sum_{k=0}^{+\infty} \frac{(\lambda t)^k}{k!} y(t) \right) dt.
\end{aligned}$$

Or $\sum_{k=0}^{+\infty} \frac{(\lambda t)^k}{k!} = e^{\lambda t}$,
donc

$$\mathbb{E}(A_n) = \int_{t=0}^{+\infty} \lambda t y(t)dt = \lambda \mathbb{E}(Y)$$

On a bien $1 - p_0 = \lambda \mathbb{E}(Y)$ A partir de là on peut déterminer le nombre moyen de clients dans le système et le temps moyen d'attente.

Proposition 2.2.9.1.2. *[16] Dans une file $M/G/1$, le temps moyen résiduel de service (définie comme la durée restant à un serveur pour terminer son service lorsqu'un nouveau client arrive* *dans la file)* $\overline{t_r}$ *vérifie :*

$$\overline{t_r} = \frac{1}{2}\mathbb{E}(Y) + \frac{Var(y)}{2\mathbb{E}(Y)}$$

où Y *est la durée d'un service.*

Nous allons établir des formules établissant le temps d'attente moyen et le nombre moyen de clients dans une file $M/G/1$. Notons $\tau = \lambda\mathbb{E}(Y)$. On suppose que $\tau < 1$ pour que la file d'attente ne s'allonge pas indéfiniment. D'après la propriété PASTA, un nouveau clients voit $\overline{N_f}$ clients dans la file, qui seront servis avant lui sous la discipline de service FIFO. Chacun de ces clients aura un temps de service moyen de $\mathbb{E}(Y)$.

Donc le temps moyen d'attente d'un nouveau client est : $\overline{N_f}E(Y)$, augmenté de son temps résiduel moyen d'attente, le temps résiduel d'attente étant la différence entre l'instant de la fin du service courant et l'instant d'arrivée.
Le temps résiduel d'attente est 0 si le serveur est inoccupé lorsque le nouveau client arrive (avec une certaine probabilité $1 - \tau$ d'après la proposition et la propriété PASTA) et égal au temps résiduel de service sinon (avec une probabilité complémentaire τ).

Donc : $\overline{T_f} = \overline{N_f}.E(Y) + \tau.\overline{t_r}$. D'autre part, avec la loi de Little, $\overline{N_f} = \lambda\overline{T_f}$, d'ou

$$\overline{T_f} = \frac{\tau\overline{t_r}}{1 - \lambda E(y)}$$

et

$$\overline{N_f} = \frac{\lambda\overline{t_r}}{1 - \lambda E(y)}$$

Comme le temps moyen de séjour $\overline{T}$ (temps d'attente + temps de service) vérifie $\overline{T} = \overline{T_f} + E(Y)$

et le nombre moyen de client dans le système vérifie $\overline{N} = \lambda \overline{T}$, on conclut :

Proposition 2.2.9.1.3. *[16] (Formules de Pollaczek Khinchine). Le nombre moyen de clients dans une file $M/G/1$ avec des arrivées Poissonniennes de taux λ et une durée de service Y de loi quelconque (admettant au moins une espérance et une variance.) est :*

$$\overline{N} = \tau + \frac{\tau^2(1 + Var(y)/\mathbb{E}(y)^2)}{2(1-\tau)}$$

où $\tau = \lambda \mathbb{E}(y)$

Le temps moyen de séjour dans la file est :

$$\overline{T} = \mathbb{E}(y)(1 + \frac{\tau(1 + Var(y)/\mathbb{E}(y)^2)}{2(1-\tau)}$$

Remarque : Pour le cas particulier $M/M/1$, on a $Var(y)/\mathbb{E}(y)^2 = 1$ et on retrouve les formules déjà établies.

2.2.10 Modèle d'attente $GI/GI/1$

2.2.10.1 Description du modèle

Pour ce modèle, on étudie une file d'attente composée d'un seul serveur. Les clients arrivent aux instants $t_n, n \in \mathbb{N}$ et le $n^{\text{ième}}$ client obtient son service à σ_n. Les suites des interarrivées $(\tau_n) = (t_{n+1} - t_n)$ des clients à la file d'attente et de leurs services (σ_n) sont supposées être i.i.d. et indépendantes ; τ_n est la durée entre les arrivées des clients d'indice n et $n-1$. Les arrivées multiples sont exclues et donc $\mathbb{P}(\tau_0 > 0) = 1$. Sous cette hypothèse, le processus ponctuel d'arrivée est donc un processus de renouvellement. La charge de la file sera notée

$$\rho = \frac{\mathbb{E}(\sigma)}{\mathbb{E}(\tau)} = \lambda \mathbb{E}(\sigma)$$

Pour $n > 0$ W_n est le temps d'attente du $n-$ ième client quand le client d'indice o attend la quantité w. À l'instant $t_n + W_n + \sigma_n$ le $n-$ ième client quitte la file d'attente, l'instant de début de service du $n+1-$ ième client, $t_{n+1} + W_{n+1}$ est donc égal à $max(t_{n+1}, t_n + W_{n+1} + \sigma_n)$. La suite (W_n) vérifie donc la relation suivante,

$$W_o = w, \quad W_n = (W_{n-1} + \sigma_n - \tau_n)^+, \quad \text{quand} \quad n > 1 \tag{2.2}$$

avec $a^+ = \max(O, a)$ pour $a \in \mathbb{R}$.

En posant $Xn = \sigma_n - \tau_n$, (S_n) désigne la suite des sommes partielles associées, pour $n > 1$,

$$S_n = \sum_{k=1}^{n} X_k$$

et $S_o = o$. La tribu engendrée par les variables aléatoires $\sigma_1, \ldots, \sigma_n \quad \text{et} \quad \tau_l, \ldots, \tau_n$ est notée $\mathcal{F}_n$.

Le calcul explicite de la loi de la suite (W_n) et celle de son éventuelle limite est le principal sujet d'étude de ce chapitre. Le premier résultat concerne l'existence de cette limite.

Proposition 2.2.10.1.1. *[31]*

1. *Si $\rho < 1$, (Wn) est une chaîne de Markov ergodique. Cette suite converge en loi vers une unique variable W telle que*

$$W = (W + X_o)^+ en\ loi, \tag{2.3}$$

avec W et $X_o = \sigma_0 - \tau_o$ indépendants. De plus $\mathbb{P}(W = 0) > 0$ et W a même loi que le *maximum de la marche aléatoire associée à* (X_n)*,*

$$W = \sup_{n>0} S_n \quad \textit{en loi.} \quad n > O$$

2. *Si* $\rho > 1$*, la chaîne de Markov est transiente et* $\mathbb{P}$*-presque sûrement,*

$$\lim W_n/n = \mathbb{E}(\sigma) - \mathbb{E}(\tau).$$

DÉMONSTRATION.

Par récurrence sur la relation (2.2),

$$W_n = (W_{n-1} + X_n)^+,$$

il est clair que (X_n, W_{n-1}) sont des variables indépendantes et donc que (W_n) a la propriété de Markov. En itérant (2.2), l'identité suivante s'obtient facilement par récurrence,

$$W_n = \sup_{2 \leq k \leq n+1} \left(\sum_{i=k}^{n} X_i \right) V \left(w + \sum_{i=1}^{n} X_i \right) \tag{2.4}$$

Si $\mathbb{E}(\sigma - \tau) < 0$, la loi des grands nombres montre que $\sum_{i=1}^{n} X_i/n$ converge p.s. vers la moyenne de X_l, $\mathbb{E}(\sigma - \tau) < 0$. En particulier, $\mathbb{P}$-presque sûrement, la quantité

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

tend vers $-\infty$ et donc ne contribue plus à la borne supérieure (2.4) à partir d'un certain rang ; $\mathbb{P}$-presque sûrement la variable W_n ne dépend plus de w pour n suffisamment grand. Les hypothèses d'indépendance et d'équidistribution des suites de variables (σ_i) et (τ_i) donnent l'identité

$$(X_l, X_2, \dots X_n) = (X_n, X_{n-1} \dots X_l) \quad \text{en loi},$$

la relation (2.4) permet d'obtenir l'égalité

$$W_n = \sup_{0 \leq k \leq n-1} S_k V(w + S_n). \tag{2.5}$$

La suite (W_n) converge donc en loi vers $\sup S_n/n > O$ qui est fini $\mathbb{P}$-p.s. puisque la marche aléatoire converge p.s. vers $-\infty$. La chaîne de Markov (W_n) a donc une probabilité invariante et toutes ces trajectoires se rejoignent indépendamment du point initial, d'ou (W_n) est une chaîne de Markov ergodique. L'équation (2.3) est l'équation de mesure invariante de cette chaîne de Markov. Si $\mathbb{P}(W = 0) = 0$, l'équation (2.3) peut s'écrire

$$W = W + X_o \quad \text{en loi},$$

en prenant la transformée de Fourier, on obtient $\mathbb{E}(\exp(\xi X)) = 1$ pour $Re(\xi) = 0$, soit $X = 0$, $\mathbb{P}$-p.s. d'après l'unicité de la transformée de Fourier, donc $\mathbb{E}(X) = 0$ ou encore $\rho = 1$, contradiction. La partie a) est montrée.

Si $\rho > 1$, l'identité (2.4) montre que $W_n > w + \sum_1^n X_i$, en utilisant encore la loi des grands nombres, $\mathbb{P}$-p.s.

$$\liminf \frac{W_n}{n} \geq \mathbb{E}(\sigma) - \mathbb{E}(\tau) > O.$$

En particulier, $\mathbb{P}$-p.s. $W_n > 0$ à partir d'un certain rang (aléatoire) n_o, en sommant la relation (2.2) entre n_o et n, il vient pour $n \geq n_o$

$$W_n = W_{n_o} + \sum_{n_o+1}^{n} X_i$$

d'où la dernière assertion de la proposition.

2.3 Exercices

Exercice 01 :

Une compagnie d'assurance est ouvert chaque jour de 9h à 17h sans pause. Une moyenne de 64 clients par jour sont accueillés dans cette compagnie. 2,5 minutes est le temps moyen du traitement d'un dossier par un seul agent. Les clients sont classés dans la file dans l'ordre de leur arrivée. Les durées des services suit une loi exponentielle et les arrivées forment un processus de Poisson.

1. Donner la notation de Kendall de cette file.
2. Donner l'expression de la probabilité invariant π_k, donner la justification de son existence.
3. Quel sont les temps moyens passés à attendre ? dans l'organisme par chaque client ?
4. Quelles sont les probabilités qu'il n'arrive aucun client entre 15H et 16H ? Que 6 clients arrivent entre 16H et 17H ?
5. Quelle est, en moyenne et par heure, la durée pendant laquelle l'employé du guichet ne s'occupe pas des clients ?
6. Quelle est la probabilité d'observer une file d'attente de 4 clients, derrière celui en cours de service ?

corrigé :

1. $M/M/1/\infty$ avec $\lambda = 8$ (clients par Heure) et $\mu = 24$ (clients par Heure).
2. On pose $a = \dfrac{\lambda}{\mu}$, soit $a = \dfrac{1}{3}$. Comme $a < 1$, la distribution stationnaire existe ; la probabilité qu'il ait k clients dans le système à un instant donné est $\pi_k = (1 - a)a$

3. Le temps , T, passé par un client dans le système (attente + service) suit une loin exponentiel de paramètre $\mu - \lambda$ i,e : $E(T) = \dfrac{1}{\mu - \lambda} = \dfrac{1}{16}$ heure.
4. $\Rightarrow$ le temps moyen d'attente dans la file est $E(T_q) = E(T) - \dfrac{1}{\mu} = \dfrac{\lambda}{\mu(\mu - \lambda)} = \dfrac{1}{48}$.
5. Probabilité qu'il n'arrive aucun client entre 15H et 16H. On sait que le nombre N de clients arrivant dans le système pendant une unité de temps est une v. a. qui suit la loi de Poisson, i.e.

$$P(N = K) = exp(-\lambda)\frac{\lambda^k}{k!}$$

Pour $k = 0$, on obtient $P(N = 0) = exp(-8)$. Pour $k = 6$, on obtient $P(N = 6) = exp(-8)\dfrac{8^6}{6!}$.

6. L'employé est inoccupé lorsque le système est dans l'état 0, ce qui correspond à la probabilité $\pi_0 = 1 - a = \dfrac{2}{3}$, soit en moyenne 40 minutes par heure.
7. S'il y a quatre usagers dans la file d'attente, il y a 5 clients en tout dans le système, ce qui correspond à la probabilité $\pi_5 = (1 - a)a^5 = \dfrac{2}{3^6}$

Exercice 02 :

Une importante maternité accueille des femmes enceintes qui sont arrivées à terme et viennent accoucher et donner naissance à leur bébé. L'occupation moyenne d'une salle de travail est de 6 heures pour un accouchement. Un statisticien a déterminé que la loi d'arrivée des futures mamans dans une maternité pour y accoucher, peut être approximée de façon satisfaisante, par une loi de Poisson, (il se présente en moyenne 24 femmes par jour) et que celle de l'occupation de la salle de travail peut l'être par une loi exponentielle. Leurs taux respectifs valent λ et μ. Le but est de déterminer le nombre N de salles de travail (et, par conséquent le nombre minimal de sages femmes devant se trouver dans la maternité), de telle sorte que la probabilité pour que toute les salles soient occupées soit inférieure à un centième : une femme qui arriverait dans ce cas serait dirigée vers une autre maternité, ce que l'on désire éviter à l'extrême.

1. Donner la valeur numérique de λ et μ.
2. Montrer que le système d'attente est un processus de naissance et de mort, comportant $N+1$ états, numérotés de 0 à N. A quoi correspondent, ici une naissance et une mort ? Donner la signification de l'état k, exprimer λ_k en fonction de λ et μ_k en fonction de μ.
3. La probabilité pour que le système soit, en régime permanent, dans l'état k est notée π_k . Calculer π_k en fonction de π_0,
4. Quel est l'état pour lequel toutes les salles sont occupées ; quelle est sa probabilité ? Trouver le nombre de salles de travail que devra comporter la clinique, de sorte que la probabilité pour qu'elles soient toutes occupées soit inférieure à 0,01.

corrigé :

1. C'est une file $M/M/s/s$ ou s est le nombre de serveurs, dans notre cas le nombre N de salles d'accouchement. Notons que $s = N$ est un nombre inconnu puisque la question est justement de calculer N pour garantir une certaine qualité de service (voir question 4). On a $\lambda = 24$ (clients par jour) et $\mu = 4$(accouchements par salle). Le système est dans l'état k lorsqu'il y a k salles occupées simultanément (k est un entier naturel).
2. Le système, $M/M/s/s$, est modélisé par un processus de naissance et de mort tel que : $\lambda k = \lambda$ et $\mu_k = k\mu$, $\forall k \geq 0$
3. La distribution stationnaire π_k est donnée par la formule (voir cours) :

$$\begin{cases} \pi_0 = \left(\displaystyle\sum_{n=0}^{N} \frac{a^n}{n!} \right) \\ \pi_k = \dfrac{a^k}{k!} \pi_0 \end{cases}$$

4. Le sytème est saturé lorsque toutes les salles sont occupées, i.e. lorsque $k = N$. D'après la formule précédente, ceci se produit avec la probabilité

$$\pi_N = \frac{a^N}{N!}\pi_0$$

que l'on note traditionnellement $B(N, a)$ et qui représente la proportion de clients refusés. On connait $a = \frac{\lambda}{\mu} = 6$ et l'inconnue est le nombre de salles N. Il faut donc résoudre l'inéquation :$B(N, a) = 0.01$. On ne peut résoudre cette équation que par ordinateur. On programme une procédure qui renvoie $B(N, a)$ et on effectue la boucle de calcul, on trouve $N = 9$.

Exercice 03 :

On considère une chaîne de Markov modélisant K machines indépendantes. Chacune d'elles peut tomber en panne avec un taux λ. Un réparateur répare une seule machine à la fois avec le taux μ.

1. Exprimer le taux de passage de l'état (n machines en panne) à l'état ($n + 1$ machines en panne).

   On fera une démonstration rigoureuse. Il sera peut-être nécessaire de démontrer au préalable un lemme sur la loi de l'inf d'un certain nombre de variables de loi exponentielle.
2. Calculer la probabilité d'avoir K machines en panne.
3. Application numérique : $K = 7$, $\lambda = 0.001$, $\mu = 1$.

Corrigé

Nous considérons une chaîne de Markov à K états. L'état i est atteint lorsque i machines parmi les K sont en pannes.

1. Lemme sur la loi de l'inf de n variables indépendantes de loi exponentielle Soit n variables

aléatoires indépendantes X_i de loi exponentielle de paramètre λ_i.

$$\begin{aligned}
P[\inf(X_i) \leq x] &= 1 - P[Inf(Xi) > x] \quad \text{(Passage au complémentaire)} \\
&= 1 - \prod_{i=1}^{n} P[X_i > x] \\
&= 1 - \prod_{i=1}^{n} e^{-\lambda_i x} \\
&= 1 - e^{\sum_{i=1}^{n} \lambda_i x}
\end{aligned}$$

Lorsque x tend vers 0, nous avons donc :

$$P[\inf(X_i) \leq x] = \sum_{i=1}^{n} \lambda_i x + o(x)$$

Expression du taux de passage de l'état n à l'état $n+1$

Nous supposons donc qu'il y a $K-n$ machines en état de marche. Soit X_i le temps qui sépare l'instant courant de la panne de la machine i. La première machine tombera en panne au bout d'un temps $\inf(X_i)$.

La probabilité de passer de l'état n à l'état $n+1$, c'est à dire $P[X(t+dt) = n+1/X(t+dt) = n]$ est

$$P[X(t+dt) = n+1/X(t) = n] = (K-n)\lambda dt + o(dt)$$

Expression du taux de passage de l'état $n+1$ à l'état n

Le réparateur est exponentiel et l'on passe donc de l'état $n+1$ à l'état n avec le taux μ.

2. On a bien un procesus sans mémoire, qui de plus, évolue uniquement par saut de 1 ou -1. C'est donc un processus de naissance et de mort.

On a donc :

$$P(n) = \frac{\lambda(0)\lambda(1)\cdots\lambda(n-1)}{\mu(1)\mu(2)\cdots\mu(n)}P(0) = \frac{K!}{(K-n)!}\left(\frac{\lambda}{\mu}\right)^n P(0)$$

La probabilité d'avoir K machines en pannes est donc :

$$P(K) = \frac{K!\left(\frac{\lambda}{\mu}\right)^K}{\sum_{n=0}^{K}\frac{K!}{(K-n)!}\left(\frac{\lambda}{\mu}\right)^n}$$

soit :

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

d'où en réindexant $(n = K - n)$:

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

3. On trouve : $P(7) = 4.7\ 10^{-11}$.

Exercice 04 :

Le nombre de clients par heure dans un magasin suit un processus de Poisson de taux 4.

1. Quelle est la probabilité qu'il y ait strictement moins de deux clients la première heure ?
2. Si dans la première heure 6 clients sont arrivés , calculer la probabilité d'avoir pendant la deuxième heure, moins de deux clients.
3. Le résponsable du magasin donne une pause aux serveurs après avoir servi dix clients. Quelle est la durée moyenne séparant deux pauses ?

4. Trois quarts des clients sont des hommes, un quart des femmes, ceci indépendamment de l'heure d'arrivé. Calculer la probabilité qu'en une heure il y ait exactement 2 hommes et 3 femmes à visiter le magasin.

Corrigé

Le nombre de clients par heure suit un processus de Poisson de taux 4.

1. Le nombre de clients au bout d'une heure suit une loi exponentielle de paramètre 4. La probabilité qu'il y ait strictement moins de 2 clients au bout d'une heure est celle qu'il y en ait 0 ou 1 :
$$p = e^{-4}\frac{4^0}{0!} + e^{-4}\frac{4^1}{1!} = 5e^{-4}$$
2. Le fait qu'il y ait eu 6 clients la première heure ne change rien à la probabilité qu'il y en ait strictement moins de deux la deuxième heure : c'est toujours $p = 5e^{-4}$.
3. Le temps nécessaire (en heures) pour avoir 10 clients est T_{10}, somme de 10 exponentielles indépendantes de paramètre 4 : c'est donc une loi $\Gamma(10, 4)$. Sa moyenne est tout simplement :
$$d = \mathbb{E}[T_{10}] = \frac{10}{4} = 2\text{h}30.$$
4. Soit p la probabilité cherchée. Avec des notations évidentes, on peut décomposer p comme suit :
$$p = \mathbb{P}(2H, 3F|N_1 = 5)P(N_1 = 5).$$
Le second terme est la probabilité qu'il y ait exactement 5 clients en 1 heure, c'est donc :
$$P(N_1 = 5) = e^{-4}\frac{4^5}{5!}.$$

Le premier terme est la probabilité que sur 5 clients, 2 soient des hommes et 3 femmes. Or, sur 5 clients, le nombre de clients hommes suit une loi binômiale $B(5, 3/4)$. Ceci donne :

$$P(2H, 3F|N_1 = 5) = C_5^2 \left(\frac{3}{4}\right)^2 \left(\frac{1}{4}\right)^3$$

Au total, la probabilité cherchée est donc :

$$p = \left(\frac{3}{4}\right) e^{-4}$$

Exercice 05 :

Dans un centre de commutation les messages arrivent de manière aléatoire à un taux moyen de 240 messages par minute (les temps des interarrivées sont exponentiel). La ligne a une vitesse de transmission de 800 caractères par seconde. La distribution de la longueur du message (y compris les caractères de contrôe) est exponentielle avec une longueur moyenne de 176 caractères.

1. Calculer le temps moyen de service μ_s et le taux d'arrivé λ.
2. Calculer Taux d'utilisation du serveur U.
3. Calculer le nombre moyen de messages dans le système L et le nombre moyen de messages dans le système L_q.
4. Calculer le temps moyen de séjour dans le système W et le temps moyen de séjour dans la file W_q.

Corrigé

1. le temps moyen de service μ_s pour transmettre un message

   $\mu_s = \frac{\text{longueur moyenne d'un message}}{\text{vitesse de transmission de la ligne}} = \frac{176}{800} = 0.22\text{secondes}$

le taux d'arrivé

$\lambda = \dfrac{240}{1\text{minute}} = 4\text{messages par seconde.}$

2. Le taux d'utilisation du serveur

   $U = \lambda \times \mu_s = 4 \times 0.22 = 0.88,$

3. Le nombre moyen de messages dans le système

   $L = \dfrac{U}{1-U} = 7.33\text{messages}$

   Le nombre moyen de messages dans la file

   $L_q = \dfrac{U^2}{1-U} = 6.45\text{messages}$

4. le temps moyen de séjour dans le système

   $W = \dfrac{\mu_s}{1-U} = 1.83\text{secondes}$

   le temps moyen de séjour dans la file

   $W_q = U \times W = 1.61\text{secondes}$

Exercice 06 :

Des camions arrivent dans une station service pour passer des tests de sécurité, suivant un processus de Poisson de taux de 6 / jour. La durée des tests pour chaque camion est une v.a exponentielle d'espérance mathématique de 1h 30 mn. On suppose que le processus d'arrivée ne s'interrompt pas et que la station travaille 24 heures sur 24.

Le système admet-il une distribution stationnaire ?

Si oui la calculer et donner le nombre moyen d'usagers dans le système, le temps moyen passé dans le système, la longueur moyenne de la file d'attente et le temps moyen passé dans la file (en régime stationnaire).

Exercice 07 :

Un centre de traitement reçoit et traite des données qui arrivent selon un processus de Poisson avec un taux de 1250 données/seconde. Le traitement des données suit la loi exponentielle. Une étude a permit d'observer que 35% du temps, le serveur est libre et aucune donnée n'a été rejetée par manque d'espace de stockage.

1. Écrire la notation de Kendall de ce système. Justifiez
2. Quel est la probabilité pour que le serveur de la file soit occupé ?
3. Déterminer la probabilité trouver moins de 03 données dans le système.
4. Calculez le temps d'attente moyen d'une donnée avant d'être traitée.
5. Calculez le nombre moyen de données dans ce système.

Corrigé

1. la notation de Kendall pour ce système est $M/M/1$ parce que :
   (a) Les rrivées sont Poissonniène.
   (b) Le service suit une loi exponentielle.
   (c) Un seul serveur (un centre de traitement).
   (d) Discipline FIFO (par défaut).
   (e) Capacité illimitée.
   (f) Population illimitée (par défaut).
2. la probabilité pour que le serveur de la file soit occupé est rien d'autre que le taux d'utilisation du serveur : le fait que 35% du temps, le serveur est libre signifie que $\pi_0 = 0.35$.
   Or $U = \sum_{i=1}^{+\infty} \pi_i = 1 - \pi_0 = \rho = 0.65$.
3. La probabilité de trouver moins de 03 données dans le système signifie que soit le système est vide, soit il contient une seule donnée soit il contient 02 données.
   Ainsi $\mathbb{P}_{i<3} = \mathbb{P}(i = 0 \vee i = 1 \vee i = 2) = \mathbb{P}_0 + \mathbb{P}_1 + \mathbb{P}_2 = \mathbb{P}_0 + \rho\mathbb{P}_0 + \rho^2\mathbb{P}_0$.
   Et donc $\mathbb{P}_{i<3} = 0.35 + 0.65 \times 0.35 + 0.652 \times 0.35 = 0.35 + 0.2275 + 0.147875 = 0.725375$.

4. Temps d'attente moyen d'une donnée avant d'être traitée :

$$W = \frac{\rho^2}{\lambda(1-\rho)} = \frac{0.65^2}{0.35 \times 1250} = 0.000965s$$

5. Nombre moyen de données dans ce système (en attente et en service) :

$$L = \frac{\rho}{1-\rho} = \frac{0.65}{0.35} = 1.857142s.$$

Exercice 08 :

On voudrait installer un certain nombre de guichets dans un bureau de poste. Les usagers arrivent suivant un processus de Poisson de taux de 35 à l'heure. On suppose que la durée de service suit une loi exponentielle d'espérance mathématique de 3 minutes. Combien de guichets sont-ils nécessaires pour que le système admette un régime stationnaire ? Quel est le nombre minimum de guichets nécessaires pour que le nombre moyen d'usagers (dans le système) en régime stationnaire ne dépasse pas 7 ? On décide d'installer deux guichets, chacun de taux de service de 30 à l'heure. Déterminer alors : la distribution stationnaire, le nombre moyen d'usagers dans le système et le nombre moyen d'usagers en attente, le temps moyen passé dans le système par usager.

Exercice 9 :

Un système clients-serveur reçoit en moyenne 1000 requêtes par seconde, arrivant selon un processus de Poisson. Il dispose d'un unique serveur pouvant traiter en moyenne 2000 clients par seconde. On suppose que le temps de service d'un client est distribué selon la loi exponentielle. Calculer la probabilité que le temps de service dépasse 2 ms.

1. Quelle est le pourcentage de clients rejetés pour un système ne comportant pas de file d'attente.

2. Même question pour un système comportant une file d'attente de 1 place. Calculer le taux d'application du serveur.

3. Même question pour un système comportant une file d'attente de 2 places.

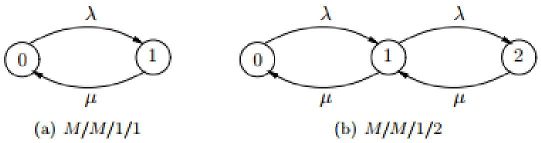


FIGURE 2.4 – (a)$M/M/1/1$ (b) $M/M/1/2$

<u>**Corrigé**</u>

1. En choisissant la milliseconde comme unité de temps on obtient un taux de naissance de 1 et un taux de décès de 2. Le temps de service TS d'un client suit une loi exponentielle de paramètre 2 donc la probabilité que le temps de service dépasse deux secondes est de 0.0183 grâce à la formule suivant :

   La chaine (a) représente une capacité dans la file de 0 place tandis que la chaine (b) représente une capacité dans la file de 1 place. Pour chaque nouvelle place il suffit de rajouter un état.

   Pour la première chaine, nous allons calculer la distribution stationnaire :

$$\begin{cases} \pi_0 = \frac{\mu}{\lambda+\mu} = \frac{2}{3} \\ \pi_1 = \frac{\lambda}{\lambda+\mu} = \frac{1}{3} \end{cases}$$

   Nous avons donc $1/3$ de clients rejetés (si on se trouve $1/3$ du temps en état 1, alors on ne peut plus recevoir des clients sur $1/3$ du fonctionnement de la file). Le taux d'occupation est $pi_0 = 2/3$.

2. Pour la deuxième chaine, nous allons calculer la distribution stationnaire :

$$\begin{cases} \pi_0 = \frac{4}{7} \\ \pi_1 = \frac{2}{7} \\ \pi_2 = \frac{1}{7} \end{cases}$$

3. Et pour une capacité de 2 dans la file d'attente nous obtenons $pi_3 = 1/15$ et $pi_0 = 8/15$.

Exercice 10 :[19]

Dans une station service, il y a une seul pompe à essence. Les voitures arrivent à la station selon un processus de Poisson à raison de 20 voitures par heure. Une voiture qui arrive à la station est trouve n voitures à quitte la station immédiatement avec une probabilité $q_n = n/4$, et rejoint la file d'attente avec une probabilité $1 - q_n, n = 0, 1, 2, 3, 4$.. Les voitures sont servies dans leur ordre d'arrivée. Le temps de service (c'est-à-dire le temps nécessaire pour le pompage et le paiement) est exponentielle. Le temps de service moyen est de 3 minutes.

1. Déterminer la distribution stationnaire du nombre de voitures à la station-service.
2. Déterminez le nombre moyen de voitures à la station service.
3. Déterminer le temps de séjour moyen (temps d'attente plus temps de service) des voitures décidant prendre de l'essence à la station.
4. Déterminer le temps de séjour moyen et le temps d'attente moyen de toutes les voitures arrivant à la station essence.

Corrigé

1. Résoudre les équations d'équilibre du système

$$\lambda(1 - q_n)p_n = \mu p_{n+1} \qquad n = 0, 1, 2, 3$$

où $\lambda = \mu = 1/3$, avec la condition de normalisation. Cela donne

$$p_0 = \frac{32}{103}, p_1 = \frac{32}{103}, p_2 = \frac{24}{103}, p_3 = \frac{12}{103}, p_4 = \frac{3}{103}$$

2. $E(L) = \frac{128}{103} \approx 1.24$

3. $E(S) = \frac{384}{71} \approx 5.41$minutes

4. $E(S) = \frac{384}{103} \approx 3.73$minutes

   $E(W) = \frac{171}{103} \approx 1.66$minutes

Chapitre 3

Réseaux de files d'attente et méthodes de renormalisation

Les réseaux de files d'attente ont été intégrés dans la modélisation de divers domaines d'activité dans les années 1960. C'est l'échange de données entre ordinateurs qui a permis une évolution rapide de la théorie des réseaux de files d'attente, notamment Les problèmes concernent les temps de traitements de requêtes sur un système centralisé ou encore les délais, le taux d'occupation des différents nœuds du réseau.
L'aparution de l'Internet et son dévelopement entre les années 70-90 ont poussé les mathématisiens à essayer de modéliser ces réseaux.
Des techniques de renormalisation ont été introduites au début des années 1990, dans le cadre d'étudier les probabilités invariantes des grands réseaux avec perte. Les méthodes de convergence de processus, de calcul stochastique ont fait ainsi progressivement leur entrée dans l'étude de ces réseaux.

Dans ce chapitre, qui a été inspiré par le document de Philippe Robert, nous introduisons un ensemble de travaux relatifs à l'étude des réseaux stochastiques. Les réseaux de files d'attente à forme produit seront donnés (les réseaux de Jackson, le Réseau de Rybko-Stolyar/Lu-Kumar, le Réseau de Bramson et les réseaux Gordon-Newel). En suite les principales définitions et notions relatives à la renormalisation d'un processus sont traitées. Enfin les limites fluides notamment seront données, qui sont les limites des processus renormalisés. et on termine le chapitre par des exercices illustrent ces définitions.

3.1 Réseaux de files d'attente à forme produit

3.2 introduction

3.2.1 Les réseaux de Jackson

Un réseau de Jackson est caractérisé par un ensemble de N files d'attente avec la displine FIFO, i.e. avec la discipline de service premier arrivé premier servi. dans la file d'attente d'ordre i avec $1 < i < N$, le service est exponentiel de paramètre μ_i et l'arrivée des clients dans le réseau à la file i est un processus de Poisson de paramètre λ Une fois le service du client par la file i est terminé, le client rejoint la file j avec probabilité P_{ij} (avec $P_{ii} = 0$ i.e. étant donné dans la file i le client ne peut jamais revenir à la file i) ou quitte définitivement le réseau avec la probabilité résiduelle voir figure 3.1.

La matrice $R = (R_{ij}, i, j = 0, \dots N)$ est définie par, si $i \neq 0$ et $j \neq 0$,

$r_{ij} = P_{ij}$;

$r_{i0} = 1 - \sum_{j=1}^{N} P_{ij}$;

$r_{00} = 1$

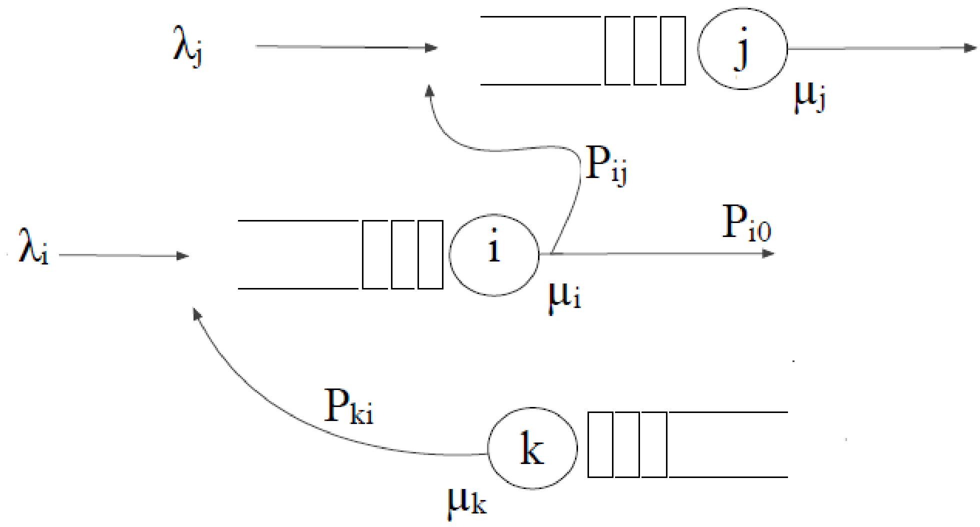


FIGURE 3.1 – Réseau de Jackson

de telle sorte que R est une matrice markovienne. On suppose que toutes les variables aléatoires utilisées sont indépendantes. Le processus de Markov associé à ce réseau de files d'attente est à valeurs dans $S = \mathbb{N}^N$. En notant pour $1 < i < N, e_i = (l_{j=i}; 1 \leq j \leq N)$ le i^{ime} vecteur unité, la matrice Q de ce processus de Markov est donnée par

$$\begin{aligned} q_{n,n+e_i-e_j} &= \mu_j \qquad n_j > 0, i, j \leq N, \\ q_{n,n-e_j} &= \mu_j \qquad n_j > 0, i, j \geq N, \\ q_{n,n+e_i} &= \lambda_i \qquad i \leq N. \end{aligned}$$

La matrice markovienne $R = (R_{ij}; i, J = 0, \ldots, N)$ est supposée avoir 0 comme unique point absorbant, si (Y_n) est une chaîne de Markov de matrice de transition R, presque sûrement (Y_n) est constante égale à 0 à partir d'un certain rang.

On fait en outre l'hypothèse que $r_{ii} = 0$ pour tout $1 \leq i \leq N$ (un client ne revient pas en fin de file d'attente après son service). Si cette condition n'est pas remplie, l'expression du générateur montre qu'il suffit de remplacer μ_i par $\mu_i/(1-r_{ii})$ et les r_{ij} par $r_{ij}/(1-r_{ii})j \neq i$, pour se ramener à cette situation.

Lemme 3.2.1. *([31]) Il existe une suite positive $(\overline{\lambda}_i; 1 \leq i \leq N)$ vérifiant les équations de trafic*

$$\overline{\lambda}_i = \lambda_i + \sum_{j=1}^{N} \overline{\lambda}_j p_{ji}. \tag{3.1}$$

Pour $1 \leq i \leq N$, on posera $\rho_i = \overline{\lambda}_i/\mu_i$

Démonstration (voir exercice 03)

3.2.2 Les réseaux de Gordon-Newel

Le réseau de Gordon-Newel est l'analogue fermé du réseau de Jackson. C'est un réseau de N files d'attente dans lequel circulent M clients. Pour $i = 1, \ldots, N$, la i^{ime} file d'attente délivre un service exponentiel de paramètre μ_i. À la sortie de cette file, un client passe à la file j avec probabilité P_{ij}. La matrice de transition Q du processus de Markov associé à ce réseau est donnée par

$$q_{n,n+e_i-e_j} = \mu_j P_{ji}, \qquad n_j > 0, i, j \leq N.$$

La matrice $P = (p_{ij}; i, j = 1, \ldots, N)$ est la matrice de transition d'une chaîne de Markov irréductible et, par conséquent, ergodique puisque l'espace d'états est fini. Si $(\nu_i; i = 1, \ldots, N)$ est la mesure invariante de cette chaîne,

$$\nu_i = \sum_{1}^{N} \nu_j p_{ji}, \qquad i = 1, \ldots, N \tag{3.2}$$

ces équations sont les analogues des équations du trafic (3.5), de la même manière les charges ρ_i sont définies par

$$\rho_i = \frac{\nu_i}{\mu_i}, \qquad i = 1, \ldots, N \tag{3.3}$$

Le processus de Markov décrivant ce réseau est à valeurs dans

$$S = \{n = (n_i, n = 1, \ldots, N) \sum_{1}^{N} n_i = M\}$$

L'irréductibilité de la matrice de transition P sur $l, \ldots, N$ entraîne celle de ce processus de Markov sur S. De la même façon que précédemment, on définit

$$\pi(n) = \prod_{1}^{N} \rho_i^{n_i}, \qquad n = (n_i, n = 1, \ldots, N) \in S$$

et la matrice du processus stationnaire renversé $\widetilde{Q}$ vaut

$$\widetilde{q}_{n,n+e_i-e_j} = \frac{\nu_i}{\rho_j} p_{ij} \quad n_j > 0, \quad i, j \leq N \tag{3.4}$$

de la même façon que dans la preuve précédente, les équations (3.2) donnent pour $n = (n_i, n = 1, \ldots, N) \in S$ et $j = 1, \ldots, N$ tel que $n_j > 0$,

$$\begin{cases} \sum_{m \in D_j} q_{n,n+m} = \sum_i \mu_j p_{ji} = \mu_j, \\ \sum_{m \in D_j} \widetilde{q}_{n,n+m} = \frac{\sum_i \nu_j p_{ij}}{\rho_j} = \frac{\nu_j}{\rho_j} = \mu_j. \end{cases}$$

Partant de l'état $n \in S$, l'état suivant du processus est nécessairement dans un des $n + D_j, j = 1, \ldots, N$. Les équations (3.2.2) donnent le théorème suivant pour les réseaux fermés.

Théorème 3.2.2.1. *[31] Le réseau fermé de files d'attente de matrice de routage (p_{ij}) décrit* *précédemment a pour mesure invariante la probabilité*

$$\pi(n) = \frac{1}{K}\prod_{1}^{N}\rho_i^{n_i} \qquad n = (n_i, n = 1, \dots, N) \in S$$

où, pour $i = 1, \dots, N$*,* $\rho_i = \nu_i/\mu_i$ *et* (ν_i) *vérifie le système d'équations*

$$\nu_i = \sum_i \nu_j p_{ji}$$

et K *est la constante de normalisation.*

La constante de normalisation K *n'est en général pas très simple à exprimer en raison de la taille de l'espace d'états. Des procédures récursives permettent cependant d'exprimer numériquement* K *assez simplement.*

3.2.3 Le Réseau de Rybko-Stolyar/Lu-Kumar

Ce réseau multi-classe a deux serveurs et quatre files d'attente. Pour $i = 1$ et 2, les clients de classe i arrivent au nœud i suivant un processus de Poisson d'intensité λ_i où ils demandent un service de distribution exponentielle de paramètre μ_{i1} puis passent à l'autre file d'attente pour être servis au taux μ_{i2}, voir figure 3.2

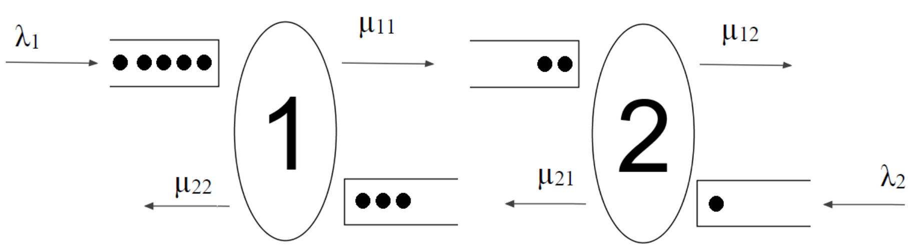


FIGURE 3.2 – Réseau de Rybko-Stolyar/Lu-Kumar

Ce réseau a été étudié par Rybko et Stolyar en 1992 et par Lu et Kumar en 1991 dans une version déterministe. Les clients de classe 1 sont prioritaires quand ils accèdent au serveur 2 : aucun client de classe 2 ne peut être servi si un client de classe 1 est présent dans la file 2. De façon symétrique les clients de classe 2 sont prioritaires dans la file 1. À l'intèrieur d'une classe, le service se fait dans l'ordre des arrivées. Les charges des nœuds 1 et 2 valent donc respectivement

$$\rho_1 = \lambda_1/\mu_{11} + \lambda_2/\mu_{22} \quad \text{et} \quad \rho_2 = \lambda_2/\mu_{21} + \lambda_1/\mu_{12}.$$

Les conditions$\rho_1 < 1$ et $\rho_2 < 1$ sont donc les conditions habituelles de stabilité rencontrées jusqu'alors. Rybko et Stolyar (1992) ont montré cependant qu'elles ne suffisent pas et qu'une condition supplémentaire croisée,

$$\lambda_1/\mu_{12} + \lambda_2/\mu_{22} < 1 \tag{3.5}$$

est nécessaire pour assurer l'ergodicité du processus de Markov associé dans $\mathbb{N}^4$. La figure 3 représente les trajectoires asymptotiques de $((L_1(t), L_2(t))$ quand les conditions $\rho_1 < 1$ et $\rho_2 < 1$ sont satisfaites mais pas la relation (3.5). On a fait l'hypothèse $\mu_{21}+\mu_{21} = +\infty$ pour simplifier, ce qui signifie qu'à leur arrivée, les clients demandent un service nul dans la première file d'attente où ils arrivent.

Habituellement, dans les réseaux à forme produit, quand la condition de charge plus petite que 1 n'est pas satisfaite, le nombre de clients d'au moins un des nœuds tend vers l'infini presque sûrement. L'instabilité des réseaux ayant une charge plus petite que 1 est différente : la condition de charge plus petite que 1 implique que chaque nœud du réseau se vide une infinité de fois presque sûrement mais la durée entre les retours à l'état vide croît linéairement avec le temps. Ainsi le nombre de clients à chacun des nœuds du réseau oscille entre des valeurs de plus en plus grandes. Globalement en effet, le nombre total de clients tend presque sûrement vers l'infini.

3.2.4 Le Réseau de Bramson

Le Réseau de Bramson. Ce réseau est constitué de deux files d'attente servant tous les clients dans l'ordre de leurs arrivées. Les clients arrivent suivant un processus de Poisson d'intensité λ à la file 1. Après cette étape, un client passe à la file 2 pour effectuer $J-2$ services exécutés séparément : après le k-ième service, $1 \leq k \leq J-2$, le client se replace en fin de file d'attente pour recevoir le $k+1$-ième service. Après l'étape $J-1$ à la file 2, le client rejoint la file 1 pour ensuite quitter définitivement le réseau. Pour $1 \leq i \leq J$, un client reçoit un service dont la durée a une distribution exponentielle de paramètre μ_i. Le service d'un client dépend donc de son étape dans son trajet à travers le réseau.

Il n'est pas difficile de constater qu'un processus de Markov de dimension finie ne peut décrire ce réseau puisqu'il faut connaître la classe de chaque client de la file d'attente pour déterminer le taux auquel il sera servi. L'étude de l'ergodicité d'un tel processus de Markov est très délicate, la condition exacte d'ergodicité n'est d'ailleurs pas connue, même dans des cas simples, dès que $J \geq 3$. Bramson a montré que le réseau était aussi instable avec des paramètres λ, J et $(\mu_j; 1 \leq j \leq J)$ pour lesquels les charges à chaque nœud $\lambda/\mu_1 + \lambda/\mu_J$ et $\lambda/\mu_2 + \ldots + \lambda/\mu_{J-1}$ sont strictement plus petites que 1. La divergence du réseau est similaire à celle du réseau de Rybko et Stolyar.

Ces exemples montrent que l'hétérogénéité seule peut déstabiliser un réseau : même si, pour chaque nœud du réseau, la charge moyenne de travail qui arrive est strictement plus petite que sa capacité, le réseau peut osciller de telle sorte que le nombre total de requêtes dans le réseau diverge. Pour ces contre-exemples, chaque nœud du réseau se vide une infinité de fois mais globalement le réseau diverge, cette situation est impossible dans les réseaux classiques. Ces réseaux avec des trafics hétérogènes sont regroupés sous l'appellation réseaux multi-classe.

Dans ce qui suit, $(X(x,t))$ est un processus markoviens de sauts càdlàg, irréductible sur un espace d'états dénombrable S qui part de $x \in S$, i.e. tel que $X(x,O) = x \in S$; on utilise aussi la notation $(X(t))$ pour ce processus s'il n'y a pas d'ambiguïté sur le point initial. Comme d'habitude la notation $\mathcal{N}_\xi(w,dx)$, $w \in \Omega$, désigne un processus de Poisson sur $\mathbb{R}$, de paramètre $\xi \in \mathbb{R}^+$, et tous les processus de Poisson utilisés sont indépendants. La topologie de Skorokhod sur l'espace des probabilités sur l'ensemble des fonctions càdlàg $D([O,T],\mathbb{R}^d)$ est utilisée constamment.

3.3 Renormalisation des processus

Définition 3.3.1. *[31] Si f est une fonction strictement positive sur S et $x \in S$, on note $(\|X\|_f(x,t))$ le processus défini par*

$$(\|X\|_f(x,t)) = \frac{f(X(x,f(x)t))}{f(x)}, \tag{3.6}$$

pour $x \in S$ et $t > o$. Si S est inclus dans un espace vectoriel sur $\mathbb{R}$, $(\overline{X}_f(x,t))$ désigne le processus $(X(x,t))$ renormalisé par la fonction f

$$(\overline{X}_f(x,t)) = \frac{1}{f(x)} X(x,f(x)t). \tag{3.7}$$

S'il n'y a pas d'ambiguïté sur le point initial x ou la fonction f, l'indice correspondant sera omis, i.e. les notations $(\overline{X}(x,t))$ et $(\|X\|(x,t))$ ou $(\overline{X}_f(t))$ et $(\|X\|_f(t))$ seront utilisées pour désigner ces processus.

Remarque 3.3.1. *À l'origine $\|X\|_f(x,O) = 1$; si S est dans un espace vectoriel et si la fonction f est une norme sur l'espace S, alors $\|X\|_f(x,t) = f(\overline{X}_f(x,t))$ pour $t > 0$, le processus $(\overline{X}_f(x,t))$ part d'un état qui est de norme 1. La renormalisation consiste à accélérer le temps d'un facteur $f(X(O))$ tout en renormalisant en espace par $1/f(X(O))$.*

Dans la suite on considère le cadre des processus de Markov à temps continu. Le cadre du temps discret est similaire : si (X_n) est une chaîne de Markov, le processus renormalisé est défini comme

$$(\|X\|_f(x,t)) = \frac{f(X_{\lfloor (x,f(x)t) \rfloor})}{f(x)}, \tag{3.8}$$

si $Xo = x \in S$ et $t \in \mathbb{R}^+$ (comme d'habitude, $\lfloor y \rfloor$ désigne la partie entière de $y \in \mathbb{R}$). Les résultats de cette section sont aussi vrais dans ce cadre.

3.4 Limites fluides

Définition 3.4.1. *[31] Une limite fluide associée au processus de Markov $(X(t))$ et à une fonction positive f est un point d'accumulation des lois de probabilité des processus*

$$\left\{ \left(\frac{f(X(x,f(x)t))}{f(x)} \right); x \in S \right\} = \{ (\|X\|_f(x,t)), x \in S \}$$

sur l'espace des fonctions càdlàg muni de la topologie de Skorokhod.

Si l'espace d'états S est inclus dans un espace vectoriel de dimension finie, un point d'accumulation des lois de probabilité de l'ensemble de processus

$$\left\{ \left(\frac{1}{f(x)} X(x,f(x)t) \right); x \in S \right\} = \left\{ \left(\overline{X}_f(x,t) \right); x \in S \right\}$$

est aussi appelé par abus de langage une limite fluide. Une limite fluide est donc une loi de probabilité $\mathbb{Q}$ d'un processus càdlàg. Quitte à agrandir l'espace de probabilité initial, il est possible de représenter une limite fluide comme un processus $(W(t))$ de loi $\mathbb{Q}$ défini sur l'espace de probabilité de base.

Les limites fluides donnent une expression asymptotique des trajectoires du processus de Markov, et donc le comportement qualitatif de celui-ci pour des grandes valeurs initiales.

3.5 Exercices

Exercice 01 :

Dans une file d'attente $M/M/1$ les arrivées des clients sont poissoniennes de taux λ, μ le taux de service et $L(t)$ le nombre de clients de la file à l'instant $t > o$.

Par la méthode des limites fluides calculer $L(t)$.

Corrigé

$f(z) = z$ si $Z \in \mathbb{N}$, le processus renormalisé est défini par

$$\overline{L}(N,t) = \frac{1}{N} L(Nt),$$

si $L(O) = N \in \mathbb{N}$, alors

$$\overline{L}(N,t) \to \left((1 + (\lambda - \mu)t)^{+}\right)$$

sur l'espace $D(\mathbb{R}_{+}, \mathbb{R})$ muni de la convergence uniforme sur les compacts. La convergence a donc lieu aussi pour la topologie de Skorokhod. La fonction $((1 + (\lambda - \mu)t)^{+})$ est donc l'unique limite fluide de ce processus de Markov.

Exercice 02 :

Deux files d'attente en tandem. Le processus d'arrivée de clients est de Poisson de paramètre λ_1 les clients se font servir dans la première file d'attente au taux μ_1 puis passent dans la deuxième file qui délivre le taux μ_2. Pour $i = 1, 2$ la quantité $L_i(t)$ désigne le nombre de clients dans la file i à l'instant $t > o$.

Calculer $L_i(t), \quad i = 1, 2$.

Corrigé

Le processus de Markov $(L(t)) = (l_1(t), l_2(t))$ est à valeurs dans $\mathbb{N}^2$. On suppose que $\lambda_1 < \mu_2 \leq \mu_1$ ce qui assure en particulier l'ergodicité du processus $(L(t))$. On se donne une suite d'états initiaux (x_N) telle que, pour $N \in \mathbb{N}$,

- $(x_N) = (l_{N,1}(0), l_{N,2}(0))$,
- $l_{N,1}(0) + l_{N,2}(0) = N$
- la suite $(l_{N,1}(0)/N)$ converge vers $y \in [0, 1]$.

Si $f(y) = \|y\| = |Y_1| + |Y_2|$, pour $y = (Y_1, Y_2) \in \mathbb{R}^2$, le processus renormalisé associé à f et partant de xN est à valeurs dans $\mathbb{R}_+^2$, il est défini par

$$\overline{L}(x_N, t) = \left(\frac{l_{N,1}(Nt)}{N}, \frac{l_{N,2}(Nt)}{N} \right)$$

La premiàre composante de $\overline{L}(x_N, t)$ correspond au processus renormalisé d'une file $M/M/1$, par conséquent on a la convergence

$$\left(\frac{l_{N,1}(Nt)}{N} \right) \to ((y + (\lambda_1 - \mu_1)t)^+)$$

quand N tend vers l'infini (pour la topologie de Skorokhod). De plus, si τ_N est le premier instant où la file 1 se vide, presque sûrement la variable τ_N/N tend vers $t_1 = y/(\mu_1 - \lambda_1)$ quand N tend vers l'infini. Tant que la file 1 est non vide, la deuxième file d'attente reçoit un flot poissonnien de paramètre μ_1, on en déduit la convergence.

$$\left(\frac{l_{N,2}(Nt)}{N} \right) \to ((1 - y + (\mu_1 - \mu_2)t))$$

pour $t < t_1$.

Après l'instant τ_N, partant de 0, la file 1 atteint l'équilibre rapidement, son processus de sortie est donc asymptotiquement de Poisson. L'échelle de temps rapide du processus renormalisé permet donc de considérer la file 2 comme une file $M/M/l$ recevant un flux poissonnien de requêtes de paramètre λ_1 et partant avec $l_{N,2}(\tau_N)$ clients initiaux, avec presque sûrement

$$y_1 = \lim_{N\to+\infty} \frac{l_{N,2}(\tau_N)}{N} = 1 - y + \frac{\mu_1 - \mu_2}{\mu_1 - \lambda_1} y = 1 + (\lambda_1 - \mu_2) t_1$$

Par conséquent, pour $t > t_1$,

$$\begin{aligned} \forall i, j \in \{1, 2, \cdots, r\}, \left(\frac{l_{N,2}(Nt)}{N}\right) &\to ((y_1 + (\lambda_1 - \mu_2)(t - t_1)))^+ \\ &= (1 + (\lambda_1 - \mu_2)t)^+ \end{aligned}$$

Pour résumer,

$$\overline{L}(N,t) \to \begin{cases} (y + (\lambda_1 - \mu_2)t, 1 - y(\mu_1 - \mu_2)t) & \text{pour} \quad t \le t_1 \\ (0, 1 + (\lambda_1 - \mu_2)t)^+ & \text{pour} \quad t \ge t_1, \end{cases}$$

avec $t_1 = y/(\mu_1 - \lambda_1)$.

Il est facile de vérifier que les seules limites possibles de l'ensemble des processus renormalisés $\overline{L}(x,t); x \in \mathbb{N}^2$ sont les fonctions déterministes linéaires par morceaux définies par (3.5) avec y parcourant l'intervalle $[0, 1]$. Noter que pour la file $M/M/1$ il n'y a qu'une seule fonction déterministe possible.

Exercice 03 :

Dans un réseau de Jackson, qui est composé de N files d'attente, dans la file d'attente d'ordre i avec $1 < i < N$, le service est exponentiel de paramètre μ_i et l'arrivée des clients dans le réseau à la file i est un processus de Poisson de paramètre λ. Une fois le service du client par la file i est terminé, le client rejoint la file j avec probabilité P_{ij} (avec $P_{ii} = 0$ ou quitte définitivement le réseau avec la probabilité résiduelle.

Montrer qu'il existe une suite positive $(\overline{\lambda}_i; 1 \leq i \leq N)$ vérifiant les équations de trafic

$$\overline{\lambda}_i = \lambda_i + \sum_{j=1}^{N} \overline{\lambda}_j p_{ji}.$$

Corrigé

Pour $1 \leq i \leq N$, on posera $\rho_i = \overline{\lambda}_i / \mu_i$. On pose $\lambda = \sum_{j=1}^{N} \lambda_i$ et $\alpha = (\lambda_i/\lambda, 1 \leq i \leq N)$. Si α est la distribution initiale de la chaîne de Markov associée à la matrice R, comme cette chaîne est transiente sur $1, \dots, N$ et l'état 0 est absorbant, le nombre de visites N_i à l'état $i = 1, \dots, N$ est intégrable et

$$\mathbb{E}(N_i) = \mathbb{P}(Y_0 = i) + \sum_{j=1}^{N} \sum_{n=0}^{+\infty} \mathbb{P}(Y_n = j, Y_{n+1} = i)$$

La propriété de Markov donne l'identité

$$\lambda\mathbb{E}(N_i) = \lambda_i + \sum_{j=1}^{N} \sum_{n=0}^{+\infty} \lambda\mathbb{P}(Y_n = j) r_{ji} = \lambda_i + \sum_{j=1}^{N} \lambda\mathbb{E}(N_j)) r_{ji},$$

le vecteur $(\lambda\mathbb{E}(N_i), 1 \leq i \leq N)$ est donc la solution de (3.5).

Chapitre 4

Appendix

4.1 Rappels : loi de Poisson et loi exponentielle

4.1.1 La loi de Poisson

Définition 4.1.1.1. *La loi de Poisson de paramètre $\lambda > 0$ de la variable aléatoire X est donnée par :*

$$\mathbb{P}(X = k) = \frac{\lambda^k e^{-\lambda}}{k!} \quad \textit{pour} \quad k = 0, 1, 2, \cdots \tag{4.1}$$

cette probabilitée est notée p_k.

Proposition 4.1.1.1. *Soit X une variable aléatoire ayant la distribution de Poisson dans (4.1) son espérance (moyenne ou premier moment) et sa variance sont égales à λ.*

Démonstration :

$$\begin{aligned}
E[X] &= \sum_{k=0}^{+\infty} k p_k \\
&= \sum_{k=0}^{+\infty} k \frac{\lambda^k e^{-\lambda}}{k!} = e^{-\lambda} \sum_{k=0}^{+\infty} k \frac{\lambda^k}{k!} \\
&= e^{-\lambda} \sum_{k=1}^{+\infty} \frac{\lambda^k}{(k-1)!} = e^{-\lambda} \lambda \sum_{k=1}^{+\infty} \frac{\lambda^{k-1}}{(k-1)!} \\
&= e^{-\lambda} \lambda \sum_{j=0}^{+\infty} \frac{\lambda^j}{(j)!} e^{-\lambda} \lambda e^{\lambda} = \lambda \qquad \blacksquare
\end{aligned}$$

Pour trouver la variance, on détermine d'abord

$$\begin{aligned}
\mathbb{E}[X(X-1)] &= \sum_{k=2}^{+\infty} k(k-1) p_k \\
&= \lambda^2,
\end{aligned}$$

Alors

$$\begin{aligned}
\mathbb{E}\left[X^2\right] &= \mathbb{E}[X(X-1)] + \mathbb{E}[X] \\
&= \lambda^2 + \lambda
\end{aligned}$$

tandis que

$$\begin{aligned}
\sigma_X^2 &= Var[X] \\
&= \mathbb{E}\left[X^2\right] - (\mathbb{E}[X])^2 \\
&= \lambda^2 + \lambda - \lambda^2 = \lambda. \qquad \blacksquare
\end{aligned}$$

Proposition 4.1.1.2. *Soit X, Y deux variables aléatoires indépendantes, $X \sim \exp(\lambda)$, $Y \sim \exp(\mu)$. Alors $Z = (X+Y) \sim \exp(\lambda+\mu)$.*

Démonstration : On dit que la variable aléatoire réelle X suit une loi exponentielle de paramètre $\lambda > 0$, et on note $t \sim \mathcal{E}(\lambda)$ si elle admet pour densité la fonction f définie par :

$$f_X(x) = \begin{cases} \lambda \exp(-\lambda x) & \text{si} \quad x \geq 0 \\ 0 & \text{sinon} \end{cases}$$

la fonction de répartition de la variable aléatoire X est

$$F(x) = \mathbb{P}(X \leq x) = 1 - e^{-\lambda x} \quad (x \geq 0),$$

Proposition 4.1.1.3. *L'espérance et la variance de X sont données par $\mathbb{E}(X) = 1/\lambda$, et $Var(X) = 1/\lambda^2$, respectivement,*

Démonstration : Pour l'espérence

$$\begin{aligned} \mathbb{E}(X) &= \int_0^{+\infty} t f(t) dt, \\ &= \int_0^{+\infty} t \lambda e^{-\lambda t} dt, \end{aligned}$$

par une intégration par partie on trouve

$$\begin{aligned} \mathbb{E}(X) &= \left[-te^{-\lambda t}\right]_0^{+\infty} - \int_0^{+\infty} -e^{-\lambda t} dt, \qquad \text{car} \qquad \left[-te^{-\lambda t}\right]_0^{+\infty} = 0, \\ &= \frac{-1}{\lambda} \int_0^{+\infty} -\lambda e^{-\lambda t} dt, \\ &= \frac{-1}{\lambda} \left[e^{-\lambda t}\right]_0^{+\infty}, \\ &= \frac{1}{\lambda}. \qquad \blacksquare \end{aligned}$$

Pour la variance

$$V(X) = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2 .$$

Or

$$\begin{aligned}
\mathbb{E}(X^2) &= \int_0^{+\infty} t^2 f(t)dt, \\
&= \int_0^{+\infty} t^2 \lambda e^{-\lambda t} dt,
\end{aligned}$$

par une intégration par partie on trouve

$$\begin{aligned}
\mathbb{E}(X^2) &= \left[-t^2 e^{-\lambda t}\right]_0^{+\infty} - \int_0^{+\infty} -2te^{-\lambda t} dt, \\
&= \int_0^{+\infty} 2te^{-\lambda t} dt, \qquad \text{car} \qquad \left[-t^2 e^{-\lambda t}\right]_0^{+\infty} = 0, \\
&= \frac{2}{\lambda} \int_0^{+\infty} t\lambda e^{-\lambda t} dt, \\
&= \frac{2}{\lambda}\mathbb{E}(X) = \frac{2}{\lambda^2}.
\end{aligned}$$

D'où la varience

$$V(X) = \frac{2}{\lambda^2} - \left(\frac{1}{\lambda}\right)^2 = \frac{1}{\lambda^2} \qquad \blacksquare$$

4.1.2 Absence de mémoire

Définition 4.1.2.1. *Une variable aléatoire X est dite sans mémoire (ou sans usure) si*

$$\forall s, t \geq 0, P(X > t + s | X > t) = P(X > s).$$

Proposition 4.1.2.1. *Une variable aléatoire de loi exponentielle est sans mémoire.*

Démonstration

La condition d'absence de mémoire (4.1.2.1) se formule de la manière suivante

$$\frac{P(X > t + s, X > t)}{P(X > t)} = P(X > t)$$

soit

$$P(X > t + s) = P(X > t)P(X > s)$$

Cette condition est évidemment satisfaite par la loi exponentielle.

Proposition 4.1.2.2. *Une variable aléatoire sans mémoire suit la loi exponentielle.*

Démonstration Soit X une variable possédant la propriété (1.4.1). On note

$$G(x) = P(X > x).$$

Nous venons d'observer que la fonction G était solution de l'équation fonctionnelle

$$\forall x, y \in \mathbb{R}_+, g(x + y) = g(x)g(y).$$

Un résultat d'analyse très classique garantit que les solutions continues (à droite) de cette équation sont de la forme

$$\forall x \in \mathbb{R}_+, g(x) = e^{-\lambda x}.$$

Ainsi, nous devons avoir

$$\forall x \in \mathbb{R}_+, P(X \leq x) = 1 - e^{-\lambda x}.$$

4.1.3 Relation entre la distribution Exponentielle et la distribution de Poisson

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

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

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

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

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

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

$\ldots$

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

4.1.4 Propriété sans mémoire de la distribution exponentielle

Quand une variable aléatoire t est exponentielle, la densité de probabilité est $f(t) = \alpha e^{-\alpha t}$, et la distribution cumulée correspondante est $F(t) = 1 - e^{-\alpha t}$.

Pour un accroissement de temps h, la probabilité que t est supérieur à h devient $P(t > h) = e^{-\alpha h}$.

De plus pour $t = (t^{'} + h)$, la probabilité de t est plus grande que $(t^{'} + h)$ est $P(t > (t^{'} + h)) = e^{-\alpha(t^{'} + h)}$

La probabilité conditionnelle de $t > (t^{'} + h)$ sachant $t > t^{'}$ est

$P(t > t' + h|t > t') = e^{-\alpha(t'+h)}/e^{-\alpha t'} = e^{-\alpha h}$

Du fait que les deux probabilités sont les mêmes, la distribution exponentielle est appelée une distribution de probabilité sans mémoire.

Bibliographie

[1] Allen, A. O. Probability, Statistics, and Queueing Theory with Computer Science Applications. Second edition, Academic Press, New York (First edition : 1978). (1990).

[2] Andersson, W.J.,Continuous Time Markov Chains :An Applications Orieted Approach,Springer New York, (1991).

[3] Andrejev, V. D., and Kudryavtsev, B. M. Markov Queueing Systems. Moscow Institute of Control (in Russian). (1982).

[4] Anisimov, V. V., Zakusilo, O. K., and Donchenko, V. S. Elements of Queueing Theory and Asymptotic System Analysis. Vishcha Shkola, Kiev (in Russian). (1987).

[5] Arnaud Guyader, Processus markoviens de sauts, (2006/2007).

[6] Baccelli, F., and Bremaud, P. Mathematical Theory of Queues. Springer-Verlag, Berlin. (1990).

[7] Bharucha-Reid, A.T. Elements of the Theory of Markov Processes and their Applications.New York : McGraw-Hill, (1960).

[8] Bhat, U. N. A Study of the Queueing Systems $M/G/l$ and $GI/M/I$. Lecture Notes in Operations Research and athematical Economics 2, Springer - Verlag, New York. (1968).

[9] Borovkov, A. A. 1976. Stochastic Processes in Queueing Theory. Springer-Verlag,Berlin (Russian original : Nauka, Moscow). (1972).

[10] Bunday, B. D. Basic Queueing Theory. Edward Arnold, London. (1986).

[11] Cheprasov, V. P. Elements of Queueing Theory. Kazan Aviation Institute (in Russian). (1985).

[12] Claudie Chabriac. Processus stochastiques et modélisation. (2012-2013).

[13] Cooper, R. B. Introduction to Queueing Theory. Second edition, North-Holland Publishing Company, New York (First edition : Macmillan, New York, 1972). Republished by the Continuing Engineering Education Program, The George Washington University, Washington D.C., 1990.

[14] David A. Levin Yuval Peres Elizabeth L. Markov Chains and Mixing Times, Wilmer University of Oregon.

[15] Fedosejev, Yu. N. Methods of Queueing Systems Analysis. Moscow Institute of Engineering and Physics (in Russian). (1982).

[16] Frédéric Sur. Programmation dynamique, chaï¿$\frac{1}{2}$nes de Markov, files d'attente. école des Mines de Nancy (2013)-(2014).

[17] Gelenbe, E.,and Pujolle, G. Introduction to Queuing Networks. John Wiley and Sons, Chichester, England (French Original : Introduction aux réseaux de files d'attente, Edition Eyrolles ; Paris). (1987).

[18] Gross,D.and C.M,Harris,Fundamentals of Queueing Theory ,Willcy,New york, (1975 1984).

[19] Ivo Adan and Jacques Resing, Queueing Systems, Eindhoven University of Technology, MB Eindhoven, The Netherlands (March 26, 2015).

[20] Howard M. Taylor, Samuel Karlin, An Introduction to Stochastic Modeling, Third Edition, Academic Press United states of america, (1998).

[21] Dr. J*á*nos Sztrik. Basic Queuing theory. p 25.

[22] Jean-Christophe Breton, Processus stochastiques M2 Mathématiques Université de Rennes 1 Octobre, Décembre (2020).

[23] Jean Louis Poss, Probabilité et statistique version 2.1. p74. Mai (2003).

[24] Kashyap, B. R. K., and Chaudhry, M. L. An Introduction to Queueing Theory. Aarkay, Calcutta, India. (1988).

[25] Kendall, D. G. Stochastic processes occurring in the theory of queues and their analysis by the method of the imbedded Markov chain. Annals of Mathematical Statistics, 24(3),338 -354. (1953).

[26] Khintchine, A. Y Mathematical Methods in the Theory of Queueing. Second edition, Hafner Publishing Company, New York (First edition : Griffin,London, 1960 ; Russian original : 1955). (1969).

[27] Lak*á*t*ó*s. L*á*szl*ó* Szeidl. Mikl*ó*s Telek. Introduction to Queueing Systems with Telecommunication Applications. Springer, p 88. (2010).

[28] Neuts, M. F. Structured Stochastic Matrices of $M/G/I$ Type and Their Applications. Marcel Dekker, Inc., New York. (1989).

[29] Newell, G. F. Applications of Queueing Theory. Second edition, Chapman and Hall, London (First edition : 1971). (1982).

[30] Norris, J.R. Markov Chains. Cambridge : Cambridge University Press, (1997).

[31] Philippe Robert, Réseaux et files d'attente : méthodes probabilistes, (2000).

[32] Saaty, T. L. Elements of queueing theory with applications. New York : McGraw Hill. (1961).

[33] Samuel Karlin. Howard, M. Taylor. A first course in stochastic processes, second edition. Academic press New york San fransisco. (1975).

[34] Sébastien Loustau, Chaînes de Markov et Processus markoviens de sauts. Applications aux files d'attente. Ecole Centrale de Marseille, Année (2008-2009).

[35] van Doom, E. Stochastic Monotonicity and Queueing Applications of Birth-Death Processes. Lecture Notes in Statistics 4, Springer-Verlag, New York, (1981).

[36] Walrand, J. Communication Networks : A First Course. The Aksen Associates Series in Electrical and Computer Engineering. Richard D. Irwin, Inc., and Aksen Associates, Inc., Homewood, IL and Boston, MA, (1991).

[37] Y. Velenik, Probabilités et Statistique, université de Genève, pp 195-202 Version du 24 mai (2012).

[38] Yves Caumel. Probabilité et Processus Stochastiques. Springer-Verlag FRance, (2011).

[39] Zakhar Kabluchko, Stochastic Processes (Stochastik II), University of Ulm Institute of Stochastics, p 23 (2013-2014).