

Université des Sciences et Technologies de Lille
U.F.R. de Mathématiques Pures et Appliquées

Files d'attente et fiabilité

**Michel Roussignol
Daniel Flipo**



D.E.S.S. Ingénierie Statistique et Numérique

Table des matières

Introduction	iii
1 Présentation générale des files d'attente	1
1.1 File d'attente	1
1.1.1 Description du modèle	1
1.1.2 Comportement de la file d'attente	3
1.2 Réseau de files d'attente	4
1.2.1 Description du modèle	4
1.2.2 Comportement du réseau	5
1.3 Simulation	6
2 Processus markoviens de sauts	9
2.1 Introduction	9
2.2 Processus de Poisson	10
2.3 Définition d'un processus markovien de sauts	18
2.3.1 Définitions	19
2.3.2 Exemples	19
2.3.3 Propriété de Markov	21
2.4 Matrice génératrice	22
2.4.1 Définitions	22
2.4.2 Exemples	24
2.4.3 Équations de Kolmogorov	25
2.5 Propriétés asymptotiques	26
2.5.1 Mesures invariantes, réversibles	26
2.5.2 Exemples	27
2.5.3 Théorème ergodique	27

3 Files d'attente	29
3.1 Introduction	29
3.2 Files M/M/.	29
3.2.1 Cas M/M/1	30
3.2.2 Cas M/M/ ∞	32
3.2.3 Cas M/M/ s	32
3.3 Files M/G/1	37
4 Réseaux de files d'attente	41
4.1 Exemples simples	41
4.1.1 Un système série-parallèle	41
4.1.2 Un système avec boucle	43
4.1.3 Un système fermé	44
4.2 Réseaux de Jackson	45
5 Fiabilité de systèmes simples	51
5.1 Définitions générales	51
5.2 Test de Laplace	54
5.3 Exemples de lois utiles en fiabilité	55
5.3.1 Loi exponentielle	55
5.3.2 Loi $\Gamma(\alpha, \beta)$	55
5.3.3 Loi de Weibull	55
5.3.4 Loi lognormale	56
5.4 Systèmes à un élément	58
5.5 Systèmes simples	60
5.5.1 Éléments en série	61
5.5.2 Éléments en parallèle	61
6 Fiabilité de systèmes complexes	65
6.1 Représentation de la logique d'un système	65
6.1.1 Diagramme de fiabilité	65
6.1.2 Arbre de défaillance	66
6.1.3 Coupes minimales	67

6.2	Systèmes markoviens	69
6.3	Systèmes non markoviens	71
6.3.1	Méthode des états fictifs	71
6.3.2	Méthode des variables complémentaires	72
	Transformation de Laplace	75
	Bibliographie	77

Introduction

Ce texte correspond à un enseignement de 32 heures dans le cadre du D.E.S.S. d'Ingénierie Statistique et Numérique de l'Université des Sciences et Technologies de Lille. Il s'adresse à des étudiants titulaires d'une maîtrise de mathématiques appliquées ayant reçu un cours de base en probabilités.

Le but du cours est de montrer comment on étudie des réseaux de files d'attente ou la fiabilité de certains systèmes à l'aide de techniques markoviennes. On présente d'abord les outils fondamentaux : chaînes de Markov, processus markoviens de sauts. À l'aide de ces outils on étudie des files d'attente et des réseaux de files d'attente dont les lois de service sont exponentielles et dont les flots d'arrivée sont poissonniens. Dans ce cadre on peut donner explicitement les conditions de stabilité et décrire les états stationnaires. Une légère excursion hors de ce cadre est effectuée dans le cas où, soit le service est exponentiel, soit le flot d'arrivée est poissonnien, car on peut encore utiliser une chaîne de Markov pour mener l'étude.

On présente ensuite les notions de base de la fiabilité et quelques outils d'analyse des systèmes complexes. Mais là aussi, l'essentiel du cours concerne les systèmes décrits par des processus de Markov. Il s'agit des systèmes réparables dont les taux de défaillance et de réparation sont constants. Dans ce cas on montre comment calculer la disponibilité et la fiabilité du système ainsi que les grandeurs moyennes qui leur sont attachées. Un paragraphe évoque des méthodes qui permettent de ramener dans le cadre markovien des systèmes dont le taux de réparation n'est pas constant.

Ce cours a été écrit en WORD par Michel ROUSSIGNOL qui a eu la gentillesse de me laisser ses originaux quand je lui ai succédé. Je l'en remercie bien vivement. Je me suis contenté de convertir son texte en L^AT_EX 2_E et d'y faire quelques modifications et compléments.

Daniel FLIPO



Chapitre 1

Présentation générale des files d'attente

1.1 File d'attente

La salle de réservation dans une grande gare SNCF donne une bonne représentation d'une file d'attente. Elle comprend un certain nombre de guichets et des clients qui sont, soit en train d'être servis, soit en attente. La salle de réservation forme le système. Elle peut être de capacité totale finie ou infinie. Les clients arrivent de manière aléatoire selon un flot régulier. Les temps de service sont également aléatoires. On va chercher à savoir si la longueur de la file d'attente a un comportement stationnaire et dans ce cas à calculer sa loi. Cela permet d'optimiser le nombre de guichets nécessaires pour satisfaire les clients.

On rencontre des files d'attente dans de nombreux cas. On vient de voir une file d'attente dans un organisme fournissant un service à un certain public. On rencontre une file d'attente à l'entrée d'une unité d'un système informatique lorsque les travaux qui arrivent se mettent en attente avant d'être traités par cette unité. Il en est de même à l'entrée d'un poste de travail dans une usine pour les pièces qui doivent être traitées à ce poste. On mesure sur ces exemples l'intérêt de la connaissance du mécanisme de formation d'une file d'attente pour optimiser un système informatique ou un atelier de production.

1.1.1 Description du modèle

Une file d'attente est décrite par plusieurs éléments : la loi des temps d'arrivée des clients, la loi des temps de service, le nombre de serveurs, la longueur maxi-

mum de la file d'attente, l'ordre dans lequel les clients sont servis. Reprenons chacun de ces éléments.

Les *instants d'arrivée des clients* sont en général aléatoires. Pour pouvoir calculer des grandeurs caractéristiques de la file d'attente, on doit connaître les lois probabilistes de ces temps d'arrivée, ou tout au moins faire certaines hypothèses sur ces lois qui seront vérifiées en pratique et qui rendront les calculs possibles. La première hypothèse faite est qu'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 variables aléatoires qui ont même loi. Cette hypothèse est vérifiée lorsque l'on étudie une file d'attente durant une période où les conditions qui amènent les clients sur la file sont semblables. Une autre hypothèse qui sera faite dans ce cours est l'indépendance probabiliste des temps d'interarrivée des clients; cette hypothèse simplifie notablement les calculs probabilistes mais elle est moins claire à justifier en pratique; on peut vérifier sa validité en faisant des tests statistiques. Enfin on supposera connue la loi des temps d'interarrivée. Le cas le plus courant est celui où cette loi est une loi exponentielle. Dans ce cas le modèle des temps d'arrivée est appelé un processus de Poisson. Cette hypothèse a deux mérites. D'abord elle est souvent vérifiée en pratique. Ensuite elle se prête bien au calcul probabiliste. Évidemment d'autres cas peuvent se présenter : temps d'interarrivée constants, de loi uniforme, de loi lognormale, de loi gamma...

On suppose que les *durées de service* sont des variables aléatoires indépendantes de même loi, indépendantes du processus des arrivées. Ces hypothèses sont souvent vérifiées en pratique. Il est courant de supposer que cette loi est exponentielle. Cette hypothèse simplifie les calculs, mais elle n'est pas toujours vérifiée. On peut être amené à considérer des temps de services constants, de loi uniforme, de loi gamma, de loi lognormale...

Le *nombre de serveurs* est évidemment un paramètre important du modèle.

La *longueur maximum de la file d'attente* est également un paramètre du modèle. On peut supposer que la file d'attente peut être aussi longue que l'on veut. C'est raisonnable dans certains cas. Cependant dans d'autres cas la longueur est limitée et lorsque la longueur de la file est maximum, un client qui arrive ne peut pas se mettre en file d'attente et repart. Il faut tenir compte de ce phénomène dans les calculs.

Le plus souvent les clients sont servis dans leur ordre d'arrivée. C'est la *discipline de service FIFO* (First In First Out). Mais d'autres disciplines peuvent être utilisées. Par exemple on peut servir en priorité certains types de clients, ou les clients demandant un service de courte durée.

Pour résumer les caractéristiques d'une file d'attente, on utilise classiquement

les *notations de Kendall* :

loi d'interarrivée / loi de service / nombre de serveurs / longueur max. de la file

Les lois d'interarrivées et les lois de services sont notées symboliquement : M pour une loi exponentielle (M pour Markov on verra pourquoi plus tard), D pour une loi déterministe (variable aléatoire constante), U pour une loi uniforme, G pour une loi quelconque (générale). Par exemple une file M/M/s/ ∞ signifie que le flot d'arrivées des clients est poissonnien, la loi de service est exponentielle, il y a s serveurs et la capacité de la salle d'attente est illimitée. Lorsqu'on ne spécifie pas le dernier paramètre celui-ci est infini. Sauf avis contraire la discipline de service est FIFO.

Deux grandeurs interviennent pour caractériser le modèle : le *nombre moyen d'arrivées par unité de temps* et la *durée moyenne d'un service*. Ces nombres se déduisent des lois d'interarrivée et de service. Par exemple si le processus d'arrivée est poissonnien, le temps moyen d'interarrivée vaut l'inverse du paramètre de la loi exponentielle et le nombre moyen d'arrivées par unité de temps vaut ce paramètre. Si la loi du temps de service est exponentielle, la durée moyenne de service vaut l'inverse du paramètre de la loi exponentielle.

1.1.2 Comportement de la file d'attente

On cherche à étudier l'évolution dans le temps de la longueur de la file d'attente. Deux cas peuvent se produire lorsque la longueur de la file n'est pas limitée. Dans le premier cas la file peut grandir de plus en plus et sa longueur tendre vers l'infini. Dans le deuxième cas la file se stabilise ; il y a des fluctuations mais un *état stationnaire* se met en place. La premier problème à résoudre est de prévoir selon les paramètres de modèle dans quel cas on se trouve. Intuitivement on peut donner une réponse. Si λ est le nombre moyen d'arrivées par unité de temps, si μ est le nombre moyen de clients que chaque serveur peut servir par unité de temps et si s est le nombre de serveurs, la file d'attente se stabilise s'il n'arrive pas trop de clients pour saturer l'offre de service, c'est-à-dire si λ est plus petit que $s\mu$. Il faut donner un sens précis à ces concepts et démontrer ce résultat. Lorsque la longueur de la file est limitée, un régime stationnaire s'établit toujours.

Dans le cas où un régime stationnaire s'établit, il est intéressant de connaître la loi de probabilité ou tout au moins la valeur moyenne d'un certain nombre de grandeurs caractéristiques. On peut citer la *longueur de la file d'attente*, le *temps passé par un client en attente*, le *nombre de serveurs occupés*. Dans le cas d'une file d'attente M/M/s, nous calculerons analytiquement ces lois de probabilités et les valeurs moyennes associées. Dans d'autres cas, nous pourrons seulement

calculer les valeurs moyennes. Dans tous les cas il est possible de simuler le fonctionnement de la file d'attente pour obtenir des informations sur ces lois de probabilités par des techniques statistiques.

1.2 Réseau de files d'attente

Le fonctionnement d'un atelier de production peut être modélisé par plusieurs files d'attente en interaction. Chaque poste de travail constitue une entité où un service est rendu, les éléments nécessaires pour ce service étant en attente, les produits réalisés partant vers d'autres postes de travail, puis vers l'extérieur à la fin de la chaîne de traitement. Ceci constitue un système de files d'attente en interaction, un réseau de files d'attente. Un système informatique est aussi constitué de plusieurs éléments qui traitent plusieurs programmes à la fois, chaque programme attendant avant d'être traité par un élément que son tour arrive, puis partant vers un autre élément et ainsi de suite jusqu'à la fin du traitement.

1.2.1 Description du modèle

La description d'un réseau de files d'attente comprend la *description de chaque file d'attente* du réseau (on dira souvent chaque station du réseau) et la *description de la circulation des clients* entre ces stations et entre l'extérieur et le réseau.

La description de chaque station est celle vue au paragraphe précédent. Chaque client qui entre dans une station se met en attente et reçoit un service selon les règles de la station.

Le modèle le plus simple de circulation des clients entre les stations est le suivant. Chaque station est éventuellement alimentée de l'extérieur par un flot régulier d'arrivée de clients et reçoit de la sortie d'un certain nombre de stations un flot de clients. À la sortie de chaque station, chaque client se dirige soit vers l'extérieur, soit vers une autre station du réseau selon un tirage au sort de probabilité donnée.

Un tel réseau de files d'attente est donc décrit par les *caractéristiques internes de chaque station* (nombre de serveurs, loi de service, longueur maximum de la file, discipline de service), par la *loi des temps d'arrivée de l'extérieur* en chaque station et par les *probabilités de routage* à la sortie de chaque station vers l'extérieur ou vers chacune des autres stations.

Le *modèle d'arrivée* le plus courant est le processus de Poisson. On suppose que pour chaque flot d'arrivée de clients de l'extérieur vers une station, les temps d'interarrivée des clients forment une suite de variables aléatoires indépendantes

de même loi exponentielle. Ce modèle est souvent observé en pratique et est commode pour faire des calculs probabilistes. Mais on peut observer d'autres modèles d'arrivée de clients avec des temps d'interarrivée constants ou suivant d'autres lois de probabilité. La paramètre le plus important est le nombre moyen d'arrivées par unité de temps en chaque station.

Le *routage à la sortie* de chaque station est caractérisé par la probabilité de sortie vers l'extérieur et les probabilités de transit vers les stations du réseau. Certaines de ces probabilités peuvent être nulles. Ce sont les probabilités non nulles qui déterminent les transits pouvant effectivement se réaliser.

On distingue deux types de réseaux : les *réseaux fermés* et les *réseaux ouverts*. Un réseau fermé n'a aucun lien avec l'extérieur. Un réseau ouvert a au moins un flux d'entrée et un flux de sortie.

1.2.2 Comportement du réseau

Comme pour une seule file d'attente, la première étape de l'étude d'un réseau consiste à savoir si le réseau va saturer, c'est-à-dire si la longueur d'au moins une des files d'attente va tendre vers l'infini. En général il est assez facile de *calculer les flux moyens* qui s'établissent entre les stations si un régime stationnaire se met en place. En effet il suffit d'écrire que les flots moyens d'entrée et de sortie en chaque station sont égaux pour obtenir un système linéaire qui permet de calculer le flux moyen de clients sur chaque lien du réseau en régime stationnaire. En comparant pour chaque station le flux moyen d'entrée et l'offre moyenne de service, on peut déterminer s'il y a *saturation* ou non.

Lorsqu'il n'y a pas saturation s'établit un *régime stationnaire*. Il est alors intéressant de connaître les lois de probabilité ou tout au moins les valeurs moyennes des grandeurs caractéristiques en chaque station : *longueur de la file d'attente*, *temps passé par un client en attente*, *nombre de serveurs occupés*. Dans le cas où toutes les files d'attente sont M/M/s et où les routages sont suffisamment simples¹, on sait calculer ces lois de probabilité. Dans les autres cas, on a la plupart du temps recours aux techniques de simulation.

Cette étude permet de guider des choix pour *optimiser le comportement du réseau*. Bien sûr en pratique les réseaux étudiés sont souvent plus complexes que ceux évoqués ci-dessus. Il peut y avoir plusieurs classes de clients. Diverses synchronisations entre les services peuvent exister. Les études analytiques donnent alors des résultats insuffisants et on utilise les *techniques de simulation*.

1. Ce point sera précisé au chapitre 4 traitant des réseaux de Jackson.

1.3 Simulation

Lorsque le modèle étudié est trop compliqué pour faire des calculs analytiques, on a recours aux techniques de simulation. On simule le modèle sur ordinateur et on traite les résultats observés par des techniques statistiques.

Pour illustrer ceci on va donner un exemple simple de simulation. Un atelier se compose de quatre postes dont trois machines d'usinage (Mo, M1, M2) et un poste de contrôle manuel (C). Il est organisé comme indiqué ci-dessous.

Les quatre postes sont précédés de files d'attente (stocks tampons) dans lesquels s'accumulent les pièces en attente de traitement. L'atelier produit deux types de pièces P1 et P2 dont les caractéristiques sont les suivantes :

- P1 arrive en Mo sur lequel elle est usinée en 4 mn, puis passe en M1 sur lequel elle est usinée en 7 mn; ensuite elle passe au poste de contrôle manuel C; le temps de contrôle est variable et peut être considéré comme uniformément distribué entre 3 et 6 mn; enfin P1 sort de l'atelier.
- P2 arrive en Mo sur lequel elle est usinée en 6 mn; elle passe en M2 sur lequel elle est usinée en 15 mn; puis elle passe au contrôle C; le temps de contrôle est du même type que pour P1.

Les pièces arrivent dans l'atelier de manière aléatoire. On a pu estimer qu'elles arrivent selon un processus de Poisson, le temps moyen d'interarrivée étant de 5 mn. Il y a 70% de pièces P1 et 30% de pièces P2 et on suppose que lorsqu'une pièce arrive, elle est de type P1 avec probabilité 0.7 et du type P2 avec probabilité 0.3.

Au poste de contrôle, 10% des pièces sont envoyées au rebut. Partout les pièces sont traitées par ordre d'arrivée.

Le problème est d'évaluer les performances de ce système de production. On peut déjà faire une analyse des flux moyens à travers ce système si un état stationnaire s'établit. Il arrive en moyenne 12 pièces par heure, dont 8.4 pièces P1 et 3.6 pièces P2. Dans la station Mo la demande moyenne de service pendant une heure vaut : $8.4 \times 4 + 3.6 \times 6 = 55.2$ mn, soit un taux d'utilisation de $55.2/60 = 0.92$. En M1 la demande moyenne de service par heure vaut $8.4 \times 7 = 58.8$ mn, soit un taux d'utilisation de $58.8/60 = 0.98$. En M2 la demande moyenne de service par heure vaut $3.6 \times 15 = 54$ mn, soit un taux d'utilisation de $54/60 = 0.90$. En C la demande moyenne de service par heure vaut $12 \times 4.5 = 54$ mn, soit un taux d'utilisation de $54/60 = 0.90$. Aucun poste n'est saturé en moyenne. Un régime stationnaire doit s'établir. On veut connaître pour ce régime les tailles moyennes des stocks intermédiaires et les temps de séjour moyens des pièces dans le système.

Plusieurs logiciels existent sur le marché pour effectuer ce genre de simulation, par exemple QNAP distribué par la société SIMULOG, filiale de l'INRIA, ou WITNESS distribué par le groupe LANNER.

Malheureusement, les tarifs proposés par les éditeurs aux établissements d'enseignement nous ont fait renoncer à les présenter dans le cadre de ce cours.

Les résultats de la simulation sous QNAP sont donnés dans le tableau suivant où les quantités suivantes interviennent :

SERVICE temps moyen de service dans la station pour un type de client,
 BUSY PCT taux d'utilisation de la station,
 CUST NB nombre moyen de clients dans la station,
 RESPONSE temps moyen passé dans la station pour un type de client,
 SERV NB nombre total de clients ayant transité dans la station.

TIME = 10000.00					
NAME	SERVICE	BUSY PCT	CUST NB	RESPONSE	SERV NB
C (P1) (P2)	4.501	0.8730	1.197	6.173	1939
	4.499	0.6049	0.8289	6.165	1344
	4.505	0.2681	0.3685	6.191	595
Mo (P1) (P2)	4.611	0.9029	5.555	28.35	1958
	4.000	0.5441	3.838	28.20	1360
	6.000	0.3588	1.717	28.71	598
M1 (P1)	7.000	0.9417	6.318	46.36	1345
	7.000	0.9417	6.318	46.36	1345
M2 (P2)	15.00	0.8941	3.412	57.22	596
	15.00	0.8941	3.412	57.22	596

Chapitre 2

Processus markoviens de sauts

2.1 Introduction

Dans ce chapitre nous allons étudier des processus de Markov qui permettent de modéliser et d'étudier des files d'attente. Nous commencerons par donner des propriétés du processus de Poisson qui modélise les arrivées de clients dans une file. Puis nous définirons les processus markoviens de sauts et étudierons certaines de leurs propriétés.

Un processus stochastique est une famille $(X_t)_{t \geq 0}$ de variables aléatoires indexées par le temps t . Les variables aléatoires X_t que nous étudierons dans ce cours prennent leurs valeurs dans une espace E dénombrable. Par exemple E est égal à \mathbb{N} lorsque X_t représente le nombre de clients présents dans une file d'attente au temps t . Dans un réseau de quatre stations, E est égal à \mathbb{N}^4 lorsque le processus X_t vaut $(X_1(t), X_2(t), X_3(t), X_4(t))$ où $X_i(t)$ représente le nombre de clients présents dans la station i , ($i = 1, 2, 3, 4$) au temps t .

Nous allons étudier des processus de Markov. Il s'agit de processus X_t tels que la loi du processus après un instant t dépend du passé du processus uniquement à travers la connaissance de X_t . Lorsque X_t représente le nombre de clients présents dans une file d'attente, la propriété de Markov est vérifiée si le flux d'entrée est Poissonnien et si les temps de service ont des lois exponentielles. L'intérêt des processus de Markov est que l'on peut obtenir une description complète de l'état stationnaire quand il existe.

2.2 Processus de Poisson

Nous allons étudier un modèle appelé processus de Poisson. Ce processus représente par exemple les instants d'arrivées successifs de clients dans une file d'attente.

On note T_1 l'instant d'arrivée du premier client, T_2 le deuxième, T_k le k -ème et ainsi de suite. La suite $(T_k)_{k \geq 1}$ est une suite croissante de variables aléatoires à valeurs dans \mathbf{R}_+ .

On note τ_i les durées des intervalles interarrivées : pour $i \geq 2$, $\tau_i = T_i - T_{i-1}$ et $\tau_1 = T_1$. La suite $(\tau_i)_{i \geq 1}$ est donc également une suite de variables aléatoires à valeurs dans \mathbf{R}_+ .

DÉFINITION 2.2.1 *Si les variables aléatoires $(\tau_i)_{i \geq 1}$ sont indépendantes et de même loi exponentielle de paramètre a ($a > 0$), on dit que la suite $(T_k = \sum_{i=1}^k \tau_i)_{k \geq 1}$ est la suite des instants de saut d'un processus de Poisson de paramètre a .*

Rappelons la densité de la loi exponentielle : $f(x) = ae^{-ax} \mathbf{1}_{x>0}$. L'espérance et la variance d'une variable exponentielle τ de paramètre a sont données par

$$\mathbf{E}(\tau) = \frac{1}{a}, \quad \mathbf{Var}(\tau) = \frac{1}{a^2}.$$

Les propriétés suivantes des variables exponentielles, absence de mémoire, minimum et somme de variables indépendantes, joueront un rôle crucial dans la suite.

PROPOSITION 2.2.2 (absence de mémoire) *Une variable aléatoire τ à densité est exponentielle si et seulement si*

$$\forall s > 0, \forall t > 0, \quad \mathbf{P}(\tau > t + s \mid \tau > t) = \mathbf{P}(\tau > s).$$

Cette propriété permet de poser $T_0 = 0$, c'est-à-dire de supposer qu'un client arrive à l'instant 0.

Démonstration. À faire en exercice. □

PROPOSITION 2.2.3 *Si τ_1 et τ_2 sont deux variables exponentielles indépendantes de paramètres a_1 et a_2 , alors la variable $\inf(\tau_1, \tau_2)$ suit une loi exponentielle de paramètre $a_1 + a_2$. De plus,*

$$\mathbf{P}(\tau_1 < \tau_2) = \frac{a_1}{a_1 + a_2}.$$

Démonstration. À faire en exercice. □

PROPOSITION 2.2.4 *La somme de n variables aléatoires indépendantes de loi exponentielle de paramètre a suit une loi (appelée loi Gamma de paramètres n et a) dont la densité est*

$$f_n(x) = a \frac{(ax)^{n-1}}{(n-1)!} e^{-ax} \mathbf{1}_{x \geq 0}.$$

En particulier, pour tout $n \geq 1$, la variable aléatoire T_n a la densité f_n ci-dessus.

Démonstration. $f_n(x)$ est la convoluée n -fois de la densité exponentielle $f_1(x)$. La démonstration est facile par récurrence sur n . \square

Pour un temps t fixé, on va s'intéresser à la variable aléatoire N_t égale au nombre de clients arrivés dans l'intervalle $]0, t]$:

$$N_t = \sum_{m \geq 1} \mathbf{1}_{T_m \leq t}.$$

N_t est une variable aléatoire discrète à valeurs dans \mathbf{N} . On l'appelle *fonction de comptage* du processus au temps t .

PROPOSITION 2.2.5 *La variable aléatoire N_t , fonction de comptage d'un processus de Poisson de paramètre a , suit une loi de Poisson de paramètre at .*

Démonstration. L'événement $\{N_t \geq n\}$ est égal à $\{T_n \leq t\}$ et on connaît la densité de T_n . On a donc :

$$\begin{aligned} \mathbf{P}(N_t \geq n) &= \mathbf{P}(T_n \leq t) \\ &= \int_0^t \frac{a^n x^{n-1} e^{-ax}}{(n-1)!} dx \end{aligned}$$

et en intégrant par parties en posant $du = \frac{x^{n-1}}{(n-1)!} dx$ et $v = a^n e^{-ax}$:

$$\begin{aligned} &= \frac{a^n t^n e^{-at}}{n!} + \mathbf{P}(T_{n+1} \leq t) \\ &= \frac{a^n t^n e^{-at}}{n!} + \mathbf{P}(N_t \geq n+1). \end{aligned}$$

Finalement

$$\mathbf{P}(N_t = n) = \mathbf{P}(N_t \geq n) - \mathbf{P}(N_t \geq n+1) = \frac{a^n t^n}{n!} e^{-at}. \quad \square$$

La loi des instants de saut T_1, T_2, \dots, T_k d'un processus de Poisson se déduit facilement de celle des $(\tau_i)_{1 \leq i \leq k}$:

PROPOSITION 2.2.6 *La loi du vecteur (T_1, T_2, \dots, T_k) des k premiers instants de saut d'un processus de Poisson a pour densité*

$$f(t_1, t_2, \dots, t_k) = a^k e^{-at_k} \mathbf{1}_{0 < t_1 < t_2 < \dots < t_k}.$$

Démonstration. La loi de (T_1, T_2, \dots, T_k) se déduit de celle de $(\tau_1, \tau_2, \dots, \tau_k)$ par le changement de variables $(u_1, u_2, \dots, u_k) \rightarrow (t_1, t_2, \dots, t_k)$ défini par

$$\left\{ \begin{array}{lcl} t_1 & = & u_1 \\ t_2 & = & u_1 + u_2 \\ \vdots & & \vdots \\ t_k & = & u_1 + u_2 + \dots + u_k \end{array} \right. \quad \left\{ \begin{array}{lcl} u_1 & = & t_1 \\ u_2 & = & t_2 - t_1 \\ \vdots & & \vdots \\ u_k & = & t_k - t_{k-1} \end{array} \right.$$

C'est un difféomorphisme de classe \mathcal{C}^1 de $D = \{0 < t_1 < t_2 < \dots < t_k\}$ dans $\Delta = (\mathbf{R}_+^*)^k$. Son jacobien vaut 1 (système linéaire triangulaire), le théorème de changement de variables donne pour h fonction borélienne positive

$$\begin{aligned} \mathbf{E}[h(T_1, T_2, \dots, T_k)] &= \mathbf{E}[h(\tau_1, \tau_1 + \tau_2, \dots, \tau_1 + \dots + \tau_k)] \\ &= \int_{(\mathbf{R}_+^*)^k} h(u_1, u_1 + u_2, \dots, u_1 + \dots + u_k) \prod_{i=1}^k a e^{-au_i} du_1 \dots du_k \\ &= \int_D h(t_1, t_2, \dots, t_k) e^{-at_k} a^k dt_1 \dots dt_k \\ &= \int_{(\mathbf{R}_+^*)^k} h(t_1, t_2, \dots, t_k) \mathbf{1}_{0 < t_1 < t_2 < \dots < t_k} e^{-at_k} a^k dt_1 \dots dt_k. \end{aligned} \quad \square$$

Les instants de saut d'un processus de Poisson vérifient la propriété intéressante suivante : lorsqu'on sait qu'exactement n sauts ont eu lieu dans l'intervalle $[0, t]$, la distribution de ces sauts dans $[0, t]$ est celle de n points tirés uniformément au hasard dans cet intervalle :

THÉORÈME 2.2.7 *Conditionnellement en $\{N_t = n\}$, la loi du vecteur (T_1, T_2, \dots, T_n) des n premiers instants de saut d'un processus de Poisson est la même que celle du réordonnement par ordre croissant d'une suite de variables aléatoires indépendantes de même loi uniforme sur $[0, t]$.*

Autrement dit : si on tire des variables aléatoires X_1, X_2, \dots, X_n indépendantes de loi uniforme sur $[0, t]$ et si on les réordonne en posant

$$Y_1 = \min_{1 \leq i \leq n} X_i < Y_2 < \dots < Y_n = \max_{1 \leq i \leq n} X_i,$$

alors la loi conditionnelle de (T_1, T_2, \dots, T_n) sachant $N_t = n$ est la même que celle de (Y_1, Y_2, \dots, Y_n) .

Démonstration. La densité de probabilité du vecteur aléatoire (T_1, T_2, \dots, T_k) étant donnée par la

proposition 2.2.6, on calcule pour h fonction borélienne positive

$$\begin{aligned} \mathbf{E}[h(T_1, T_2, \dots, T_n) \mathbf{1}_{N_t=n}] &= \mathbf{E}[h(T_1, T_2, \dots, T_n) \mathbf{1}_{T_n \leq t < T_{n+1}}] \\ &= \int_{(\mathbf{R}_+)^{n+1}} h(t_1, t_2, \dots, t_n) \mathbf{1}_{t_n \leq t < t_{n+1}} \\ &\quad e^{-at_{n+1}} a^{n+1} \mathbf{1}_{0 < t_1 < t_2 < \dots < t_{n+1}} dt_1 dt_2 \dots dt_{n+1}. \end{aligned}$$

En intégrant en t_{n+1} on obtient

$$\begin{aligned} \mathbf{E}[h(T_1, T_2, \dots, T_n) \mathbf{1}_{N_t=n}] &= \int_{(\mathbf{R}_+)^n} h(t_1, t_2, \dots, t_n) \mathbf{1}_{0 < t_1 < t_2 < \dots < t_n \leq t} dt_1 dt_2 \dots dt_n \\ &\quad \int_t^\infty e^{-at_{n+1}} a^{n+1} dt_{n+1} \\ &= \int_{(\mathbf{R}_+)^n} h(t_1, t_2, \dots, t_n) \mathbf{1}_{0 < t_1 < t_2 < \dots < t_n \leq t} \\ &\quad e^{-at} a^n dt_1 dt_2 \dots dt_n. \end{aligned}$$

On divise cette expression par $\mathbf{P}(N_t = n) = e^{-at} (at)^n / n!$, d'où l'espérance conditionnelle

$$\mathbf{E}[h(T_1, T_2, \dots, T_n) | N_t = n] = \int_{\mathbf{R}^n} h(t_1, t_2, \dots, t_n) \frac{n!}{t^n} \mathbf{1}_{0 < t_1 < t_2 < \dots < t_n < t} dt_1 \dots dt_n.$$

La densité du vecteur (T_1, T_2, \dots, T_n) conditionnellement en $N_t = n$ est donc

$$g(t_1, t_2, \dots, t_n) = \frac{n!}{t^n} \mathbf{1}_{0 < t_1 < t_2 < \dots < t_n < t}.$$

Il reste à vérifier que g est aussi la densité du réordonnement de n points tirés uniformément au hasard sur $[0, t]$. La densité du vecteur (X_1, X_2, \dots, X_n) est t^{-n} sur le pavé $[0, t]^n$, on subdivise celui-ci en $n!$ secteurs $(S_k)_{1 \leq k \leq n!}$ correspondant aux différentes permutations des (x_i) et on fait sur chaque secteur le changement de variable $(x_1, \dots, x_n) \mapsto (y_1, \dots, y_n)$ (réordonnement par ordre croissant), dont le jacobien vaut 1 :

$$\begin{aligned} \mathbf{E}[h(Y_1, \dots, Y_n)] &= \sum_{1 \leq k \leq n!} \int_{S_k} h(x_{\sigma(1)}, \dots, x_{\sigma(n)}) \frac{1}{t^n} dx_1 \dots dx_n \\ &= n! \int_{0 < y_1 < \dots < y_n} h(y_1, \dots, y_n) \frac{1}{t^n} dy_1 \dots dy_n, \end{aligned}$$

d'où le résultat. \square

Le résultat précédent est à la base du test de Laplace utilisé en fiabilité (voir chapitre 5). On l'utilise également pour calculer les probabilités $\mathbf{P}(N_u = k | N_t = n)$ avec $0 \leq k \leq n$ et $0 < u < t$. Les X_i étant tirées uniformément au hasard sur $[0, t]$, chacune d'elles a probabilité u/t de tomber dans $[0, u]$. La loi du nombre d'entre elles qui tombe dans $[0, u]$ est donc binomiale de taille n et de paramètre u/t .

$$\mathbf{P}(N_u = k | N_t = n) = C_n^k \left(\frac{u}{t}\right)^k \left(1 - \frac{u}{t}\right)^{n-k}.$$

Plus généralement, conditionnellement en $\{N_t = n\}$, la répartition des n instants de saut dans les k intervalles $]t_{i-1}, t_i]_{1 \leq i \leq k}$, où $t_0 = 0 < t_1 < \dots < t_k = t$, suit une loi multinomiale :

$$(2.1) \quad \mathbf{P}(N_{t_1} = n_1, N_{t_2} - N_{t_1} = n_2, \dots, N_{t_k} - N_{t_{k-1}} = n_k \mid N_t = n) \\ = \frac{n!}{n_1! n_2! \dots n_k!} \prod_{i=1}^k \left(\frac{t_i - t_{i-1}}{t} \right)^{n_i} \quad \text{avec} \quad n = n_1 + n_2 + \dots + n_k.$$

Nous allons maintenant nous intéresser à N_t en tant qu'application de \mathbf{R}_+ dans l'ensemble des variables aléatoires à valeurs dans \mathbf{N} . À chaque temps t on associe une variable aléatoire discrète à valeurs dans \mathbf{N} . On dit que $(N_t)_{t \geq 0}$ est une processus à temps continu dont l'espace d'états est \mathbf{N} . Pour chaque expérience ω , la fonction de \mathbf{R}_+ dans \mathbf{N} qui à t associe $N_t(\omega)$ est une fonction croissante, constante par morceaux, qui part de 0 et qui saute de $+1$ en chaque instant d'arrivée $T_n(\omega)$.

Voici une deuxième caractérisation du processus de Poisson faisant intervenir la fonction de comptage au lieu des intervalles interarrivées :

THÉORÈME 2.2.8 *La fonction de comptage $(N_t)_{t \in \mathbf{R}_+}$ d'un processus de Poisson de paramètre a est à accroissements indépendants :*

$\forall k \geq 2, \forall t_0 < t_1 < \dots < t_k$, les v.a. $(N_{t_i} - N_{t_{i-1}})_{1 \leq i \leq k}$ sont indépendantes

et pour tous $t > 0$ et $s > 0$ la variable $N_{t+s} - N_s$ suit une loi de Poisson de paramètre at .

Réiproquement toute fonction de comptage ayant ces deux propriétés définit un processus de Poisson de paramètre a : les intervalles interarrivées sont des variables indépendantes exponentielles de paramètre a .

Un processus de Poisson de paramètre a peut donc être défini

- soit à partir de ses instants de sauts, en disant que la suite des $\tau_i = T_i - T_{i-1}$ est une suite de variables aléatoires indépendantes de même loi exponentielle de paramètre a ,
- soit à partir de sa fonction de comptage $N(t)$, en disant que $N(t)$ est à accroissements indépendants et que pour tout couple (t, s) de réels ($t \geq 0, s > 0$) l'accroissement $N(t+s) - N(t)$ suit une loi de Poisson de paramètre as .

Démonstration. On veut montrer que $\forall k \geq 1$ pour toute suite croissante de réels $0 = t_0 < t_1 <$

$t_2 < \dots < t_k$ et pour toute suite (n_1, n_2, \dots, n_k) d'entiers ≥ 0 ,

$$\begin{aligned} \mathbf{P}(\mathbf{N}_{t_1} = n_1 \cap \mathbf{N}_{t_2} - \mathbf{N}_{t_1} = n_2 \cap \dots \cap \mathbf{N}_{t_k} - \mathbf{N}_{t_{k-1}} = n_k) \\ = \frac{(at_1)^{n_1}}{n_1!} e^{-at_1} \times \frac{[a(t_2 - t_1)]^{n_2}}{n_2!} e^{-a(t_2 - t_1)} \times \dots \times \frac{[a(t_k - t_{k-1})]^{n_k}}{n_k!} e^{-a(t_k - t_{k-1})}. \end{aligned}$$

Posons $n = n_1 + n_2 + \dots + n_k$ et conditionnons par $\{\mathbf{N}_{t_k} = n\}$ pour utiliser l'égalité (2.1) :

$$\begin{aligned} \mathbf{P}(\mathbf{N}_{t_1} = n_1 \cap \mathbf{N}_{t_2} - \mathbf{N}_{t_1} = n_2 \cap \dots \cap \mathbf{N}_{t_k} - \mathbf{N}_{t_{k-1}} = n_k) \\ = \mathbf{P}(\mathbf{N}_{t_1} = n_1 \cap \mathbf{N}_{t_2} - \mathbf{N}_{t_1} = n_2 \cap \dots \cap \mathbf{N}_{t_k} - \mathbf{N}_{t_{k-1}} = n_k \mid \mathbf{N}_{t_k} = n) \mathbf{P}(\mathbf{N}_{t_k} = n) \\ = \left(\frac{n!}{n_1! n_2! \dots n_k!} \prod_{i=1}^k \left(\frac{t_i - t_{i-1}}{t_k} \right)^{n_i} \right) \frac{(at_k)^n}{n!} e^{-at_k} \quad (\text{en posant } t_0 = 0) \\ = \left(\prod_{i=1}^k \frac{a^{n_i} (t_i - t_{i-1})^{n_i}}{n_i!} \right) e^{-at_k} = \prod_{i=1}^k \left(\frac{a^{n_i} (t_i - t_{i-1})^{n_i}}{n_i!} e^{-a(t_i - t_{i-1})} \right). \end{aligned}$$

Réiproque : on suppose la fonction de comptage \mathbf{N} à accroissements indépendants et on calcule pour $0 < t_1 < t_2 < \dots < t_k$, la fonction

$$\begin{aligned} G(t_1, t_2, \dots, t_k) &= \mathbf{P}(0 < T_1 \leq t_1 < T_2 \leq t_2 < \dots < T_{k-2} \leq T_{k-1} < T_{k-1} < T_k \leq t_k) \\ &= \mathbf{P}(\mathbf{N}_{t_1} - \mathbf{N}_0 = 1 \cap \mathbf{N}_{t_2} - \mathbf{N}_{t_1} = 1 \cap \dots \\ &\quad \cap \mathbf{N}_{t_{k-1}} - \mathbf{N}_{t_{k-2}} = 1 \cap \mathbf{N}_{t_k} - \mathbf{N}_{t_{k-1}} \geq 1) \\ &= at_1 e^{-at_1} \times a(t_2 - t_1) e^{-a(t_2 - t_1)} \times \dots \\ &\quad \times a(t_{k-1} - t_{k-2}) e^{-a(t_{k-1} - t_{k-2})} \times (1 - e^{-a(t_k - t_{k-1})}) \\ &= a^{k-1} t_1 (t_2 - t_1) \dots (t_{k-1} - t_{k-2}) (e^{-at_{k-1}} - e^{-at_k}). \end{aligned}$$

La suite des (T_j) étant p.s. strictement croissante, la loi du vecteur T_1, T_2, \dots, T_k est portée par le sous-ensemble $\{0 < t_1 < t_2 < \dots < t_k\}$ de \mathbf{R}_+^k . Sur cette partie, G est de classe \mathcal{C}^∞ en t_1, t_2, \dots, t_k et représente « essentiellement » la fonction de répartition F du vecteur T_1, T_2, \dots, T_k , précisons les choses dans le cas $k = 3$:

$$\begin{aligned} G(t_1, t_2, t_3) &= \mathbf{P}(0 < T_1 \leq t_1 < T_2 \leq t_2 < T_3 \leq t_3) \\ &= \mathbf{P}(T_1 \leq t_1 \cap T_2 \leq t_2 \cap T_3 \leq t_3 \cap (T_2 > t_1 \cap T_3 > t_2)) \\ &= \mathbf{P}(T_1 \leq t_1 \cap T_2 \leq t_2 \cap T_3 \leq t_3) \\ &\quad - \mathbf{P}((T_1 \leq t_1 \cap T_2 \leq t_2 \cap T_3 \leq t_3) \cap (T_2 \leq t_1 \cup T_3 \leq t_2)) \\ &= F(t_1, t_2, t_3) - \mathbf{P}((T_2 \leq t_1 \cap T_3 \leq t_3) \cup (T_1 \leq t_1 \cup T_3 \leq t_2)) \\ &= F(t_1, t_2, t_3) \\ &\quad - \mathbf{P}(T_2 \leq t_1 \cap T_3 \leq t_3) - \mathbf{P}(T_1 \leq t_1 \cap T_3 \leq t_2) + \mathbf{P}(T_2 \leq t_1 \cap T_3 \leq t_2) \end{aligned}$$

ce qui établit, dans le cas $k = 3$, que G et F ont donc même dérivée par rapport à t_3, t_2, t_1 . Il en est de même dans le cas général :

$$\frac{\partial^k F(t_1, t_2, \dots, t_k)}{\partial t_1 \partial t_2 \dots \partial t_k} = \frac{\partial^k G(t_1, t_2, \dots, t_k)}{\partial t_1 \partial t_2 \dots \partial t_k} = a^k e^{-at_k}.$$

La loi du vecteur T_1, T_2, \dots, T_k admet donc une densité sur \mathbf{R}_+^k de la forme

$$f(t_1, t_2, \dots, t_k) = a^k e^{-at_k} \mathbf{1}_{0 < t_1 < t_2 < \dots < t_k},$$

on reconnaît la loi trouvée à la proposition 2.2.6, ce qui établit la réiproque. \square

En reformulant le résultat de la proposition précédente on montre que pour toute suite strictement décroissante d'instants $s, t, t_1, t_2, \dots, t_k$ et toute suite décroissante d'entiers $j, i, i_1, i_2, \dots, i_k$:

$$\mathbf{P}(N_s = j \mid N_t = i, N_{t_1} = i_1, \dots, N_{t_k} = i_k) = \mathbf{P}(N_s = j \mid N_t = i) = \mathbf{P}(N_{s-t} = j - i)$$

Le processus à temps continu $(N_t)_{t \geq 0}$ vérifie une propriété de perte de mémoire identique à celle vue pour les chaînes de Markov. Il s'agit de la propriété de définition des processus de Markov homogènes en temps continu que nous retrouverons dans le paragraphe suivant. Ceci constitue une troisième caractérisation du processus de Poisson :

THÉORÈME 2.2.9 *Une fonction de comptage $(N_t)_{t \in \mathbf{R}_+}$ est celle d'un processus de Poisson si et seulement si elle vérifie*

(2.2)

$$\forall k \in \mathbf{N}^*, \forall s > 0, \forall t_1 < t_2 < \dots < t_k \text{ (réels } > 0\text{), } \forall j, \forall i_1 \leq i_2 \dots \leq i_k \leq i \text{ (entiers),}$$

$$\begin{aligned} \mathbf{P}(N_{t+s} = i + j \mid N_t = i, N_{t_k} = i_k, \dots, N_{t_1} = i_1) &= \mathbf{P}(N_{t+s} = i + j \mid N_t = i) \\ &= \frac{(as)^j}{j!} e^{-as}. \end{aligned}$$

Démonstration. Pour toute suite strictement décroissante d'instants t_1, t_2, \dots, t_k, t et toute suite croissante d'entiers i_1, i_2, \dots, i_k, i , calculons la probabilité :

$$\begin{aligned} &\mathbf{P}(N_{t+s} = i + j \mid N_t = i, N_{t_k} = i_k, \dots, N_{t_1} = i_1) \\ &= \mathbf{P}(N_{t+s} - N_t = j \mid N_t - N_{t_k} = i - i_k, \dots, N_{t_2} - N_{t_1} = i_2 - i_1, N_{t_1} - N_0 = i_1) \end{aligned}$$

N étant à accroissements indépendants :

$$= \mathbf{P}(N_{t+s} - N_t = j) = \mathbf{P}(N_{t+s} - N_t = j \mid N_t - N_0 = i) = \mathbf{P}(N_{t+s} = i + j \mid N_t = i).$$

Enfin, les accroissements suivant des lois de Poisson, on a

$$\mathbf{P}(N_{t+s} - N_t = j) = \frac{(as)^j}{j!} e^{-as}.$$

Réciiproquement, si N vérifie la propriété de Markov ci-dessus, en utilisant l'égalité $\mathbf{P}(A_n \cap A_{n-1} \cap \dots \cap A_1) = \mathbf{P}(A_1) \mathbf{P}(A_2 \mid A_1) \dots \mathbf{P}(A_n \mid A_{n-1} \cap \dots \cap A_1)$ on obtient :

$$\begin{aligned} &\mathbf{P}(N_{t_k} - N_{t_{k-1}} = n_k \cap \dots \cap N_{t_2} - N_{t_1} = n_2 \cap N_{t_1} = n_1) \\ &= \mathbf{P}(N_{t_1} = n_1) \times \dots \times \mathbf{P}(N_{t_2} - N_{t_1} = n_2 \mid N_{t_1} = n_1) \\ &\quad \times \mathbf{P}(N_{t_k} - N_{t_{k-1}} = n_k \mid N_{t_{k-1}} - N_{t_{k-2}} = n_{k-1} \cap \dots \cap N_{t_1} = n_1) \\ &= \mathbf{P}(N_{t_1} = n_1) \times \dots \times \mathbf{P}(N_{t_2} = n_1 + n_2 \mid N_{t_1} = n_1) \\ &\quad \times \mathbf{P}(N_{t_k} = n_k + \dots + n_1 \mid N_{t_{k-1}} = n_{k-1} + \dots + n_1 \cap \dots \cap N_{t_1} = n_1) \end{aligned}$$

soit, en utilisant (2.2) :

$$= \frac{[at_1]^{n_1}}{n_1!} e^{-at_1} \times \frac{[a(t_2 - t_1)]^{n_2}}{n_2!} e^{-a(t_2 - t_1)} \times \cdots \times \frac{[a(t_k - t_{k-1})]^{n_k}}{n_k!} e^{-a(t_k - t_{k-1})}$$

donc N est bien un processus de Poisson (cf. théorème 2.2.8). \square

Le processus de Poisson se retrouve dans de nombreuses applications. On a déjà parlé des flots d'arrivée de clients dans les files d'attente. Le processus de Poisson modélise également les temps de désintégration d'une particule radioactive. Ce modèle se retrouve dans les appareils de radiographie en médecine nucléaire. Les temps de passage de voitures sur une route à grande circulation non embouteillée constituent également un processus de Poisson.

Terminons en donnant deux résultats sur la superposition de processus de Poisson indépendants et sur l'opération inverse de décomposition.

THÉORÈME 2.2.10 *Soient (N_t^1) et (N_t^2) deux processus de Poisson indépendants (i.e. tout événement lié à l'un est indépendant de tout événement lié à l'autre), de paramètres respectifs a_1 et a_2 , on appelle superposition de ces deux processus le processus somme :*

$$\forall t \in \mathbf{R}_+, \quad N_t = N_t^1 + N_t^2.$$

Alors (N_t) est un processus de Poisson de paramètre $a = a_1 + a_2$.

Inversement, si (N_t) est un processus de Poisson de paramètre a , on associe aux instants de saut $(T_n)_{n \in \mathbf{N}}$ de ce processus une suite de variables de Bernoulli $(Y_n)_{n \in \mathbf{N}}$ indépendantes, indépendantes du processus (N_t) et de même loi :

$$\forall n \in \mathbf{N}, \quad \mathbf{P}(Y_n = 1) = p, \quad \mathbf{P}(Y_n = 0) = 1 - p.$$

Soit (N_t^1) (resp. (N_t^2)) le processus dont les instants de saut sont les (T_n) tels que $Y_n = 1$ (resp. $Y_n = 0$), alors (N_t^1) et (N_t^2) sont des processus de Poisson de paramètres $a_1 = ap$ et $a_2 = a(1 - p)$ respectivement.

Exemples d'application :

- sur une voie à double sens, si les flux de véhicules venant de droite et de gauche sont deux processus de Poisson de paramètres a_1 et a_2 respectivement, le flux global de véhicules est un processus de Poisson de paramètre $a_1 + a_2$;
- si le flux total de voitures est un processus de Poisson de paramètre a et que 5% des voitures sont rouges, le flux de voitures rouges est aussi un processus de Poisson (de paramètre $0,05a$).

Démonstration. 1) Superposition : il est clair que le processus somme (N_t) est à accroissements indépendants. D'autre part la somme de deux variables de Poisson indépendantes, de paramètres λ et μ est une variable de Poisson de paramètre $\lambda + \mu$ (à vérifier directement ou en utilisant les fonctions génératrices), donc pour tous $t, s > 0$, $N_{t+s}^3 - N_t^3$ somme de deux variables de Poisson de paramètres $a_1 s$ et $a_2 s$, suit une loi de Poisson de paramètre $(a_1 + a_2)s$ et d'après le théorème 2.2.8 (N_t^3) est un processus de Poisson de paramètre $a_1 + a_2$.

2) Décomposition : il est facile de voir que la suite des intervalles interarrivées du processus (N_t^1) (ou (N_t^2)) est une suite de variables indépendantes. Il reste à vérifier que ces intervalles interarrivées suivent des lois exponentielles de même paramètre. Supposons que $Y_n = 1$ et plaçons nous à l'instant T_n , notons τ l'intervalle interarrivée suivant pour le processus (N_t^1), sa loi est donnée par :

$$\begin{aligned}
 \mathbf{P}(\tau > t) &= \sum_{k=1}^{\infty} \mathbf{P}(\tau > t \cap Y_{n+1} = 0 \cdots \cap Y_{n+k-1} = 0 \cap Y_{n+k} = 1) \\
 &= \sum_{k=1}^{\infty} \mathbf{P}(\tau_1 + \cdots + \tau_k > t \cap Y_{n+1} = 0 \cdots \cap Y_{n+k-1} = 0 \cap Y_{n+k} = 1) \\
 &= \sum_{k=1}^{\infty} \mathbf{P}(\tau_1 + \cdots + \tau_k > t) (1-p)^{k-1} p \quad (\text{indépendance}) \\
 &= \sum_{k=1}^{\infty} (1-p)^{k-1} p \int_t^{\infty} a \frac{(au)^{k-1}}{(k-1)!} e^{-au} du \quad (\text{loi gamma}) \\
 &= \int_t^{\infty} ap e^{-au} \sum_{k=1}^{\infty} \frac{(a(1-p)u)^{k-1}}{(k-1)!} du \quad (\text{Tonelli-Fubini}) \\
 &= \int_t^{\infty} ap e^{-au} e^{a(1-p)u} du \\
 &= \int_t^{\infty} ap e^{-apu} du.
 \end{aligned}$$

Donc τ est bien exponentielle de paramètre ap et (N_t^1) est un processus de Poisson de paramètre ap . \square

2.3 Définition d'un processus markovien de sauts

Nous allons chercher à modéliser un système qui est à valeurs dans un espace E fini ou dénombrable et qui évolue dans le temps. À chaque instant, le système est décrit par une variable aléatoire à valeurs dans E . Alors que pour une chaîne de Markov le temps est supposé discret, pour un processus markovien de sauts le temps est continu. On note X_t la variable aléatoire représentant le système au temps t ($t \in \mathbf{R}_+$). On appelle processus aléatoire la famille de variables aléatoires $(X_t)_{t \in \mathbf{R}_+}$. Un processus de sauts est un processus qui évolue par sauts successifs à des instants aléatoires. Un processus de Markov est un processus aléatoire dont la loi du processus après un instant t dépend du passé du processus uniquement à travers la connaissance de X_t .

2.3.1 Définitions

DÉFINITION 2.3.1 *La famille $(X_t)_{t \geq 0}$ de variables aléatoires à valeurs dans un espace fini ou dénombrable E est appelée un processus de sauts si pour tout ω de l'espace de probabilité, la fonction de \mathbf{R}_+ dans E qui à t associe $X_t(\omega)$ est constante par morceaux, continue à droite, limitée à gauche et la suite des instants de saut de cette fonction est une suite infinie qui tend vers l'infini.*

L'hypothèse de continuité à droite et de limite à gauche est assez naturelle. On exclut de notre étude des processus dont les temps de sauts s'accumuleraient sur un temps fini.

Notons $(T_n)_{n \geq 0}$ la suite des instants de saut du processus $(X_t)_{t \geq 0}$. Nous allons nous intéresser à une classe de processus de sauts appelés processus markoviens de sauts. Ces processus évoluent de la manière suivante. Si le processus se trouve dans l'état x à l'issue du saut intervenant à l'instant T_n , le temps de séjour dans l'état x ($T_{n+1} - T_n$) suit une loi exponentielle de paramètre $\lambda(x)$. Le paramètre de cette loi dépend de l'état x où le processus se trouve. À part cette dépendance de l'état x , $T_{n+1} - T_n$ est indépendante du passé du processus. À l'instant T_{n+1} , le processus saute de l'état x à l'état y ($x \neq y$) avec une probabilité $Q(x, y)$ indépendante de $T_{n+1} - T_n$ et du passé. L'évolution du processus est donc déterminée par une suite de réels $(\lambda(x))$, $(x \in E)$ et une matrice markovienne $(Q(x, y))$, $(x, y \in E^2)$, telle que $Q(x, x) = 0$ pour tout état x . On supposera toujours que la suite $(\lambda(x))_{x \in E}$ est bornée et formée de nombres strictement positifs. De manière plus formelle, on peut donner la définition suivante :

DÉFINITION 2.3.2 *Soit $(X_t)_{t \geq 0}$ un processus de sauts à valeurs dans un espace E , d'instants de sauts $(T_n)_{n \in \mathbf{N}}$. $(X_t)_{t \geq 0}$ est un processus markovien de sauts s'il existe*

- *une suite bornée de nombres réels $(\lambda(x), x \in E)$ strictement positifs,*
- *une matrice markovienne $(Q(x, y), (x, y) \in E^2)$ de diagonale nulle,*

tel que pour tout entier n , pour tous $x_{n+1}, x_n, x_{n-1}, \dots, x_0 \in E$ et pour tous $u_{n+1}, u_n, u_{n-1}, \dots, u_0 \in \mathbf{R}_+$:

$$\begin{aligned} \mathbf{P}(X_{T_{n+1}} = x_{n+1}, T_{n+1} - T_n > u_{n+1} \mid X_{T_n} = x_n, T_n - T_{n-1} > u_n, \\ X_{T_{n-1}} = x_{n-1}, T_{n-1} - T_{n-2} > u_{n-2}, \dots, X_{T_1} = x_1, T_1 - T_0 > u_0, X_{T_0} = x_0) \\ &= \exp(-\lambda(x_n)u_{n+1})Q(x_n, x_{n+1}). \end{aligned}$$

2.3.2 Exemples

Processus de Poisson

Nous avons étudié au chapitre précédent le processus de Poisson. Nous avons vu que le processus de comptage $(N_t)_{t \geq 0}$ est un processus aléatoire à valeurs dans \mathbf{N} ,

qui vaut 0 au temps 0 et qui saute de 1 à chacun des instants aléatoires T_i qui caractérisent le modèle. Quand le processus est dans l'état i , il saute en $i + 1$ au bout d'un temps aléatoire de loi exponentielle de paramètre λ .

Processus à deux états

Considérons une machine qui peut être soit en état de marche, soit en panne. On note X_t la variable aléatoire égale à 0 si la machine est en panne à l'instant t et égale à 1 si elle est en état de marche à l'instant t .

Nous faisons l'hypothèse que les temps de fonctionnement de cette machine sont des variables aléatoires de loi exponentielle de paramètre λ et que les temps de réparation sont des variables aléatoires de loi exponentielle de paramètre μ . Nous supposons également toutes ces variables aléatoires indépendantes.

Le processus $(X_t)_{t \geq 0}$ est un processus markovien de sauts à valeurs dans $\{0, 1\}$. Les paramètres qui interviennent sont : $\lambda(0) = \mu$, $\lambda(1) = \lambda$, $Q(0, 1) = 1$, $Q(1, 0) = 1$.

File d'attente à un serveur

Considérons un guichet où un serveur rend un service. Les clients arrivent à des instants successifs selon un processus de Poisson de paramètre λ . Ils se mettent en file d'attente et sont servis selon leur ordre d'arrivée. Le temps de service pour chaque client est une variable aléatoire de loi exponentielle de paramètre μ . Toutes les variables aléatoires qui interviennent sont indépendantes.

Considérons le processus $(X_t)_{t \geq 0}$ qui représente le nombre de clients en attente (y compris le client en train d'être servi) au temps t . C'est un processus de sauts à valeurs dans \mathbb{N} . Quand un client arrive, le processus saute de $+1$ et quand un client s'en va à la fin de son service, le processus saute de -1 .

Si à un certain instant le processus saute et prend la valeur i ($i > 0$), il va rester en i un temps aléatoire qui vaut $\inf(U_1, U_2)$ où U_1 est le temps nécessaire pour l'arrivée du prochain client et U_2 est le temps nécessaire pour la fin du service en cours. Or ces variables aléatoires sont indépendantes de lois exponentielles de paramètre λ et μ (cf. Prop. 2.2.2). Le temps de séjour dans l'état i sera donc une variable aléatoire de loi exponentielle de paramètre $\lambda + \mu$ (cf. Prop. 2.2.3). La probabilité que le saut suivant soit de $+1$ est la probabilité que U_1 soit plus petite que U_2 , c'est-à-dire $\lambda/(\lambda + \mu)$, tandis que la probabilité pour que le saut soit de -1 vaut de la même manière $\mu/(\lambda + \mu)$ (cf. Prop. 2.2.3).

Si à un certain instant le processus saute et prend la valeur 0, il va rester en 0 un temps aléatoire qui vaut U_1 où U_1 est le temps nécessaire pour l'arrivée du pro-

chain client, c'est-à-dire un temps aléatoire de loi exponentielle de paramètre λ . Le saut à l'issue de ce temps est nécessairement de $+1$.

On a bien la description d'un processus markovien de sauts.

2.3.3 Propriété de Markov

La définition d'un processus markovien de sauts implique la propriété de Markov suivante.

PROPOSITION 2.3.3 *Si $(X_t)_{t \geq 0}$ est un processus markovien de sauts à valeurs dans E , il satisfait la propriété de Markov suivante : pour tout $n \geq 0$, pour toute suite croissante d'instants $0 < t_1 < t_2 < \dots, t_n < t_{n+1} \in \mathbf{R}_+$ et toute suite $x_1, x_2, \dots, x_n, x_{n+1}$ d'états de E ,*

$$\begin{aligned} \mathbf{P}(X_{t_{n+1}} = x_{n+1} \mid X_{t_n} = x_n, \dots, X_{t_1} = x_1, X_0 = x_0) &= \mathbf{P}(X_{t_{n+1}} = x_{n+1} \mid X_{t_n} = x_n) \\ &= \mathbf{P}(X_{t_{n+1}-t_n} = x_{n+1} \mid X_0 = x_n). \end{aligned}$$

Nous allons juste donner une justification heuristique de cette proposition. Si T_k est l'instant de saut du processus qui précède l'instant t_n , d'après la définition, le processus perd la mémoire de ce qui s'est passé avant T_k , sachant qu'il se trouve dans l'état x_n à cet instant. Donc la seule influence éventuelle du passé de t_n est le temps aléatoire déjà passé dans l'état x_n . Mais la loi de $T_{k+1} - T_k$ est une loi exponentielle de paramètre $\lambda(x_n)$. On sait que le processus n'a pas sauté entre T_k et t_n , donc la loi de $T_{k+1} - t_n$ est la loi de $(T_{k+1} - T_k - t_n + T_k)$ sachant que $T_{k+1} - T_k$ est plus grande que $t_n - T_k$. On sait que c'est encore une loi exponentielle de paramètre $\lambda(x_n)$. Donc à partir de t_n , sachant que le processus est dans l'état x_n , le processus oublie tout le passé.

Les quantités $\mathbf{P}(X_s = y \mid X_t = x)$ ne dépendent que de x , y et $t - s$. On note alors :

$$\mathbf{P}(X_s = y \mid X_t = x) = \mathbf{P}(X_{s-t} = y \mid X_0 = x) = P_{s-t}(x, y).$$

DÉFINITION 2.3.4 *On appelle probabilités de transition la famille $(P_t)_{t \in \mathbf{R}_+}$ de matrices markoviennes définies par*

$$\forall (x, y) \in E^2, \quad P_t(x, y) = \mathbf{P}(X_t = y \mid X_0 = x).$$

Pour tout réel positif t , la matrice $P_t(.,.)$ est une matrice markovienne (on dit aussi matrice stochastique). L'hypothèse de continuité à droite sur les trajectoires entraîne que, pour tous x et y , les fonctions qui à t associent $P_t(x, y)$ sont continues à droite. En particulier, si on désigne par I la matrice identité sur $E \times E$, on a pour tous x et y de E :

$$\lim_{t \rightarrow 0_+} P_t(x, y) = I(x, y).$$

Les probabilités de transition satisfont l'équation de Chapman-Kolmogorov suivante.

PROPOSITION 2.3.5 *Si $(P_t(.,.))_{t \geq 0}$ est la famille des matrices de transition d'un processus de Markov homogène, pour tous temps s et t et pour tous états x et y , on a la relation :*

$$P_{t+s}(x, y) = \sum_{z \in E} P_t(x, z) P_s(z, y).$$

Matriciellement cette égalité s'écrit : $P_{t+s} = P_t P_s$.

Démonstration.

$$\begin{aligned} P_{t+s}(x, y) &= \mathbf{P}(X_{t+s} = y \mid X_0 = x) \\ &= \sum_{z \in E} \mathbf{P}(X_{t+s} = y \cap X_t = z \mid X_0 = x) \\ &= \sum_{z \in E} \mathbf{P}(X_{t+s} = y \mid X_t = z, X_0 = x) \mathbf{P}(X_t = z \mid X_0 = x) \\ &= \sum_{z \in E} \mathbf{P}(X_s = y \mid X_0 = z) \mathbf{P}(X_t = z \mid X_0 = x). \end{aligned}$$

□

2.4 Matrice génératrice

On va définir une nouvelle matrice, la matrice génératrice, qui caractérise le processus.

2.4.1 Définitions

PROPOSITION 2.4.1 *Considérons un processus markovien de sauts $(X_t)_{t \geq 0}$ associé à la famille bornée de paramètres strictement positifs $(\lambda(x), x \in E)$ et à la matrice markovienne de diagonale nulle $(Q(x, y), (x, y) \in E^2)$. Les probabilités de transition $t \rightarrow P_t(x, y)$ de ce processus sont dérивables à droite en 0 et on a :*

$$\begin{aligned} \forall (x, y) \in E \times E, (x \neq y) \quad \lim_{t \rightarrow 0_+} \frac{1}{t} P_t(x, y) &= \lambda(x) Q(x, y), \\ \forall x \in E, \quad \lim_{t \rightarrow 0_+} \frac{1}{t} (P_t(x, x) - 1) &= -\lambda(x). \end{aligned}$$

Démonstration. Soient x et y deux états différents. On note $\mathbf{P}_x(.)$ la probabilité conditionnelle $\mathbf{P}(. \mid X_0 = x)$ et on considère les deux premiers instants de saut T_1

$(T_1 \leq t \text{ car } x \neq y)$ et T_2 :

$$\begin{aligned}\mathbf{P}_x(X_t = y) &= \mathbf{P}_x(X_t = y, T_1 \leq t) \\ &= \mathbf{P}_x(T_1 \leq t, X_{T_1} = y, T_2 > t) + \mathbf{P}_x(T_1 \leq t, X_t = y, T_2 \leq t) \\ &= \mathbf{P}_x(\tau_1 \leq t, X_{T_1} = y, \tau_1 + \tau_2 > t) + \mathbf{P}_x(\tau_1 \leq t, X_t = y, \tau_1 + \tau_2 \leq t).\end{aligned}$$

Évaluons le premier terme en utilisant le fait que τ_1 et τ_2 sont deux variables exponentielles indépendantes :

$$\begin{aligned}\mathbf{P}_x(\tau_1 \leq t, X_{T_1} = y, \tau_1 + \tau_2 > t) &= Q(x, y) \int_{\{u \leq t, u+v > t\}} \lambda(x) e^{-\lambda(x)u} \lambda(y) e^{-\lambda(y)v} du dv \\ &= Q(x, y) \begin{cases} \frac{\lambda(x)}{\lambda(x) - \lambda(y)} (e^{-\lambda(y)t} - e^{-\lambda(x)t}) & \text{si } \lambda(x) \neq \lambda(y) \\ \lambda(x) t e^{-\lambda(x)t} & \text{si } \lambda(x) = \lambda(y) \end{cases} \\ &= Q(x, y) \lambda(x) t + o(t).\end{aligned}$$

De même montrons que le second terme est infiniment petit devant t :

$$\begin{aligned}\mathbf{P}_x(\tau_1 \leq t, X_t = y, \tau_1 + \tau_2 \leq t) &\leq \mathbf{P}_x(\tau_1 \leq t, \tau_2 \leq t) \\ &\leq \sum_{z \in E} \mathbf{P}_x(\tau_1 \leq t, X_{T_1} = z, \tau_2 \leq t) \\ &\leq \sum_{z \in E} (1 - e^{-\lambda(x)t}) Q(x, z) (1 - e^{-\lambda(z)t}) \\ &\leq \sum_{z \in E} \lambda(x) Q(x, z) \lambda(z) t^2 \\ &\leq \lambda(x) M t^2 \quad \text{avec } M \text{ majorant des } (\lambda(z), z \in E).\end{aligned}$$

En définitive on trouve :

$$\mathbf{P}(X_t = y \mid X_0 = x) = \lambda(x) Q(x, y) t + o(t).$$

D'autre part dans le cas $y = x$:

$$\mathbf{P}_x(X_t = x) = \mathbf{P}_x(T_1 > t) + \mathbf{P}_x(T_1 \leq t, T_2 \leq t, X_t = x).$$

Le premier terme vaut $\mathbf{P}_x(T_1 > t) = e^{-\lambda(x)t} = 1 - \lambda(x)t + o(t)$ et le second est infiniment petit devant t :

$$\begin{aligned}\mathbf{P}_x(T_1 \leq t, T_2 \leq t, X_t = x) &\leq \mathbf{P}_x(T_1 \leq t, T_2 \leq t) \\ &\leq \mathbf{P}_x(\tau_1 \leq t, \tau_2 \leq t) \\ &\leq \lambda(x) M t^2.\end{aligned}$$

On en déduit que :

$$\mathbf{P}(X_t = x \mid X_0 = x) - 1 = -\lambda(x)t + o(t).$$

□

DÉFINITION 2.4.2 *On appelle générateur (ou matrice génératrice) d'un processus markovien de sauts de probabilités de transition $(P_t)_{t \geq 0}$ la matrice A dérivée à droite en 0 de (P_t) :*

$$\forall (x, y) \in E^2 \quad A(x, y) = \lim_{t \rightarrow 0_+} \frac{P_t(x, y) - I(x, y)}{t} = \begin{cases} \lambda(x)Q(x, y) & \text{si } x \neq y, \\ -\lambda(x) & \text{si } x = y. \end{cases}$$

On vient de voir que pour un processus markovien de sauts cette matrice existe toujours.

2.4.2 Exemples

Processus de Poisson

Pour tout x entier, $\lambda(x)$ est égal à λ et $Q(x, x+1)$ égal à 1. D'où pour tout x :

$$\begin{aligned} \forall y < x \quad A(x, y) &= 0, \\ \forall x \quad A(x, x) &= -\lambda \quad \text{et} \quad A(x, x+1) = \lambda, \\ \forall y > x+1 \quad A(x, y) &= 0. \end{aligned}$$

On a le générateur :

$$A = \begin{pmatrix} -\lambda & \lambda & 0 & 0 & \dots \\ 0 & -\lambda & \lambda & 0 & \dots \\ 0 & 0 & -\lambda & \lambda & \dots \\ 0 & 0 & 0 & -\lambda & \ddots \\ \vdots & \ddots & \ddots & \ddots & \ddots \end{pmatrix}$$

Processus à deux états

On trouve :

$$\begin{aligned} A(0, 0) &= -\mu & A(0, 1) &= \mu \\ A(1, 0) &= \lambda & A(1, 1) &= -\lambda \end{aligned}$$

Le générateur vaut :

$$A = \begin{pmatrix} -\mu & \mu \\ \lambda & -\lambda \end{pmatrix}$$

File d'attente à un serveur

On trouve :

$$\begin{aligned} A(0, 0) &= -\lambda & A(0, 1) &= \lambda \\ A(n, n-1) &= \mu & A(n, n) &= -(\lambda + \mu) & A(n, n+1) &= \lambda \quad \text{si } n > 0. \end{aligned}$$

2.4.3 Équations de Kolmogorov

Le calcul direct des probabilités de transition est souvent difficile, voire impossible. Le théorème suivant va donner un moyen de calculer les probabilités de transition à partir du générateur.

PROPOSITION 2.4.3 (Équations de Kolmogorov) *Considérons un processus markovien de sauts de probabilités de transition $(P_t)_{t \geq 0}$ et de générateur A. On a la relation :*

$$(P_t)' = P_t A = A P_t$$

$(P_t)'$ désigne la dérivée de P_t au point t strictement positif et la dérivée à droite si t vaut 0 :

$$(P_t)'(x, y) = \lim_{s \rightarrow 0} \frac{P_{t+s}(x, y) - P_t(x, y)}{s}.$$

Démonstration. Nous ne ferons la démonstration que dans le cas où E est fini. La formule de Chapman-Kolmogorov s'écrit :

$$\forall (x, y) \in E^2, \quad P_{t+s}(x, y) = \sum_{z \in E} P_t(x, z) P_s(z, y).$$

On peut dériver cette formule par rapport à s et faire $s = 0$. On trouve $(P_t)' = P_t A$. La seconde formule s'obtient en dérivant par rapport à t et en faisant $t = 0$. \square

Si on connaît le générateur d'un processus de Markov, on peut trouver les probabilités de transition comme solutions d'une équation différentielle linéaire à coefficients constants. Si l'espace E est fini on a la formule :

$$P_t = e^{tA} = \sum_{k \geq 0} \frac{t^k}{k!} A^k.$$

Le générateur du processus détermine les probabilités de transition, donc la loi du processus. En pratique, lorsque E est fini, le calcul explicite de e^{tA} est possible si on sait diagonaliser la matrice A.

Exemple du processus à deux états : E est égal à $\{0, 1\}$ et le générateur du processus existe et vaut :

$$A = \begin{pmatrix} -\mu & \mu \\ \lambda & -\lambda \end{pmatrix}$$

En diagonalisant A (à faire en exercice) on trouve :

$$P_t = e^{tA} = \frac{1}{\lambda + \mu} \begin{pmatrix} \lambda + \mu e^{-(\lambda + \mu)t} & \mu - \mu e^{-(\lambda + \mu)t} \\ \lambda - \lambda e^{-(\lambda + \mu)t} & \mu + \lambda e^{-(\lambda + \mu)t} \end{pmatrix}$$

2.5 Propriétés asymptotiques

2.5.1 Mesures invariantes, réversibles

L'analyse du comportement asymptotique passe par l'identification d'une probabilité invariante sur l'espace des états E . Une probabilité π sur E est invariante par le processus si pour tout temps t et tout état y de E on a $\pi P_t = \pi$ soit :

$$\sum_{x \in E} \pi(x) P_t(x, y) = \pi(y).$$

c'est-à-dire que si π est la loi de probabilité de X_0 , alors pour tout temps t la loi de X_t est encore π .

En dérivant cette relation en t et en faisant t égal à 0, on arrive à la définition suivante.

DÉFINITION 2.5.1 *La probabilité π sur E est dite invariante pour un processus de Markov de générateur A si et seulement si $\pi A = 0$ soit :*

$$\forall y \in E, \quad \sum_{x \in E} \pi(x) A(x, y) = 0.$$

Grâce aux équations de Kolmogorov, on démontre que si π est une probabilité invariante au sens de la définition ci-dessus, on a bien :

$$\forall t \geq 0, \forall y \in E, \quad \sum_{x \in E} \pi(x) P_t(x, y) = \pi(y).$$

L'identification d'une probabilité invariante nécessite la résolution d'un système linéaire de taille égale au nombre d'éléments de E .

DÉFINITION 2.5.2 *La probabilité π sur E est dite réversible pour un processus de Markov de générateur A si et seulement si :*

$$\forall (x, y) \in E \times E, \quad \pi(x) A(x, y) = \pi(y) A(y, x).$$

PROPOSITION 2.5.3 *Toute probabilité réversible pour un processus est invariante pour ce processus.*

Démonstration. Si π est réversible, on a pour tout y dans E :

$$\sum_{x \in E} \pi(x) A(x, y) = \sum_{x \in E} \pi(y) A(y, x) = \pi(y) \sum_{x \in E} A(y, x) = 0. \quad \square$$

En pratique pour chercher une probabilité invariante, il est recommandé de commencer par chercher une probabilité réversible, plus facile à identifier quand elle existe.

2.5.2 Exemples

Processus de Poisson

L'équation d'invariance donne : $\forall i \geq 0, \pi(i) = \pi(i+1)$. Il n'existe pas de probabilité invariante.

Processus à deux états

La matrice génératrice vaut : $\begin{pmatrix} -\mu & \mu \\ \lambda & -\lambda \end{pmatrix}$.

On trouve comme unique probabilité invariante : $\pi(0) = \frac{\lambda}{\lambda+\mu}$ et $\pi(1) = \frac{\mu}{\lambda+\mu}$.

File d'attente à un serveur

La matrice génératrice est de la forme :

$$A = \begin{pmatrix} -\lambda & \lambda & 0 & 0 & 0 & \dots \\ \mu & -(\lambda + \mu) & \lambda & 0 & 0 & \dots \\ 0 & \mu & -(\lambda + \mu) & \lambda & 0 & \dots \\ 0 & 0 & \mu & -(\lambda + \mu) & \lambda & \ddots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots \end{pmatrix}$$

Les équations de réversibilité s'écrivent :

$$\forall x \geq 1, \pi(x-1)\lambda = \pi(x)\mu$$

d'où :

$$\forall x \geq 0, \pi(x) = \pi(0) \frac{\lambda^x}{\mu^x}.$$

Il existe une probabilité réversible si la série de terme général $\pi(x)$ est convergente, c'est-à-dire si $\lambda < \mu$. Cette probabilité réversible s'écrit alors :

$$\forall x \geq 0, \pi(x) = \left(1 - \frac{\lambda}{\mu}\right) \left(\frac{\lambda}{\mu}\right)^x.$$

2.5.3 Théorème ergodique

DÉFINITION 2.5.4 *Un processus markovien de saut de matrice génératrice A est dit irréductible si pour tous états x et y ($x \neq y$), il existe des états x_1, x_2, \dots, x_n tous différents tels que :*

$$A(x, x_1)A(x_1, x_2) \dots A(x_{n-1}, x_n)A(x_n, y) > 0.$$

Ceci signifie que l'on peut passer (en plusieurs étapes) de n'importe quel état x à n'importe quel état y avec une probabilité strictement positive. Cela est vrai pour le processus à deux états dont les éléments de la matrice génératrice sont différents de 0 et pour le processus associé à une file d'attente. Ce n'est pas vrai pour le processus de Poisson (le processus X_t est croissant).

Nous admettrons le théorème limite suivant.

THÉORÈME 2.5.5 *Supposons que le processus markovien de sauts $(X_t)_{t \geq 0}$ est irréductible sur E et admet une probabilité invariante π , alors :*

- cette probabilité invariante est unique,
- $\forall (x, y) \in E^2, \lim_{t \rightarrow +\infty} P_t(x, y) = \pi(y),$
- $\forall y \in E, \lim_{t \rightarrow +\infty} \frac{1}{t} \int_{[0, t]} \mathbf{1}_{X_s=y} ds = \pi(y) \text{ p.s.}$

Le comportement stationnaire du processus est donc décrit par l'unique probabilité invariante quand elle existe.

Ce résultat exprime qu'au bout d'un certain temps, le système se stabilise dans un régime d'équilibre appelé « régime stationnaire » et que

- la probabilité d'être dans l'état y en régime stationnaire est donnée par $\pi(y)$,
- la proportion de temps passé dans l'état y entre les instants 0 et t est également donnée par $\pi(y)$ lorsque t est assez grand.

Chapitre 3

Files d'attente

3.1 Introduction

Nous avons vu dans le chapitre 1 qu'une file d'attente est décrite par la loi des temps d'arrivée des clients, la loi des temps de service, le nombre de serveurs et la longueur maximum de la file d'attente. Dans ce chapitre nous supposerons que les clients arrivent selon un processus de Poisson. Nous supposerons que les temps de service sont des variables aléatoires indépendantes de même loi, indépendantes du processus des arrivées. Dans un premier temps nous supposerons que la loi de service est exponentielle. Nous pourrons alors utiliser les techniques vues au chapitre précédent et calculer explicitement les caractéristiques de l'état stationnaire quand il existe. Dans le cas plus général d'une loi de service quelconque, nous verrons qu'en utilisant une chaîne de Markov on peut obtenir des résultats.

3.2 Files M/M/1.

Dans ces files les clients arrivent selon un processus de Poisson de paramètre λ . Les temps de service sont des variables aléatoires indépendantes, de mêmes lois exponentielles de paramètre μ et indépendantes du processus des arrivées. On note N_t le nombre de clients en attente, y compris ceux qui sont en train d'être servis, au temps t .

PROPOSITION 3.2.1 *Le nombre N_t de clients présents (en attente ou en service) à l'instant t dans une file M/M/s (s pouvant valoir l'infini) est un processus mar-*

kovien de sauts à valeurs dans \mathbf{N} de matrice génératrice :

$$\begin{aligned} A(n, n+1) &= \lambda \\ A(n, n-1) &= \mu \inf(n, s) \quad \text{si } n > 0 \\ A(n, m) &= 0 \quad \text{si } m \neq n \pm 1, \text{ et } m \neq n. \end{aligned}$$

Démonstration. Le processus N_t est à l'évidence un processus de sauts. Les propriétés de perte de mémoire du processus de Poisson et de la loi exponentielle permettent de montrer que ce processus est markovien. Les valeurs de la matrice génératrice se calculent sans problème. \square

Il est clair que ce processus est irréductible. Nous avons vu au chapitre précédent, au moins dans le cas $s = 1$, qu'il admet une mesure réversible unique à une constante multiplicative près. Pour chaque valeur de s , suivant les valeurs de λ et μ cette mesure réversible est finie ou non. Lorsqu'elle est finie, il existe une probabilité réversible, donc invariante et le théorème limite 2.5.5 s'applique. Le processus $(N_t)_{t \geq 0}$ se stabilise sur un état dont la loi est cette probabilité réversible. Les caractéristiques de la longueur de la chaîne dans l'état stationnaire sont donc données par cette probabilité. Nous allons faire ces calculs pour différentes valeurs de s .

3.2.1 Cas M/M/1

Dans ce cas il y a un seul serveur. C'est la file d'attente la plus simple. Le générateur est donné par :

$$\begin{aligned} \forall n \in \mathbf{N}, \quad A(n, n+1) &= \lambda \\ \forall n \in \mathbf{N}^*, \quad A(n, n-1) &= \mu. \end{aligned}$$

On obtient comme mesure réversible : $m_n = m_0 \left(\frac{\lambda}{\mu} \right)^n$.

Cette mesure est finie si $\lambda < \mu$ et alors l'unique probabilité réversible vaut :

$$\pi_n = \left(1 - \frac{\lambda}{\mu} \right) \left(\frac{\lambda}{\mu} \right)^n.$$

On a montré la proposition :

PROPOSITION 3.2.2 *Dans le cas M/M/1, le processus $(N_t)_{t \geq 0}$ admet une unique probabilité invariante lorsque le paramètre λ du processus de Poisson des arrivées est strictement inférieur au paramètre μ de la loi exponentielle des services (c'est-à-dire par unité de temps le nombre moyen d'arrivées est inférieur au nombre moyen de clients servis). Cette probabilité est une loi géométrique sur \mathbf{N} de paramètre $a = \lambda / \mu$.*

On peut montrer que lorsque λ est strictement plus grand que μ la longueur de la file tend vers l'infini. Lorsque λ est égal à μ , le processus n'admet pas de probabilité invariante.

Lorsque λ est strictement plus petit que μ , la longueur (nombre de clients présents) de la file d'attente en régime stationnaire a pour moyenne et variance :

$$E(N) = \frac{\frac{\lambda}{\mu}}{1 - \frac{\lambda}{\mu}} \quad \text{Var}(N) = \frac{\frac{\lambda}{\mu}}{\left(1 - \frac{\lambda}{\mu}\right)^2}.$$

Le *facteur d'utilisation* du serveur (« mean busy period » en anglais) représente en régime stationnaire le pourcentage du temps où le serveur est occupé, c'est donc la moyenne de la variable aléatoire qui vaut 0 avec probabilité π_0 et 1 avec probabilité $1 - \pi_0$. Ce facteur vaut donc $1 - \pi_0$, soit λ/μ .

Une variable aléatoire intéressante est le *temps de réponse* du système c'est-à-dire le temps total passé dans le système par les différents clients. En régime stationnaire toutes ces variables aléatoires ont même loi. Si le client trouve à son arrivée n personnes dans la file il passera dans le système un temps aléatoire égal à la somme des temps de service de ces n personnes plus son temps de service à lui. Ce temps est égal à la somme de $n + 1$ variables aléatoires indépendantes de loi exponentielle de paramètre μ (le temps nécessaire pour finir de servir la personne en train d'être servie quand le client arrive est encore une loi exponentielle de paramètre μ , d'après la propriété d'absence de mémoire). Cette loi admet la densité sur \mathbf{R}_+ (cf. prop. 2.2.4) :

$$g_{n+1}(x) = \frac{\mu(\mu x)^n e^{-\mu x}}{n!}.$$

En notant N le nombre de clients en attente à l'arrivée du client test et R le temps total d'attente plus service de ce client on a :

$$\begin{aligned} \mathbf{P}(R \leq t) &= \sum_{n=0}^{\infty} \mathbf{P}(N = n) \mathbf{P}(R \leq t \mid N = n) \\ &= \sum_{n=0}^{\infty} \pi_n \int_0^t g_{n+1}(x) dx \\ &= \sum_{n=0}^{\infty} \left(1 - \frac{\lambda}{\mu}\right) \frac{\lambda^n}{\mu^n} \int_0^t \frac{\mu(\mu x)^n e^{-\mu x}}{n!} dx \\ &= \int_0^t \left(1 - \frac{\lambda}{\mu}\right) \mu e^{-\mu x} \sum_{n=0}^{\infty} \frac{\lambda^n}{\mu^n} \frac{(\mu x)^n}{n!} dx \quad (\text{Fubini, fonction positive}) \\ &= \int_0^t (\mu - \lambda) e^{-\mu x} e^{\lambda x} dx. \end{aligned}$$

On a établi le résultat suivant :

PROPOSITION 3.2.3 *La variable aléatoire temps de réponse du système (temps d'attente plus service d'un client en régime stationnaire) suit une loi exponentielle de paramètre $\mu - \lambda$, sa densité vaut :*

$$f(x) = (\mu - \lambda) e^{-(\mu - \lambda)x} \mathbf{1}_{x \geq 0}.$$

Le temps moyen passé dans le système par un client en régime stationnaire vaut $1/(\mu - \lambda)$. Cette quantité dépend de λ et de μ et pas uniquement du quotient λ/μ . Un système peut être presque saturé avec un quotient λ/μ proche de 1 et une longueur de file d'attente grande en moyenne mais avoir un temps de réponse moyen faible.

3.2.2 Cas M/M/ ∞

Dans ce cas il y a une infinité de serveurs. Le générateur est donné par :

$$\begin{aligned} \forall n \in \mathbf{N}, \quad A(n, n+1) &= \lambda, \\ \forall n \in \mathbf{N}^*, \quad A(n, n-1) &= n\mu. \end{aligned}$$

PROPOSITION 3.2.4 *Dans le cas M/M/ ∞ , le processus $(N_t)_{t \geq 0}$ possède une probabilité invariante unique qui suit une loi de Poisson de paramètre λ/μ .*

Démonstration. Une mesure réversible satisfait la relation :

$$\forall n, \quad m_n \lambda = m_{n+1} \mu (n+1), \quad \text{d'où} \quad m_n = m_0 \frac{1}{n!} \left(\frac{\lambda}{\mu}\right)^n.$$

Cette mesure est de masse totale finie pour tous λ et μ . En la normalisant on trouve le résultat. \square

Dans le cas M/M/ ∞ le *facteur d'utilisation* (ou la charge de service) est le nombre moyen de serveurs occupés. Il est égal au nombre moyen de clients dans la file puisqu'il n'y a pas d'attente. Ce nombre vaut donc λ/μ .

Le *temps moyen de réponse* est égal au temps moyen de service et vaut $1/\mu$.

3.2.3 Cas M/M/s

Il y a s serveurs. Le générateur est donné par :

$$\begin{aligned} \forall n \in \mathbf{N}, \quad A(n, n+1) &= \lambda, \\ \forall n \in \mathbf{N}^*, \quad A(n, n-1) &= \mu \inf(n, s). \end{aligned}$$

PROPOSITION 3.2.5 *Dans le cas M/M/s, le processus $(N_t)_{t \geq 0}$ possède une probabilité invariante unique si $\rho = \frac{\lambda}{s\mu}$ est strictement plus petit que 1. Cette probabilité vaut :*

$$\begin{aligned}\pi_j &= \frac{1}{j!} \left(\frac{\lambda}{\mu}\right)^j \pi_0 && \text{si } j \leq s, \\ \pi_j &= \frac{s^s}{s!} \left(\frac{\lambda}{s\mu}\right)^j \pi_0 && \text{si } j \geq s, \\ \text{avec } \frac{1}{\pi_0} &= \sum_{j=0}^{s-1} \frac{1}{j!} \left(\frac{\lambda}{\mu}\right)^j + \left(1 - \frac{\lambda}{s\mu}\right)^{-1} \frac{1}{s!} \left(\frac{\lambda}{\mu}\right)^s.\end{aligned}$$

Démonstration. Une mesure réversible satisfait : $m_{n-1} \lambda = m_n \mu \inf(n, s)$. On en déduit :

$$m_i = m_0 \frac{1}{i!} \left(\frac{\lambda}{\mu}\right)^i \text{ pour } i \leq s \quad \text{et} \quad m_{s+j} = m_0 \frac{1}{s!s^j} \left(\frac{\lambda}{\mu}\right)^{s+j}.$$

Si $\rho = \lambda/s\mu < 1$, cette mesure est finie et en la normalisant on trouve la probabilité invariante annoncée. \square

Lorsque $\rho = \lambda/s\mu$ est strictement plus grand que 1, le processus $(N_t)_{t \geq 0}$ tend vers l'infini.

Lorsque λ est égal à $s\mu$ le processus n'admet pas de probabilité invariante.

Lorsque $\rho = \lambda/s\mu$ est strictement plus petit que 1, il existe donc un régime stationnaire. Sous ce régime, on peut calculer la probabilité pour que tous les serveurs soient occupés. Cette probabilité vaut :

$$C = \sum_{j=s}^{\infty} \pi_j = \frac{s a^s}{(s-a)s!} \pi_0 = \frac{\rho^s}{1-\rho} \frac{s^s}{s!} \pi_0 \quad \text{avec} \quad a = \frac{\lambda}{\mu} \quad \text{et} \quad \rho = \frac{\lambda}{s\mu}.$$

Ceci constitue la deuxième formule d'Erlang (terminologie européenne) ou « Erlang delay formula » (terminologie USA).

La charge de service (ou *facteur d'utilisation*) représente le nombre moyen de serveurs occupés en régime stationnaire et vaut :

$$a' = \sum_{j=1}^{s-1} j \pi_j + s \sum_{j=s}^{\infty} \pi_j = a \sum_{j=1}^{s-1} \frac{a^{j-1}}{(j-1)!} \pi_0 + s C = a \pi_0 \left(\sum_{k=0}^{s-2} \frac{a^k}{k!} + \frac{s a^{s-1}}{(s-a)(s-1)!} \right) = a.$$

Donc la charge de service en régime stationnaire vaut encore $a = \lambda/\mu$, chaque serveur est occupé en moyenne ρ % de son temps.

On peut calculer le nombre moyen de clients présents (en attente ou en service) :

$$\begin{aligned}
 \mathbf{E}(N) &= \sum_{j=0}^s j \frac{a^j \pi_0}{j!} + \sum_{j=s+1}^{\infty} j \frac{a^j \pi_0}{s! s^{j-s}} \\
 &= a \left(1 - \frac{a^s \pi_0}{(s-1)!(s-a)} \right) + \sum_{i=1}^{\infty} (i+s) \frac{a^{i+s} \pi_0}{s! s^i} \\
 &= a + \frac{\pi_0 a^{s+1}}{(s-1)!(s-a)^2}.
 \end{aligned}$$

Il est intéressant de comparer le comportement d'une file d'attente M/M/2 de paramètres λ et μ avec le comportement d'une file M/M/1 de paramètres λ et 2μ . Les deux files ont la même condition de stationnarité : $\lambda < 2\mu$. La longueur moyenne de la file M/M/1 est $\lambda/(2\mu - \lambda)$. Calculons la quantité analogue pour la file M/M/2. Pour cette file la probabilité stationnaire vaut :

$$\pi_0 = \frac{1}{1 + a + \frac{a^2}{2-a}} = \frac{2-a}{2+a} \quad \text{et} \quad \pi_j = \frac{a^j}{2^{j-1}} \pi_0 \quad \text{pour } j \geq 1.$$

Alors on trouve :

$$\mathbf{E}(N) = \frac{4a}{4-a^2} = \frac{4\lambda\mu}{(2\mu-\lambda)(2\mu+\lambda)} > \frac{\lambda}{2\mu-\lambda}.$$

La file d'attente avec un serveur est plus petite en moyenne que la file d'attente avec deux serveurs. C'est normal intuitivement car la capacité de service maximale du système à deux serveurs est moins utilisée que celle à un serveur.

Il est possible de trouver une formule simple pour la loi du temps d'attente d'un client en régime stationnaire. Le temps d'attente est le temps passé par un client entre son arrivée et son début de service. Appelons W ce temps d'attente. On a déjà calculé $\mathbf{P}(W > 0)$, c'est la probabilité pour que tous les serveurs soient occupés. Calculons $\mathbf{P}(W > t)$. Si un client arrive et trouve s personnes en train d'être servies et aucune en attente, il attendra un temps aléatoire exponentiel de paramètre μs avant de commencer à être servi (cf. prop. 2.2.2 et 2.2.3). Si un client arrive et trouve $s+j$ personnes dans le système, il attendra avant de commencer à être servi un temps égal à la somme de $j+1$ variables aléatoires indépendantes exponentielles de paramètre μs . En notant N le nombre de clients en attente à l'arrivée du client test, on a :

$$\begin{aligned}
 \mathbf{P}(W > t) &= \sum_{j \geq 0} \mathbf{P}(N = j+s) \mathbf{P}(W > t | N = j+s) \\
 &= \sum_{j \geq 0} \pi_{s+j} \int_t^{\infty} \frac{(s\mu)^{j+1} x^j}{j!} e^{-\mu s x} dx
 \end{aligned}$$

$$\begin{aligned}
 &= \sum_{j \geq 0} \pi_0 \frac{a^{j+s}}{s!s^j} \int_t^\infty \frac{(s\mu)^{j+1}x^j}{j!} e^{-\mu sx} dx \quad \text{avec } a = \frac{\lambda}{\mu} \\
 &= \pi_0 \frac{a^s \mu s}{s!} \int_t^\infty e^{-\mu sx} \sum_{j \geq 0} \frac{(a\mu x)^j}{j!} dx \quad (\text{Fubini, fonction positive}) \\
 &= \pi_0 \frac{a^s \mu s}{s!} \int_t^\infty e^{-\mu x(s-a)} dx \\
 &= \mathbf{P}(W > 0) \int_t^\infty \mu(s-a) e^{-\mu(s-a)x} dx.
 \end{aligned}$$

On a établi le résultat suivant :

PROPOSITION 3.2.6 *Conditionnellement en ($W > 0$), le temps d'attente W d'un client en régime stationnaire suit la loi v exponentielle de paramètre $(\mu s - \lambda)$. La loi de la variable aléatoire W vaut donc :*

$$(1 - \mathbf{P}(W > 0)) \delta_0 + \mathbf{P}(W > 0) v.$$

Le temps d'attente moyen avant de commencer à être servi vaut :

$$\mathbf{E}(W) = \mathbf{P}(W > 0) \frac{1}{\mu(s-a)} = \frac{a^s \pi_0}{(s-1)! \mu(s-a)^2}.$$

Le temps moyen de réponse vaut alors : $\mathbf{E}(W) + 1/\mu$. On peut calculer en exercice la loi du temps de réponse. On trouve une densité égale à :

$$\mu e^{-\mu x} \left(1 - \frac{\mathbf{P}(W > 0)}{1-s+a} \right) + \mu(s-a) e^{-\mu(s-a)x} \frac{\mathbf{P}(W > 0)}{1-s+a}.$$

Si R est le temps de réponse pour un client, on constate que pour une file $M/M/s$ on a la formule :

$$\boxed{\mathbf{E}(N) = \lambda \mathbf{E}(R)}$$

Cette formule, également vraie pour une file $M/M/\infty$ est très générale. Elle porte le nom de *formule de Little*¹. Ce qu'on appelle en général *formule de Little* est plutôt l'égalité

$$\boxed{\mathbf{E}(L) = \lambda \mathbf{E}(W)}$$

où L désigne le nombre de clients en attente et W le temps d'attente d'un client. L'équivalence des deux formules résulte des égalités $R = W + \sigma$ et $N = L + S$ où S

1. Ce n'est pas $\mathbf{E}(R) = \mathbf{E}(\sigma) \mathbf{E}(N)$ mais $\mathbf{E}(R) = \mathbf{E}(\tau) \mathbf{E}(N)$!

désigne le nombre de serveurs occupés. $E(S)$ est le facteur d'utilisation et on a vu que $E(S) = \rho = \lambda E(\sigma)$, ce qui établit l'équivalence au moins pour les files M/M/1.

Voici une justification heuristique de la formule de LITTLE : on considère une période test commençant à l'arrivée d'un client et se terminant au début de service du même client. On suppose que l'ordre de service est FIFO. λ est le nombre moyen de clients arrivant par unité de temps, donc $\lambda E(W)$ est le nombre moyen de clients arrivant dans la file pendant la période test, de durée moyenne $E(W)$. Si le système est en équilibre, le nombre moyen de clients sortant de la file d'attente doit être égal au nombre de clients entrant pendant la même période. Or les clients sortant de la file d'attente pendant la période test sont ceux qui étaient présents à l'arrivée du client test, leur nombre est donc L et a pour moyenne $E(L)$. D'où l'égalité annoncée.

On va maintenant s'intéresser au flot de sortie d'une file d'attente M/M/s.

THÉORÈME 3.2.7 (Burke) *Dans une file d'attente M/M/s (s fini ou non) en régime stationnaire, le flot de sortie est un processus poissonnien de même paramètre que le flot d'entrée.*

Démonstration. Soit $(X_t)_{t \geq 0}$ le processus de naissance et mort représentant le nombre de clients dans le système. On suppose que ce processus est stationnaire, c'est-à-dire que la loi de X_t pour tout t vaut π , unique probabilité réversible qui existe dans les conditions d'étude de la file d'attente.

Fixons un temps T très grand et définissons un processus sur $[0, T]$ par : $t \mapsto X_{T-t}$. Changeons les valeurs de ce processus aux points de saut de manière à ce que chaque trajectoire soit continue à droite. Appelons $(Y_t)_{0 \leq t \leq T}$ le processus obtenu. Ce processus est un processus markovien de saut homogène. En effet si on note $P_t(\cdot, \cdot)$ les probabilités de transition de $(X_t)_{t \geq 0}$, pour toute suite strictement croissante d'instants $0 = t_0 < t_1 < t_2 < \dots < t_k < t_{k+1} < T$:

$$\begin{aligned}
 & P(Y_{t_{k+1}} = a_{k+1} \mid Y_{t_k} = a_k, \dots, Y_0 = a_0) \\
 &= \frac{P(Y_{t_{k+1}} = a_{k+1}, Y_{t_k} = a_k, \dots, Y_0 = a_0)}{P(Y_{t_k} = a_k, \dots, Y_0 = a_0)} \\
 &= \frac{P(X_T = a_0, \dots, X_{T-t_k} = a_k, X_{T-t_{k+1}} = a_{k+1})}{P(X_T = a_0, \dots, X_{T-t_{k-1}} = a_{k-1}, X_{T-t_k} = a_k)} \\
 &= \frac{P(X_T = a_0, \dots, X_{T-t_k} = a_k \mid X_{T-t_{k+1}} = a_{k+1}) P(X_{T-t_{k+1}} = a_{k+1})}{P(X_T = a_0, \dots, X_{T-t_{k-1}} = a_{k-1} \mid X_{T-t_k} = a_k) P(X_{T-t_k} = a_k)}
 \end{aligned}$$

soit en utilisant la propriété de Markov de (X_t) et en simplifiant

$$= P_{t_{k+1}-t_k}(a_{k+1}, a_k) \frac{P(Y_{t_{k+1}} = a_{k+1})}{P(Y_{t_k} = a_k)}$$

Comme le processus $(X_t)_{t \geq 0}$ est stationnaire, les probabilités de transition du processus $(Y_t)_{0 \leq t \leq T}$ valent (poser $a_k = i, a_{k+1} = j$) :

$$Q_t(i, j) = P_t(j, i) \frac{\pi_j}{\pi_i}.$$

Si A est le générateur de $(X_t)_{t \geq 0}$, le générateur de $(Y_t)_{0 \leq t \leq T}$ aux points i et j vaut

$$B(i, j) = A(j, i) \frac{\pi_j}{\pi_i} = A(i, j)$$

puisque la probabilité π est réversible. Le processus Y a même générateur que X , il a donc même loi. Le processus des instants de sauts +1 de Y a même loi que le processus des temps de saut +1 de X , soit la loi d'un processus de Poisson de paramètre λ . Or les sauts +1 de Y correspondent sur $[0, T]$ aux sauts -1 de X , soit aux temps de sortie des clients du système. Ceci démontre bien que ces temps de sortie suivent un processus de Poisson de paramètre λ . \square

3.3 Files M/G/1

Dans ce paragraphe, nous allons étudier une file d'attente à un serveur telle que le flot d'arrivée des clients est un processus de Poisson de paramètre λ et telle que les temps de service des différents clients sont des variables aléatoires indépendantes de même loi toutes indépendantes des arrivées. La discipline de service est toujours premier arrivé, premier servi. Lorsque la loi de service n'est pas exponentielle, le nombre X_t de personnes dans le système au temps t n'est pas un processus de Markov. En effet le temps de service restant à fournir pour le client en cours de service dépend a priori du temps déjà passé à être servi. On ne peut donc plus utiliser les techniques des processus markoviens de sauts. Le fait que les arrivées soient un processus de Poisson va cependant nous permettre d'utiliser une technique markovienne. Nous allons mettre en évidence une chaîne de Markov incluse dans le processus, le nombre de personnes dans le système aux instants de sortie de chaque client.

Soient $(S_n)_{n \geq 1}$ les instants de sorties des clients. Appelons Y_n le nombre X_{S_n} de clients présents à l'instant S_n ou encore *juste après*² l'instant S_n .

PROPOSITION 3.3.1 *Dans une file d'attente M/GI/1, la suite $(Y_n)_{n \geq 1}$ des nombres de clients présents dans le système aux instants de fin de service est une chaîne de Markov à valeurs dans \mathbf{N} .*

2. Se souvenir que les trajectoires du processus X_t sont continues à droite, donc $X_{S_n} = X_{S_n^+}$.

Démonstration. Appelons K_n le nombre de clients arrivés pendant le service du n -ième client. On a la relation :

$$Y_{n+1} = Y_n - \mathbf{1}_{Y_n > 0} + K_n$$

en posant Y_0 égale à 0. La variable aléatoire K_n est indépendante de Y_n vu les propriétés du processus de Poisson des arrivées et l'indépendance des temps de service par rapport à ce processus. Ceci démontre la propriété de Markov. \square

La loi de K_n , lorsque le temps de service σ du n -ième client vaut t , est une loi de Poisson de paramètre λt (loi conditionnelle sachant $\sigma = t$). La loi de K_n s'obtient en intégrant cette loi conditionnelle par rapport à la loi de σ :

$$(3.1) \quad q_k = \mathbf{P}(K_n = k) = \int_0^\infty e^{-\lambda t} \frac{(\lambda t)^k}{k!} d\mathbf{P}_\sigma(t)$$

avec \mathbf{P}_σ loi de service. La chaîne de Markov $(Y_n)_{n \in \mathbb{N}}$ a comme matrice de transition :

$$P = \begin{pmatrix} q_0 & q_1 & q_2 & q_3 & q_4 & \dots \\ q_0 & q_1 & q_2 & q_3 & q_4 & \dots \\ 0 & q_0 & q_1 & q_2 & q_3 & \dots \\ 0 & 0 & q_0 & q_1 & q_2 & \ddots \\ 0 & 0 & 0 & q_0 & q_1 & \ddots \\ \vdots & \vdots & \ddots & \ddots & \ddots & \ddots \end{pmatrix}$$

Comme tous les q_k sont strictement positifs, il est clair que cette chaîne est irréductible sur \mathbb{N} . Sauf au point 0, cette chaîne est identique à une marche aléatoire : en effet, si $Y_n > 0$, $Y_{n+1} = Y_n + (K_n - 1)$ et les K_n sont indépendantes. Les propriétés des marches aléatoires sur \mathbb{Z} sont bien connues. Si la moyenne de la loi des sauts ($\mathbf{E}(K_n - 1)$ ici) est nulle, la marche est récurrente nulle. Si la moyenne est strictement positive, la marche est transiente et tend vers $+\infty$. Si la moyenne est strictement négative, la marche est transiente et tend vers $-\infty$. Il est clair que si la moyenne de la marche est strictement positive, la chaîne comme la marche tend vers plus l'infini et est transiente. Si la moyenne est négative ou nulle, la chaîne partant de n'importe quel point de \mathbb{N} atteint presque sûrement le point 1 qui est donc récurrent et la chaîne est récurrente. Il reste à calculer la moyenne $\mathbf{E}(K_n - 1)$. En permutant la somme et l'intégrale (Tonelli), on obtient :

$$\mathbf{E}(K_n) = \sum_{k \in \mathbb{N}} k q_k = \lambda \mathbf{E}(\sigma).$$

On pose $\rho = \lambda \mathbf{E}(\sigma)$ (quotient des temps moyens de service et d'interarrivée), la moyenne du saut de la marche vaut $\rho - 1$.

On a donc le résultat suivant :

PROPOSITION 3.3.2 *La chaîne de Markov $(Y_n)_{n \in \mathbb{N}}$ est récurrente si $\rho \leq 1$ et transiente dans le cas contraire.*

Si $\rho \leq 1$, il existe une mesure invariante, unique à une constante multiplicative près. L'équation d'invariance s'écrit :

(3.2)

$$\begin{aligned}\pi_0 &= \pi_1 q_0 + \pi_0 q_0 \\ \pi_1 &= \pi_2 q_0 + \pi_1 q_1 + \pi_0 q_1 \\ \pi_2 &= \pi_3 q_0 + \pi_2 q_1 + \pi_1 q_2 + \pi_0 q_2 \\ \vdots &= \ddots \quad \ddots \quad \ddots \quad \ddots \quad \ddots \\ \pi_k &= \pi_{k+1} q_0 + \pi_k q_1 + \pi_{k-1} q_2 + \cdots + \pi_1 q_k + \pi_0 q_k\end{aligned}$$

On démontre le résultat suivant :

PROPOSITION 3.3.3 *La chaîne de Markov $(Y_n)_{n \in \mathbb{N}}$ admet une probabilité invariante si et seulement si $\rho < 1$.*

Il est possible de calculer certaines caractéristiques de la probabilité invariante. En régime stationnaire, $Y - \mathbf{1}_{Y>0} + K$ a même loi que Y (avec des notations évidentes), d'où :

$$\mathbf{E}(Y) = \mathbf{E}(Y) - \mathbf{P}(Y > 0) + \mathbf{E}(K),$$

soit : $\mathbf{P}(Y = 0) = 1 - \rho$. Comme la chaîne est récurrente et irréductible, la probabilité invariante charge tous les points, donc $\mathbf{P}(Y = 0) > 0$ et donc nécessairement $\rho < 1$. On constate qu'il ne peut pas exister de probabilité invariante si $\rho = 1$.

En éllevant au carré et en prenant l'espérance on trouve (Y et K sont indépendantes) :

$$\mathbf{E}(Y^2) = \mathbf{E}(Y^2) + \mathbf{P}(Y > 0) + \mathbf{E}(K^2) - 2\mathbf{E}(Y) + 2\mathbf{E}(Y)\mathbf{E}(K) - 2\mathbf{P}(Y > 0)\mathbf{E}(K).$$

Or $\mathbf{E}(K)$ et $\mathbf{P}(Y > 0)$ valent ρ . Quant à $\mathbf{E}(K^2)$, on le déduit de :

$$\mathbf{E}(K(K-1)) = \int_0^\infty \sum_{k \in \mathbb{N}} k(k-1) e^{-\lambda t} \frac{(\lambda t)^k}{k!} d\mathbf{P}_\sigma(t) = \int_0^\infty \lambda^2 t^2 d\mathbf{P}_\sigma(t) = \lambda^2 \mathbf{E}(\sigma^2).$$

D'où $\mathbf{E}(K^2) = \mathbf{E}(K(K-1)) + \mathbf{E}(K) = \lambda^2 \mathbf{E}(\sigma^2) + \rho$ et

$$\mathbf{E}(Y) = \rho + \frac{\lambda^2 \mathbf{E}(\sigma^2)}{2(1-\rho)}.$$

Il est possible aussi de calculer la fonction génératrice g de la loi invariante π en fonction de celle de la loi des K_n (notée h) : en multipliant la j -ième équation

du système (3.2) par s^{j+1} et en sommant en diagonale (à faire en exercice), on obtient

$$g(s) = \frac{h(s)(s-1)}{s-h(s)} \pi_0.$$

En faisant tendre s vers 1, on obtient $\pi_0 = 1 - \rho$ ($g(1) = h(1) = 1$ et $h'(1) = E(K)$).

Il reste à remarquer que h s'exprime facilement en fonction de la transformée de Laplace \mathcal{L}_σ de la loi des services (utiliser (3.1) et permutez somme et intégrale) :

$$h(s) = \sum_{k=0}^{\infty} s^k q_k = \int_0^{\infty} e^{-\lambda(1-s)t} d\mathbf{P}_\sigma(t) = \mathcal{L}_\sigma(\lambda - \lambda s).$$

Finalement, lorsque $\rho < 1$, la fonction génératrice de la loi invariante s'écrit :

$$g(s) = \frac{(s-1)\mathcal{L}_\sigma(\lambda - \lambda s)}{s - \mathcal{L}_\sigma(\lambda - \lambda s)} (1 - \rho).$$

Cherchons à identifier le temps de séjour moyen R (attente + service) d'un client en régime stationnaire. Les personnes présentes dans le système au départ d'un client sont celles qui sont arrivées durant son temps de séjour, donc si ce temps de séjour vaut t , le nombre de clients pendant ce temps suit une loi de Poisson de paramètre λt et on peut écrire :

$$\mathbf{P}(Y = k) = \int_0^{\infty} e^{-\lambda t} \frac{(\lambda t)^k}{k!} d\mathbf{P}_R(t)$$

d'où en sommant sur k , $E(Y) = \lambda E(R)$. On retrouve la formule de LITTLE vue dans le cas M/M/s.

La théorie des processus de renouvellement permet de montrer que le comportement du processus $(X_t)_{t \geq 0}$ est le même que celui de la chaîne $(Y_n)_{n \geq 0}$. Dans le cas stationnaire, le processus a la même loi invariante que la chaîne.

Nous avons étudié en détail le cas markovien M/M/. dans lequel nous avons pu faire explicitement la plupart des calculs de lois. Dans le cas M/GI/1 nous avons mis en évidence une chaîne de Markov sous-jacente qui nous a permis, là aussi, de faire des calculs explicites de lois. Le cas GI/M/1 se traite de la même manière. On peut mettre en évidence une chaîne de Markov, cette fois non pas aux instants de départ du système, mais aux instants d'arrivées et faire des calculs sur cette chaîne.

Dans le cas général GI/GI/1, les techniques markoviennes ne fonctionnent plus. Il faut faire appel aux techniques liées aux processus de renouvellement. On démontre qu'il existe un régime stationnaire lorsque le produit du nombre moyen d'arrivées par unité de temps par le temps moyen de service est strictement plus petit que 1. On démontre que la formule de LITTLE est encore vraie. On montre que, en régime stationnaire, le flot des sorties du système a la même loi que le flot des entrées.

Chapitre 4

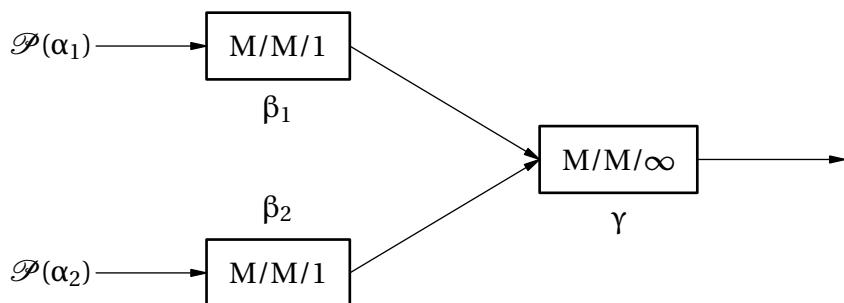
Réseaux de files d'attente

Nous passons maintenant à l'étude des réseaux de files d'attente. Cette étude va évidemment être plus compliquée que celle d'une seule file d'attente. Pour pouvoir utiliser les processus markoviens de sauts, nous ne considérerons que des réseaux alimentés par des flux extérieurs poissonniens et dont les lois de service sont exponentielles. Dans ce cas nous pourrons expliciter les probabilités invariantes.

4.1 Exemples simples

4.1.1 Un système série–parallèle

Considérons un réseau de trois files d'attente. Les files 1 et 2 sont du type $M/M/1$. Les paramètres de service de ces files sont β_1 et β_2 . Les entrées dans ces files se font selon des processus de Poisson indépendants de paramètres α_1 et α_2 . Les flots de sortie des files 1 et 2 se réunissent et constituent les entrées de la file 3. La file 3 est du type $M/M/\infty$. Le paramètre de service de la file 3 est γ . La sortie de cette file se fait vers l'extérieur.



Notons $N_1(t)$, $N_2(t)$ et $N_3(t)$ les nombres de clients présents (en service et en attente) dans chacune des unités. Les processus $N_1(t)$ et $N_2(t)$ sont des processus markoviens indépendants. Si α_1/β_1 et α_2/β_2 sont tous deux strictement plus petits que 1, un régime stationnaire s'établit dans les files 1 et 2. Il sort de ces files deux processus de Poisson de paramètres α_1 et α_2 (théorème de BURKE), le flot d'entrée dans la file 3 est donc un processus de Poisson de paramètre $\alpha_1 + \alpha_2$ (cf. théorème 2.2.10). Le processus $N_3(t)$ est également markovien et admet toujours un régime stationnaire. Il sort de cette file un flot poissonien de paramètre $\alpha_1 + \alpha_2$. A priori le processus $N_3(t)$ n'est pas indépendant de $N_1(t)$ et $N_2(t)$. Nous allons voir que c'est le cas en régime stationnaire.

Notons π_1 , π_2 et π_3 les probabilités invariantes de N_1 , N_2 et N_3 . Nous avons :

$$\begin{aligned}\pi_1(n) &= \left(1 - \frac{\alpha_1}{\beta_1}\right) \left(\frac{\alpha_1}{\beta_1}\right)^n \quad \pi_2(n) = \left(1 - \frac{\alpha_2}{\beta_2}\right) \left(\frac{\alpha_2}{\beta_2}\right)^n \\ \pi_3(n) &= \frac{1}{n!} \left(\frac{\alpha_1 + \alpha_2}{\gamma}\right)^n \exp\left(-\frac{\alpha_1 + \alpha_2}{\gamma}\right).\end{aligned}$$

Si on note $N(t)$ le processus de sauts à valeurs dans \mathbf{N}^3 de composantes $N_1(t)$, $N_2(t)$, $N_3(t)$, ce processus est markovien et son générateur est donné par :

$$\begin{aligned}A((n_1, n_2, n_3), (n_1 + 1, n_2, n_3)) &= \alpha_1 \\ A((n_1, n_2, n_3), (n_1 - 1, n_2, n_3 + 1)) &= \beta_1 \mathbf{1}_{n_1 > 0} \\ A((n_1, n_2, n_3), (n_1, n_2 + 1, n_3)) &= \alpha_2 \\ A((n_1, n_2, n_3), (n_1, n_2 - 1, n_3 + 1)) &= \beta_2 \mathbf{1}_{n_2 > 0} \\ A((n_1, n_2, n_3), (n_1, n_2, n_3 - 1)) &= \gamma n_3 \\ A((n_1, n_2, n_3), (n_1, n_2, n_3)) &= -(\alpha_1 + \beta_1 \mathbf{1}_{n_1 > 0} + \alpha_2 + \beta_2 \mathbf{1}_{n_2 > 0} + \gamma n_3).\end{aligned}$$

Les autres termes du générateur sont nuls.

On constate que le processus $N(t)$ est irréductible. Il n'admet pas de mesure réversible (pourquoi?). Les équations d'invariance s'écrivent $\pi A = 0$, soit pour tout triplet (n_1, n_2, n_3) de \mathbf{N}^3 :

$$\begin{aligned}\pi(n_1, n_2, n_3) (\alpha_1 + \beta_1 \mathbf{1}_{n_1 > 0} + \alpha_2 + \beta_2 \mathbf{1}_{n_2 > 0} + \gamma n_3) \\ = \pi(n_1 - 1, n_2, n_3) \alpha_1 \mathbf{1}_{n_1 > 0} + \pi(n_1 + 1, n_2, n_3 - 1) \beta_1 \mathbf{1}_{n_3 > 0} + \pi(n_1, n_2 - 1, n_3) \alpha_2 \mathbf{1}_{n_2 > 0} \\ + \pi(n_1, n_2 + 1, n_3 - 1) \beta_2 \mathbf{1}_{n_3 > 0} + \pi(n_1, n_2, n_3 + 1) \gamma (n_3 + 1).\end{aligned}$$

On vérifie que la probabilité

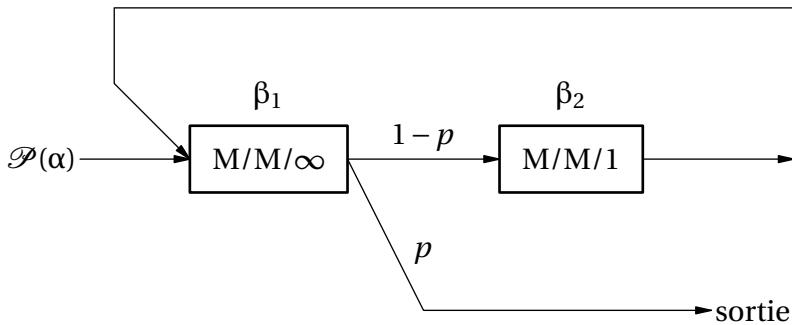
$$\pi(n_1, n_2, n_3) = \pi_1(n_1) \pi_2(n_2) \pi_3(n_3)$$

est invariante.

4.1.2 Un système avec boucle

Considérons un réseau (voir schéma page suivante) formé de deux files 1 (de type $M/M/\infty$) et 2 (de type $M/M/1$). Les deux unités fournissent des services indépendants dont les durées suivent des lois exponentielles de paramètres β_1 et β_2 .

Les clients se présentent à l'entrée suivant un flot de Poisson de paramètre α et entrent dans le file 1. À la sortie de la file 1, chaque client a la probabilité p de quitter le réseau et la probabilité $1 - p$ de se diriger vers l'unité 2. Les choix de directions sont indépendants entre eux et indépendants des services et des arrivées. À la sortie de l'unité 2, le client rentre dans la file 1.



Notons $N_1(t)$ et $N_2(t)$ le nombre de clients présents au temps t dans les unités 1 et 2. Le processus $N(t) = (N_1(t), N_2(t))$ est un processus markovien de sauts à valeurs dans \mathbb{N}^2 . Son générateur est donné par :

$$\begin{aligned} A((n_1, n_2), (n_1 + 1, n_2)) &= \alpha \\ A((n_1, n_2), (n_1 - 1, n_2)) &= pn_1\beta_1 \\ A((n_1, n_2), (n_1 - 1, n_2 + 1)) &= (1 - p)n_1\beta_1 \\ A((n_1, n_2), (n_1 + 1, n_2 - 1)) &= \beta_2 \mathbf{1}_{n_2 > 0} \\ A((n_1, n_2), (n_1, n_2)) &= -(\alpha + n_1\beta_1 + \beta_2 \mathbf{1}_{n_2 > 0}). \end{aligned}$$

Les autres termes du générateur sont nuls.

Ce processus est irréductible. Pour trouver une probabilité invariante, supposons que l'on soit dans un état stationnaire et évaluons les flots d'entrée moyens λ_1 et λ_2 dans chaque unité de service. Nous obtenons :

$$\begin{aligned} \lambda_1 &= \alpha + \lambda_2 \\ \lambda_2 &= (1 - p)\lambda_1. \end{aligned}$$

D'où $\lambda_1 = \frac{\alpha}{p}$ et $\lambda_2 = \frac{\alpha(1-p)}{p}$.

Ces flots ne sont pas poissonniens et les processus $N_1(t)$ et $N_2(t)$ ne sont pas markoviens. Cependant faisons comme si ils l'étaient, supposons $\lambda_2 < \beta_2$ pour

qu'ils admettent un régime stationnaire et posons :

$$\begin{aligned}\pi_1(n) &= \frac{1}{n!} \left(\frac{\lambda_1}{\beta_1} \right)^n \exp \left(-\frac{\lambda_1}{\beta_1} \right) & (\text{file M/M}/\infty) \\ \pi_2(n) &= \left(1 - \frac{\lambda_2}{\beta_2} \right) \left(\frac{\lambda_2}{\beta_2} \right)^n & (\text{file M/M}/1) \\ \pi(n_1, n_2) &= \pi_1(n_1) \pi_2(n_2).\end{aligned}$$

Les équations d'invariance s'écrivent pour tous (n_1, n_2) de \mathbf{N}^2 :

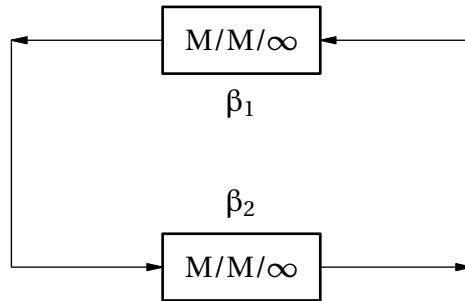
$$\begin{aligned}\pi(n_1, n_2)(\alpha + n_1\beta_1 + \beta_2 \mathbf{1}_{n_2>0}) &= \pi(n_1 - 1, n_2)\alpha \mathbf{1}_{n_1>0} + \pi(n_1 + 1, n_2)\beta_1(n_1 + 1)p \\ &\quad + \pi(n_1 + 1, n_2 - 1)\beta_1(n_1 + 1)(1 - p) \mathbf{1}_{n_2>0} + \pi(n_1 - 1, n_2 + 1)\beta_2 \mathbf{1}_{n_1>0}.\end{aligned}$$

On vérifie que la mesure $\pi(n_1, n_2) = \pi_1(n_1)\pi_2(n_2)$ est invariante.

Donc si $\lambda_2 < \beta_2$, c'est-à-dire si $\alpha < \beta_2 p / (1 - p)$, il existe un régime stationnaire décrit par la probabilité π .

4.1.3 Un système fermé

Considérons un réseau fermé constitué de deux unités 1 et 2 comportant chacune une infinité de serveurs. Les temps de service dans les deux unités sont indépendants et suivent des loi exponentielles de paramètres β_1 et β_2 . Il y a N clients dans le réseau. Les clients qui sortent de la station 1 entrent dans la station 2 et vice versa.



Notons $N_1(t)$ et $N_2(t)$ le nombre de clients présents au temps t dans les unités 1 et 2. On a $N_1(t) + N_2(t) = N$. Les processus N_1 et N_2 ne sont évidemment pas indépendants. Chaque processus de saut $N_i(t)$ ($i = 1$ ou 2) est markovien. Étant

irréductible sur un espace fini, il admet une unique probabilité invariante. Le générateur de N_1 est donné par :

$$\begin{aligned} A(n, n+1) &= \beta_2 (N - n), \\ A(n, n-1) &= \beta_1 n. \end{aligned}$$

On obtient comme probabilité invariante une loi binomiale $\mathcal{B}(N, \beta_2 / (\beta_1 + \beta_2))$, soit sur l'espace des couples (n_1, n_2) de nombres entiers positifs de somme égale à N :

$$\pi(n_1, n_2) = N! \frac{1}{n_1!} \left(\frac{\beta_2}{\beta_1 + \beta_2} \right)^{n_1} \frac{1}{n_2!} \left(\frac{\beta_1}{\beta_1 + \beta_2} \right)^{n_2}.$$

On constate que la probabilité invariante est là aussi de forme produit.

4.2 Réseaux de Jackson

Les exemples 1 et 2 du paragraphe précédent de réseaux ouverts ont donné des résultats semblables. Lorsqu'il y a un régime stationnaire, ce régime est décrit par une probabilité produit. Nous allons étudier dans ce paragraphe une classe de réseaux qui ont la même propriété.

Nous considérons K stations. Chaque station a un seul serveur et le temps de service à la station i est une variable aléatoire exponentielle de paramètre μ_i . Chaque station i est alimentée par un flot poissonnien exogène de paramètre α_i et par des flots qui viennent des autres stations du réseau et éventuellement aussi de i . Tous les flots exogènes et les temps de service sont indépendants. À la sortie de chaque station, un client soit sort du réseau, soit retourne dans une station du réseau. Ce choix se fait au hasard, indépendamment de toutes les autres variables aléatoires qui interviennent; à la sortie de la station i , il y a une probabilité r_{ij} ($1 \leq j \leq K$) d'aller à la station j et une probabilité r_i de sortir du réseau ($r_i + \sum_{j=1}^K r_{ij} = 1$). Un tel réseau est appelé réseau de Jackson.

On note $X(t) = (X_1(t), \dots, X_K(t))$ l'état du réseau à l'instant t . $X_i(t)$ représente le nombre de clients présents dans la station i au temps t . Le processus $(X(t))_{t \geq 0}$ est un processus markovien de sauts à valeurs dans \mathbb{N}^K .

On note $n = (n_1, \dots, n_K)$ et $e_i = (0, \dots, 0, 1, 0, \dots, 0)$ le i -ème vecteur de base de \mathbb{R}^K . Le générateur de ce processus est donné par :

$$\begin{aligned} A(n, n + e_i) &= \alpha_i, \\ A(n, n - e_i) &= \mu_i r_i \mathbf{1}_{n_i > 0}, \\ A(n, n - e_i + e_j) &= \mu_i r_{ij} \mathbf{1}_{n_i > 0}. \end{aligned}$$

Les termes diagonaux s'obtiennent en remarquant que la somme des termes d'une même ligne est nulle :

$$A(n, n) = - \sum_{i=1}^K (\alpha_i + \mu_i \mathbf{1}_{n_i > 0}).$$

On constate que le processus $X(t)$ est irréductible sur \mathbf{N}^K si :

- il est ouvert, c'est-à-dire alimenté extérieurement : il existe j tel que $\alpha_j > 0$;
- il est sans capture, c'est-à-dire si à partir de toute station il existe une possibilité de sortir définitivement du réseau : $\forall i \exists (i_1, i_2, \dots, i_k, j) r_{ii_1} r_{i_1 i_2} \dots r_{i_k j} r_j > 0$.

L'équation d'invariance $\pi A = 0$ s'écrit :

$$\begin{aligned} \pi(n) \sum_{i=1}^K (\alpha_i + \mu_i \mathbf{1}_{n_i > 0}) &= \sum_{i=1}^K (\pi(n + e_i) \mu_i r_i + \pi(n - e_i) \alpha_i \mathbf{1}_{n_i > 0}) \\ &+ \sum_{i=1}^K \sum_{j=1}^K \pi(n + e_i - e_j) \mu_i r_{ij}. \end{aligned}$$

Pour deviner une solution π à ce système, nous allons étudier les flots de clients à travers le réseau. Supposons que le réseau fonctionne en régime stationnaire. Appelons λ_i le nombre moyen de clients qui entrent dans la station i par unité de temps. Comme on est en équilibre, c'est aussi le nombre moyen de clients qui sortent de la station i par unité de temps. On a alors l'équation suivante, dite *équation du trafic* :

$$\forall i (1 \leq i \leq K), \quad \lambda_i = \alpha_i + \sum_{j=1}^K \lambda_j r_{ji}.$$

Ceci constitue un système de K équations à K inconnues. On peut démontrer que sous les hypothèses d'un réseau sans capture, ce système admet une unique solution strictement positive. Posons alors $\rho_i = \frac{\lambda_i}{\mu_i}$ et supposons que tous les nombres ρ_i sont strictement plus petits que 1. Par analogie avec les exemples vus dans le paragraphe précédent posons : $\pi_i(n_i) = (1 - \rho_i)(\rho_i)^{n_i}$ (probabilité invariante d'une file M/M/1) et $\pi(n) = \prod_{i=1}^K \pi_i(n_i)$. Nous allons démontrer que π est bien une probabilité invariante pour $X(t)$.

Pour cela on peut faire une vérification en remplaçant dans l'équation d'invariance l'expression de π . Il est plus adroit d'utiliser la notion de réversibilité avec la proposition suivante.

PROPOSITION 4.2.1 *Soit $(X_t)_{t \geq 0}$ un processus markovien de sauts à valeurs dans E de générateur $A(., .)$. Soient π une distribution de probabilité sur E strictement*

positive et $B(.,.)$ la matrice définie par :

$$\forall (i, j) \in E^2, \quad \pi(i)A(i, j) = \pi(j)B(j, i).$$

Si pour tout i dans E on a l'égalité :

$$\sum_{\substack{j \in E \\ j \neq i}} B(i, j) = \sum_{\substack{j \in E \\ j \neq i}} A(i, j),$$

alors π est une probabilité invariante pour $(X_t)_{t \geq 0}$ et pour tout T , $B(.,.)$ est le générateur du processus retourné $(X_{T-t})_{0 \leq t \leq T}$.

Démonstration. L'équation $\sum_{j \neq i} B(i, j) = -A(i, i)$ n'est rien d'autre que l'équation d'invariance. Donc π est bien une probabilité invariante. Le reste de la proposition a été vu dans la démonstration du théorème de BURKE (proposition 3.2.7). \square

Dans le cas des réseaux de Jackson, calculons la matrice $B(.,.)$:

$$\begin{aligned} \pi(n)B(n, n + e_i) &= \pi(n + e_i)A(n + e_i, n) &= \pi(n + e_i)\mu_i r_i, \\ \pi(n)B(n, n - e_i) &= \pi(n - e_i)A(n - e_i, n) &= \pi(n - e_i)\alpha_i \mathbf{1}_{n_i > 0}, \\ \pi(n)B(n, n - e_i + e_j) &= \pi(n - e_i + e_j)A(n - e_i + e_j, n) = \pi(n - e_i + e_j)\mu_j r_{ji} \mathbf{1}_{n_i > 0}. \end{aligned}$$

Par définition de π :

$$\pi(n) = \prod_{i=1}^K \pi_i(n_i) = \prod_{i=1}^K (1 - \rho_i)(\rho_i)^{n_i}$$

$$\text{d'où } \frac{\pi(n + e_i)}{\pi(n)} = \rho_i = \frac{\lambda_i}{\mu_i}, \quad \frac{\pi(n - e_i)}{\pi(n)} = \frac{1}{\rho_i}, \quad \frac{\pi(n - e_i + e_j)}{\pi(n)} = \frac{\rho_j}{\rho_i} \quad \text{et :}$$

$$\begin{aligned} B(n, n + e_i) &= \lambda_i r_i, \\ B(n, n - e_i) &= \frac{\alpha_i}{\rho_i} \mathbf{1}_{n_i > 0}, \\ B(n, n - e_i + e_j) &= r_{ji} \frac{\lambda_j}{\rho_i} \mathbf{1}_{n_i > 0}. \end{aligned}$$

On en déduit :

$$\begin{aligned} \sum_{\substack{n' \in E \\ n' \neq n}} B(n, n') &= \sum_{i=1}^K \left(\lambda_i r_i + \frac{\alpha_i}{\rho_i} \mathbf{1}_{n_i > 0} + \frac{1}{\rho_i} \mathbf{1}_{n_i > 0} \sum_{j=1}^K \lambda_j r_{ji} \right) \\ &= \sum_{i=1}^K \left(\lambda_i r_i + \frac{\alpha_i \mu_i}{\lambda_i} \mathbf{1}_{n_i > 0} + \frac{\mu_i}{\lambda_i} \mathbf{1}_{n_i > 0} (\lambda_i - \alpha_i) \right) \quad (\text{équation du trafic}) \\ &= \sum_{i=1}^K (\lambda_i r_i + \mu_i \mathbf{1}_{n_i > 0}). \end{aligned}$$

D'autre part on a vu que :

$$\sum_{\substack{n' \in E \\ n' \neq n}} A(n, n') = -A(n, n) = \sum_{i=1}^K (\alpha_i + \mu_i \mathbf{1}_{n_i > 0})$$

En sommant sur i les équations du trafic on trouve que : $\sum_{i=1}^K \alpha_i = \sum_{i=1}^K \lambda_i r_i$. On constate que l'on a bien l'égalité : $\sum_{n' \neq n} A(n, n') = \sum_{n' \neq n} B(n, n')$.

Ceci démontre que π est une probabilité invariante pour le processus $X(t)$. On en déduit également que le processus retourné a comme générateur $B(.,.)$. En régime stationnaire, ce processus est un processus de Markov associé à un réseau de Jackson de paramètres $(\alpha_i)', (\mu_i)', (r_i)'$ et $(r_{ij})'$ tels que :

$$(\alpha_i)' = \lambda_i r_i \quad (\mu_i)'(r_i)' = \frac{\alpha_i}{\rho_i} \quad (\mu_i)'(r_{ij})' = r_{ij} \frac{\lambda_j}{\rho_i}.$$

Les instants de sorties définitives du réseau initial sont les instants d'entrée du réseau correspondant au processus retourné. Ce sont donc des flots poissonniers indépendants entre eux de paramètres $\lambda_i r_i$.

Nous avons démontré le théorème suivant :

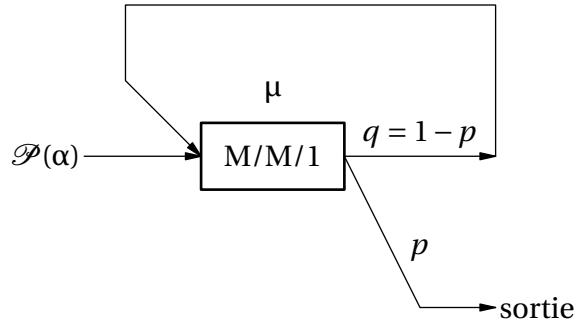
THÉORÈME 4.2.2 (Jackson) *Dans un réseau de Jackson ouvert et sans capture, le processus markovien de sauts $X(t)$ correspondant aux nombres de clients présents dans chaque station du réseau à l'instant t est irréductible et admet une probabilité invariante π de forme produit : $\pi(n) = \prod_i \pi_i(n_i)$. Les probabilités marginales $\pi_i(.)$ sont les probabilités invariantes correspondant à la station numéro i fonctionnant seule avec un flot extérieur poissonnien de paramètre λ_i . Ces paramètres λ_i sont calculés à l'aide des équations de trafic.*

Les flots de sortie du réseau sont des flots poissonniers indépendants.

Les flots d'entrée et de sortie du réseau sont poissonniers. En revanche les flots qui traversent les stations du réseau n'ont aucune raison d'être poissonniers.

Pour s'en convaincre étudions le réseau à une station suivant : la station a un flot d'entrée extérieur poissonnien de paramètre α , un temps de service exponentiel de paramètre μ et un serveur; à la sortie de la station chaque client sort du réseau avec probabilité p et retourne dans la file d'attente de la station avec probabilité q ($p + q = 1$).

Le flot moyen λ à travers la station satisfait : $\lambda = \alpha + q\lambda$, d'où $\lambda = \alpha/p$. Il y a un régime stationnaire si $\lambda < \mu$, c'est-à-dire si $\alpha < p\mu$. Ce régime stationnaire est décrit par la probabilité invariante : $m(n) = (1 - \rho)\rho^n$ avec $\rho = \lambda/\mu = \alpha/(p\mu)$.



On se place à l'instant d'arrivée d'un client dans la station. Notons X le nombre de clients présents dans la station juste avant cette arrivée, T_e le temps nécessaire pour que le prochain client venant de l'extérieur arrive, T_r le temps nécessaire pour qu'un client présent dans la station y revienne et T le temps qui s'écoulera jusqu'à l'arrivée du client suivant (extérieur ou recyclé).

$$\begin{aligned}
 \mathbf{P}(T > x) &= \mathbf{P}(T_e > x \cap T_r > x) \\
 &= \sum_{n=0}^{\infty} \mathbf{P}(T_e > x \cap T_r > x \mid X = n) \mathbf{P}(X = n) \\
 &= \sum_{n=0}^{\infty} \mathbf{P}(T_e > x) m(n) \mathbf{P}(T_r > x \mid X = n) \\
 &= \sum_{n=0}^{\infty} e^{-\alpha x} \left(1 - \frac{\alpha}{p\mu}\right) \left(\frac{\alpha}{p\mu}\right)^n \mathbf{P}(T_r > x \mid X = n).
 \end{aligned}$$

Calculons la probabilité $\mathbf{P}(T_r > x \mid X = n)$. T_r est l'instant de sortie du premier client recyclé, ou $+\infty$ si les $n+1$ clients présents quittent tous le réseau à la fin de leur service; décomposons donc l'événement $T_r > x$ selon le numéro du premier client recyclé :

$$\begin{aligned}
 \mathbf{P}(T_r > x \mid X = n) &= \sum_{j=1}^{n+1} \left(p^{j-1} q \mathbf{P}\left(\sum_{i=1}^j \sigma_i > x\right) \right) + p^{n+1} \\
 &= \sum_{j=1}^{n+1} \left(p^{j-1} q \int_x^{\infty} \mu e^{-\mu t} \frac{(\mu t)^{j-1}}{(j-1)!} dt \right) + p^{n+1} \quad (\text{cf. prop. 2.2.4}).
 \end{aligned}$$

On reporte ce résultat dans l'expression de $\mathbf{P}(T > x)$:

$$\begin{aligned}
 \mathbf{P}(T > x) &= e^{-\alpha x} \left(1 - \frac{\alpha}{p\mu}\right) \sum_{n=0}^{\infty} \left(\frac{\alpha}{p\mu}\right)^n \left(p^{n+1} + \sum_{j=1}^{n+1} p^{j-1} q \mu \int_x^{\infty} e^{-\mu t} \frac{(\mu t)^{j-1}}{(j-1)!} dt \right) \\
 &= e^{-\alpha x} \left(1 - \frac{\alpha}{p\mu}\right) \left(\frac{p}{1 - \frac{\alpha}{\mu}} + \sum_{n=0}^{\infty} \sum_{k=0}^n \left(\frac{\alpha}{p\mu}\right)^n p^k q \mu \int_x^{\infty} e^{-\mu t} \frac{(\mu t)^k}{k!} dt \right)
 \end{aligned}$$

soit en appliquant Fubini pour permuter les trois sommes (fonction positive)

$$\begin{aligned}
 &= e^{-\alpha x} \left(1 - \frac{\alpha}{p\mu} \right) \left(\frac{p}{1 - \frac{\alpha}{\mu}} + q\mu \int_x^\infty e^{-\mu t} \sum_{k=0}^{\infty} p^k \frac{(\mu t)^k}{k!} \sum_{n=k}^{\infty} \left(\frac{\alpha}{p\mu} \right)^n dt \right) \\
 &= e^{-\alpha x} \left(1 - \frac{\alpha}{p\mu} \right) \left(\frac{p}{1 - \frac{\alpha}{\mu}} + q\mu \int_x^\infty e^{-\mu t} \sum_{k=0}^{\infty} p^k \frac{(\mu t)^k}{k!} \left(\frac{\alpha}{p\mu} \right)^k \frac{1}{1 - \frac{\alpha}{p\mu}} dt \right) \\
 &= e^{-\alpha x} \left(1 - \frac{\alpha}{p\mu} \right) \left(\frac{p}{1 - \frac{\alpha}{\mu}} + q\mu \int_x^\infty e^{-\mu t} \sum_{k=0}^{\infty} \frac{(\alpha t)^k}{k!} \frac{1}{1 - \frac{\alpha}{p\mu}} dt \right) \\
 &= e^{-\alpha x} \left(1 - \frac{\alpha}{p\mu} \right) \left(\frac{p}{1 - \frac{\alpha}{\mu}} + q\mu \int_x^\infty e^{-\mu t} e^{\alpha t} \frac{1}{1 - \frac{\alpha}{p\mu}} dt \right) \\
 &= e^{-\alpha x} \left(1 - \frac{\alpha}{p\mu} \right) \left(\frac{p}{1 - \frac{\alpha}{\mu}} + q\mu \frac{e^{(\alpha - \mu)x}}{\mu - \alpha} \frac{1}{1 - \frac{\alpha}{p\mu}} \right) \\
 &= \frac{p\mu - \alpha}{\mu - \alpha} e^{-\alpha x} + \frac{q\mu}{\mu - \alpha} e^{-\mu x}.
 \end{aligned}$$

On constate que T ne suit pas une loi exponentielle et donc le flot n'est pas poissonnien. On vérifie bien que : $E(T) = 1/\lambda$.

On peut étudier de la même manière des réseaux de Jackson plus généraux avec des stations à plusieurs serveurs. On peut également étudier des réseaux fermés. On peut aussi avoir le cas de plusieurs types de clients qui évoluent dans des réseaux de Jackson. Tout ceci se traite par les mêmes méthodes et on trouve des probabilités produits pour décrire les régimes stationnaires.

Chapitre 5

Fiabilité de systèmes simples

La fiabilité d'un dispositif est une caractéristique de ce dispositif exprimée par la probabilité qu'il accomplisse une fonction requise sans défaillance, dans des conditions données, pendant une durée donnée.

Les études de fiabilité se sont beaucoup développées ces vingt dernières années. Les problèmes de sécurité (industries aéronautique et spatiale, électro-nucléaire) ont été un moteur de ce développement. Cependant ces techniques se répandent dans toute l'industrie car elles permettent de modéliser dans quelles conditions un dispositif fonctionne correctement ou se trouve en panne. Elles permettent ainsi d'améliorer les caractéristiques du dispositif, de donner une base rationnelle aux contrats liant les différents acteurs intervenant sur le dispositif. Elles donnent des outils pour une gestion des opérations de maintenance.

Dans ce chapitre nous allons donner des bases des calculs de fiabilité sur des systèmes simples où des calculs analytiques explicites sont possibles. Cela nous permettra de nous familiariser avec ces notions avant d'aborder l'étude de systèmes complexes dans le prochain chapitre. L'outil mathématique essentiel qui est utilisé est le calcul des probabilités et la statistique. Pour l'étude du temps de fonctionnement d'un système simple on utilise des techniques probabilistes et statistiques élémentaires. Quand on s'intéresse à un système composé d'éléments réparables et qu'alternent des périodes de fonctionnement et des périodes de réparation, on fait intervenir les processus de Markov et les processus de renouvellement.

5.1 Définitions générales

Quand on étudie un système, on définit en fonction du temps deux grandeurs : la fiabilité et la disponibilité.

DÉFINITION 5.1.1 *On appelle fiabilité $R(t)$ d'un système S devant accomplir une mission dans des conditions données la probabilité que le système S n'ait eu aucune défaillance entre les instants 0 et t . On appelle disponibilité $A(t)$ la probabilité que le système S fonctionne à l'instant t . On a donc :*

$$R(t) = \mathbf{P}(S \text{ non défaillant sur } [0, t])$$

$$A(t) = \mathbf{P}(S \text{ non défaillant à l'instant } t).$$

La fiabilité est une fonction qui décroît de 1 à 0 lorsque le temps varie de 0 à l'infini. Dans certaines conditions la disponibilité d'un système réparable en fonctionnement permanent se stabilise lorsque t grandit sur une valeur non nulle. En revanche la fiabilité tend toujours vers 0. Lorsque le système est réparable on définit également une autre notion : la maintenabilité.

DÉFINITION 5.1.2 *On appelle maintenabilité $M(t)$ d'un système réparable S la probabilité pour que le système S soit réparé avant l'instant t sachant qu'il est défaillant à l'instant 0. On a donc :*

$$M(t) = 1 - \mathbf{P}(S \text{ non réparé sur } [0, t]).$$

La maintenabilité est une fonction croissante de 0 à 1 lorsque t varie de 0 à l'infini.

On s'intéressera à diverses grandeurs moyennes dont les sigles sont les suivants :

- MTTF (mean time to failure) : durée moyenne de bon fonctionnement d'un système en état de marche à l'instant initial,
- MTTR (mean time to repair) : durée moyenne de réparation d'un système en panne à l'instant initial,
- MUT (mean up time) : durée moyenne de bon fonctionnement d'un système réparable en régime stationnaire,
- MDT (mean down time) : durée moyenne de défaillance d'un système réparable en régime stationnaire,
- MTBF (mean time between failure) : intervalle de temps moyen séparant deux défaillances consécutives d'un système réparable en régime stationnaire.

Ces sigles sont classiques, mais ne sont pas toujours utilisés avec exactement la même signification. A priori MUT est différent de MTTF car lorsqu'un système est remis en état après une défaillance tous ses composants n'ont pas été remis à neuf. De même MTTR est différent de MDT car MTTR concerne la première réparation d'un système alors que MDT représente le temps moyen de réparation du système en régime stationnaire. On a évidemment la relation : $MTBF = MUT + MDT$.

Si T représente la variable aléatoire mesurant la durée de bon fonctionnement du système, on peut écrire : $R(t) = \mathbf{P}(T > t)$

La fiabilité $R(t)$ vaut donc 1 moins la fonction de répartition de T . Nous supposons dans la suite que T admet une densité f par rapport à la mesure de Lebesgue sur \mathbf{R}_+ , presque partout continue. Cette densité vaut presque partout :

$$f(t) = -\frac{dR}{dt}(t).$$

On sait que si T admet un premier moment alors :

$$\text{MTTF} = \mathbf{E}(T) = \int_0^\infty R(t) dt.$$

Si U est la variable aléatoire représentant la durée de réparation du système, on a : $M(t) = \mathbf{P}(U \leq t)$.

$M(t)$ est la fonction de répartition de U . Nous supposerons que U admet une densité g par rapport à la mesure de Lebesgue sur \mathbf{R}_+ , presque partout continue. Cette densité vaut presque partout :

$$g(t) = \frac{dM}{dt}(t).$$

Si le temps moyen de réparation est fini, alors on a alors :

$$\text{MTTR} = \int_0^\infty (1 - M(t)) dt.$$

On introduit également les taux instantanés de défaillance et de réparation $\lambda(t)$ et $\mu(t)$:

$$\begin{aligned} \lambda(t) &= \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \mathbf{P}(t < T \leq t + \Delta t \mid T > t), \\ \mu(t) &= \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \mathbf{P}(t < U \leq t + \Delta t \mid U > t). \end{aligned}$$

En tout point de continuité des densités on a les relations :

$$\begin{aligned} \lambda(t) &= \frac{f(t)}{R(t)} = \frac{-\frac{dR}{dt}(t)}{R(t)} \\ \mu(t) &= \frac{g(t)}{1 - M(t)} = \frac{\frac{dM}{dt}(t)}{1 - M(t)} \end{aligned}$$

d'où :

$$R(t) = \exp\left(-\int_0^t \lambda(u) du\right) \quad \text{et} \quad M(t) = 1 - \exp\left(-\int_0^t \mu(u) du\right).$$

La fiabilité (resp. la maintenabilité) s'exprime en fonction du taux instantané de défaillance (resp. de réparation) et vice versa.

5.2 Test de Laplace

Le cas le plus simple à étudier est celui où les taux instantanés de défaillance et de réparation sont constants. Le test de Laplace est souvent utilisé pour vérifier ces hypothèses. Il repose sur le résultat de la proposition 2.2.7 : négligeons la durée des réparations et supposons le taux instantané de défaillance $\lambda(t)$ constant, le processus des instants de panne est alors un processus de Poisson de paramètre λ . Notons (T_i) ses instants de saut et considérons la variable aléatoire

$$U_n = \frac{1}{n} \sum_{i=1}^n T_i.$$

La loi conditionnelle de U_n sachant $N_t = n$ est la loi de

$$V_n = \frac{1}{n} \sum_{i=1}^n X_i,$$

où les X_i sont indépendantes de loi uniforme sur $[0, t]$ (proposition 2.2.7).

Pour n assez grand on approxime la loi de V_n par le théorème central limite : $\frac{V_n - E(V_n)}{\sigma(V_n)}$ suit approximativement une loi normale réduite $\mathcal{N}(0, 1)$. On a (loi uniforme)

$$\forall i, \quad E(X_i) = \frac{t}{2}, \quad \text{Var}(X_i) = \frac{t^2}{12}, \quad \text{d'où} \quad E(V_n) = \frac{t}{2}, \quad \text{Var}(V_n) = \frac{t^2}{12n}.$$

Pratiquement, on observe les n premiers instants de panne (on a alors $N(t) = n$), la variable test sera

$$W_n = \frac{U_n - E(U_n | N_t = n)}{\sigma(U_n | N_t = n)} = \frac{\frac{1}{n} \sum_{i=1}^n T_i - \frac{t}{2}}{\frac{t}{\sqrt{12n}}} = \frac{\frac{1}{n} \sum_{i=1}^n T_i - \frac{t}{2}}{t} \sqrt{12n},$$

car $E(U_n | N_t = n) = E(V_n)$ et $\sigma(U_n | N_t = n) = \sigma(V_n)$. Si l'hypothèse $\lambda(t) = \lambda$ (constante) est vérifiée, W_n doit suivre (approximativement pour n assez grand) une loi normale réduite.

Si au contraire $\lambda(t)$ est décroissante (fiabilité croissante) les instants de panne seront plus concentrés autour de 0 (se souvenir que $E(T_{i+1} - T_i) = \lambda^{-1}$, donc λ représente le nombre moyen de pannes par unité de temps).

On prendra donc comme région de rejet $\{W_n < -c\}$ pour tester

$$H_0 = \{\lambda(t) = \lambda\} \quad \text{contre} \quad \{\lambda(t) \text{ décroissante}\}$$

et $\{W_n > c\}$ pour tester

$$H_0 = \{\lambda(t) = \lambda\} \quad \text{contre} \quad \{\lambda(t) \text{ croissante}\}$$

où c est lu dans la table de Gauss, par exemple :

- $c = 1.96$ au seuil de confiance $\alpha = 0.975$,
- $c = 1.645$ au seuil de confiance $\alpha = 0.95$.

5.3 Exemples de lois utiles en fiabilité

5.3.1 Loi exponentielle

On rencontre souvent en fiabilité la loi exponentielle. Si le paramètre est α on a :

$$f(t) = \alpha e^{-\alpha t} \mathbf{1}_{t \geq 0}, \quad R(t) = e^{-\alpha t} \mathbf{1}_{t \geq 0}, \quad \lambda(t) = \alpha, \quad \text{MTTF} = \frac{1}{\alpha}.$$

5.3.2 Loi $\Gamma(\alpha, \beta)$

On rencontre également la loi $\Gamma(\alpha, \beta)$ dont la densité est

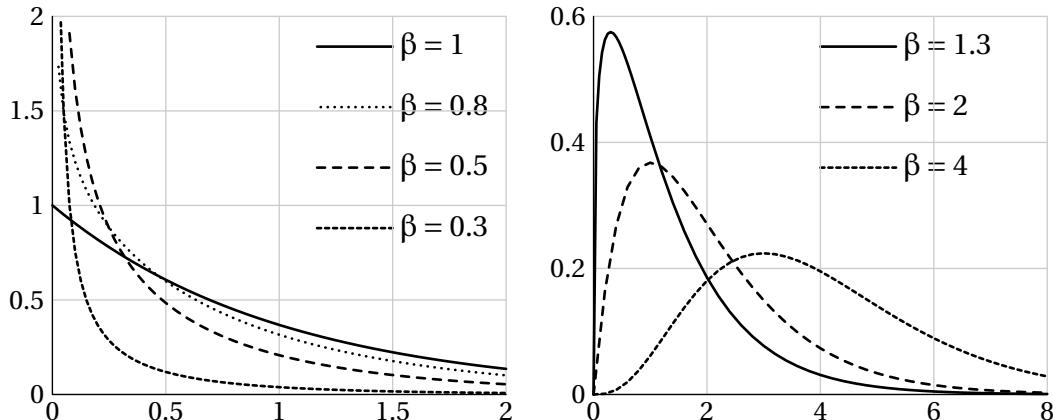
$$f(t) = \frac{\alpha}{\Gamma(\beta)} (\alpha t)^{\beta-1} e^{-\alpha t} \mathbf{1}_{t \geq 0} \quad \text{avec } \alpha > 0, \beta > 0.$$

On a alors :

$$\text{MTTF} = \frac{\beta}{\alpha}.$$

Quand $\beta = 1$ on retrouve la densité exponentielle. Lorsque $\beta \neq 1$ il n'y a pas de formule explicite pour la fiabilité et le taux de défaillance.

L'allure des densités de probabilité, selon les valeurs de β , est la suivante :



5.3.3 Loi de Weibull

La loi de Weibull, qui admet trois paramètres, est souvent utilisée. X suit une loi de Weibull de paramètres η, β, γ , si et seulement si $\left(\frac{X-\gamma}{\eta}\right)^\beta$ suit une loi exponentielle de paramètre 1. Sa densité vaut :

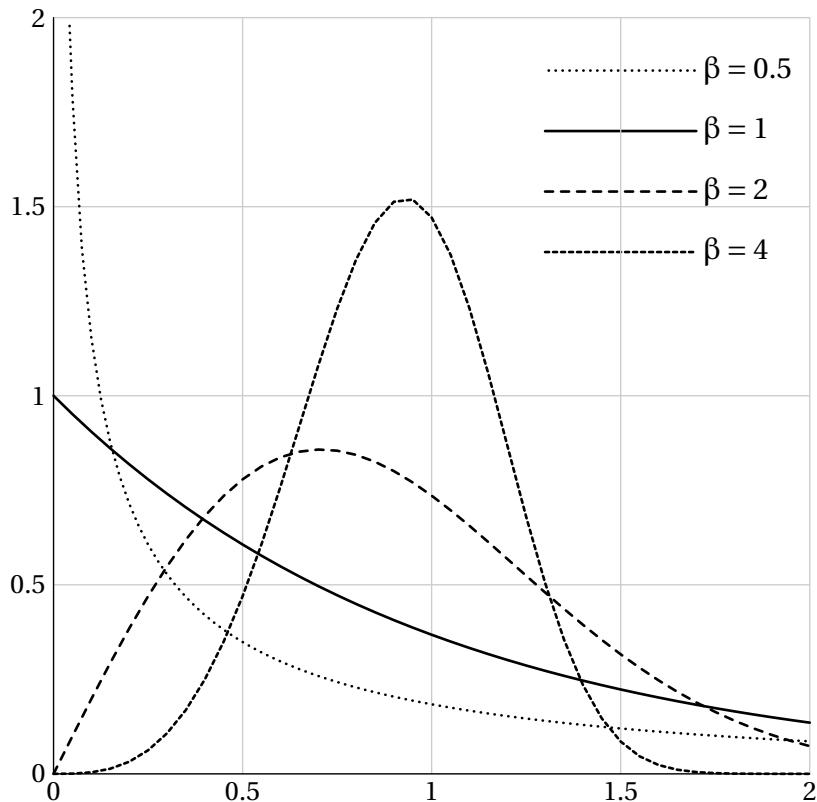
$$f(t) = \frac{\beta}{\eta} \left(\frac{t-\gamma}{\eta}\right)^{\beta-1} e^{-\left(\frac{t-\gamma}{\eta}\right)^\beta} \mathbf{1}_{t \geq \gamma} \quad \text{avec } \beta > 0, \eta > 0 \text{ et } \gamma \geq 0.$$

$$\forall t \geq \gamma, \quad R(t) = e^{-\left(\frac{t-\gamma}{\eta}\right)^\beta} \quad \text{et} \quad \lambda(t) = \frac{\beta}{\eta} \left(\frac{t-\gamma}{\eta}\right)^{\beta-1}.$$

On trouve :

$$\text{MTTF} = \gamma + \eta \Gamma(1 + \frac{1}{\beta}).$$

Le choix des paramètres donne une grande variété de comportements, c'est ce qui fait l'intérêt de cette loi. Voici l'allure des densités de probabilité, selon les valeurs de β , pour $\eta = 1$ et $\gamma = 0$:



5.3.4 Loi lognormale

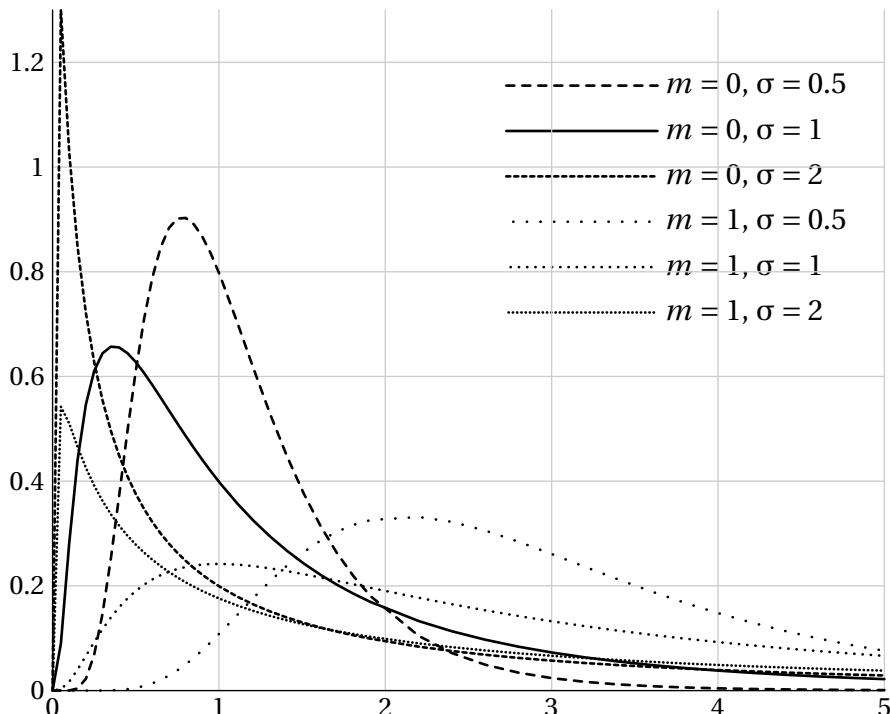
On dit que X suit une loi lognormale de paramètres m et σ si et seulement si $\ln X$ suit une loi normale $\mathcal{N}(m, \sigma)$, la loi lognormale a donc pour densité

$$f(t) = \frac{1}{t\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\ln t - m}{\sigma}\right)^2} \mathbf{1}_{t>0}.$$

Le taux de défaillance vaut 0 en 0, croît et passe par un maximum puis décroît en tendant vers 0 à l'infini. On a :

$$\text{MTTF} = \exp(m + \frac{\sigma^2}{2}).$$

Voici l'allure des densités de probabilité, pour $m = 0$ ou 1 et $\sigma = 0.5, 1$ ou 2 .



On utilise diverses lois pour modéliser la durée de bon fonctionnement ou de réparation d'un système. Le choix de la loi se fait, d'une part en cherchant à décrire le plus exactement possible la durée étudiée, d'autre part en tenant compte de la facilité de traitement qu'offrent les diverses lois. Ces deux contraintes sont souvent contradictoires et il faut faire une pondération raisonnable entre les deux. Quand on a décidé du type de loi utilisée, on estime les paramètres de la loi à l'aide des techniques classiques de statistique paramétrique, la plupart du temps à l'aide d'estimateurs du maximum de vraisemblance sur un échantillon de cette loi. On dispose souvent de données censurées, d'où l'utilisation d'estimateurs adaptés à ce type de données. Dans ce cours on suppose que cette démarche statistique a été effectuée et on s'intéresse au traitement probabiliste du modèle étudié.

5.4 Systèmes à un élément

Ce qui vient d'être dit en général s'applique évidemment à un système comportant un seul élément. Les études statistiques ont montré que la courbe représentant le taux de défaillance $\lambda(t)$ a souvent la forme d'une « courbe en baignoire ». Cette courbe comprend trois régions. La première région correspond à la période de jeunesse de l'élément au cours de laquelle le taux de défaillance décroît. La deuxième région correspond à la période de vie utile de l'élément au cours de laquelle le taux de défaillance est sensiblement constant. La dernière région correspond à la période de vieillissement pendant laquelle le taux de défaillance croît rapidement.

Pendant la période de vie utile le taux de défaillance est constant. Si on ne s'intéresse qu'à cette période, on peut supposer que le taux de défaillance est toujours constant et donc que le temps de défaillance suit une loi exponentielle. Si on suppose que le temps de réparation suit lui aussi une loi exponentielle, λ désignant le taux de défaillance et μ le taux de réparation, on a :

$$\text{MTTF} = \text{MUT} = \frac{1}{\lambda}, \quad \text{MTTR} = \text{MDT} = \frac{1}{\mu}.$$

Le processus $(X_t)_{t \geq 0}$ qui vaut 1 quand l'élément est en état de marche et 0 quand il est en panne au temps t est un processus markovien de saut. Ce processus a été étudié au chapitre 2 : il est ergodique et admet comme probabilité invariante :

$$m(0) = \frac{\lambda}{\lambda + \mu}, \quad m(1) = \frac{\mu}{\lambda + \mu}.$$

Donc $\lim_{t \rightarrow +\infty} A(t) = m(1) = \frac{\mu}{\lambda + \mu}$.

On a également calculé les probabilités de transition de ce processus en résolvant l'équation de Kolmogorov (cf. 2.4.3, p. 25). Cela donne la disponibilité si $A(0)$ vaut 1 :

$$(5.1) \quad A(t) = P_t(1, 1) = \frac{\mu}{\lambda + \mu} + \frac{\lambda}{\lambda + \mu} e^{-(\lambda + \mu)t}.$$

Si l'hypothèse « taux de défaillance constant » est raisonnable pendant la période de vie utile, l'hypothèse « taux de réparation constant » n'est pas souvent vérifiée. Elle est commode car elle permet d'utiliser les techniques markoviennes.

Considérons maintenant un processus $(X_t)_{t \geq 0}$ associé à un système formé d'un élément réparable. Comme plus haut X_t vaut 1 ou 0 suivant que l'élément est en état de marche ou non. Supposons que X_0 vaut 1, que les temps de bon fonctionnement sont des variables aléatoires indépendantes de même loi exponentielle de paramètre λ , que les temps de réparation sont des variables aléatoires

de même loi m et que les temps de fonctionnement et les temps de réparation sont indépendants. Si la loi m n'est pas exponentielle, le processus $(X_t)_{t \geq 0}$ n'est pas markovien. C'est un processus de sauts. Nous allons montrer que la disponibilité de ce processus admet des équations du type Kolmogorov :

$$\begin{aligned} A(t+h) &= \mathbf{P}(X(t+h) = 1) \\ &= \mathbf{P}(X(t+h) = 1 \mid X(t) = 1) \mathbf{P}(X(t) = 1) + \mathbf{P}(X(t+h) = 1 \cap X(t) = 0), \end{aligned}$$

soit en décomposant le second terme de la somme selon la durée u de réparation

$$= (1 - \lambda h + o(h))A(t) + \int_0^t A(t-u)(\lambda h + o(h))m(du).$$

On en déduit que $A(t)$ est dérivable et que :

$$\frac{dA}{dt}(t) = -\lambda A(t) + \lambda \int_0^t A(t-u)m(du).$$

Pour résoudre cette équation différentielle on peut prendre la transformée de Laplace des deux membres. Si on note $A^*(s)$ et $m^*(s)$ les transformées de Laplace de $A(t)$ et m

$$A^*(s) = \int_{\mathbf{R}_+} e^{-st} A(t) dt \quad \text{et} \quad m^*(s) = \int_{\mathbf{R}_+} e^{-st} m(dt),$$

on obtient (voir annexe 6.3.2) :

$$sA^*(s) - A(0) = -\lambda A^*(s) + \lambda m^*(s)A^*(s),$$

d'où : $A^*(s) = \frac{1}{s + \lambda - \lambda m^*(s)}$, on a donc démontré la proposition suivante :

PROPOSITION 5.4.1 *Si un système à un élément satisfait aux conditions suivantes :*

- *il fonctionne à l'instant 0,*
 - *les durées de bon fonctionnement sont des variables aléatoires indépendantes de même loi exponentielle de paramètre λ ,*
 - *les durées de réparation sont des variables aléatoires indépendantes de même loi de transformée de Laplace m^* ,*
 - *les durées de bon fonctionnement et de réparation sont indépendantes,*
- alors la disponibilité admet pour transformée de Laplace :*

$$A^*(s) = \frac{1}{s + \lambda - \lambda m^*(s)}.$$

En particulier si m est la loi exponentielle de paramètre μ , on trouve :

$$A^*(s) = \frac{\mu}{\lambda + \mu} \frac{1}{s} + \frac{\lambda}{\lambda + \mu} \frac{1}{s + \lambda + \mu}$$

et en inversant on retrouve le résultat de l'équation 5.1. Dans le cas général, il faut calculer $m^*(s)$ et inverser la transformée de Laplace $A^*(s)$.

Exercice : Calculer la disponibilité en régime stationnaire en fonction du temps moyen de réparation.

Si $A(t)$ a comme limite A_∞ quand t tend vers l'infini, alors $sA^*(s)$ tend également vers A_∞ lorsque s tend vers 0 : on remarque que $\int_0^\infty se^{-st} dt = 1$ et on écrit

$$sA^*(s) - A_\infty = \int_0^T se^{-st}(A(t) - A_\infty) dt + \int_T^\infty se^{-st}(A(t) - A_\infty) dt,$$

on choisit T assez grand pour que la seconde intégrale soit petite et on fait tendre s vers 0 dans la première. La réciproque est vraie si $A(t)$ n'est pas trop irrégulier au voisinage de l'infini. On démontre que c'est le cas ici.

Si on note m_1 le temps moyen de réparation ($m_1 = \int_0^\infty x m(dx)$), $m^*(s)$ a pour développement limité en $s = 0$ $m^*(s) = 1 - sm_1 + o(s^2)$ et

$$\begin{aligned} \lim_{t \rightarrow +\infty} A(t) &= \lim_{s \rightarrow 0} sA^*(s) \\ &= \lim_{s \rightarrow 0} \frac{s}{s + \lambda - \lambda(1 - sm_1 + o(s^2))} \\ &= \frac{1}{1 + \lambda m_1}. \end{aligned}$$

Si la loi m du temps de réparation est exponentielle de paramètre μ on retrouve

$$\lim_{t \rightarrow +\infty} A(t) = \frac{\mu}{\lambda + \mu}.$$

5.5 Systèmes simples

Nous allons nous intéresser à des systèmes composés de plusieurs éléments. Ces éléments peuvent être en série, en parallèle ou former un réseau plus complexe. Dans ce paragraphe nous allons nous contenter d'étudier des systèmes simples, soit du fait de leur configuration, soit du fait du petit nombre d'éléments.

5.5.1 Éléments en série

Considérons un système composé de n éléments en série : on suppose que la défaillance de l'un quelconque des n éléments entraîne la défaillance du système. Supposons que les instants $(T_i)_{1 \leq i \leq n}$ de défaillance des éléments soient des variables aléatoires indépendantes. Si T est le premier instant de défaillance du système on a : $T = \inf_{1 \leq i \leq n} T_i$. La fiabilité du système vaut :

$$R(t) = \mathbf{P}(T > t) = \mathbf{P}(\forall i (1 \leq i \leq n) \quad T_i > t) = \prod_{i=1}^n \mathbf{P}(T_i > t) = \prod_{i=1}^n R_i(t).$$

Si $\lambda_i(t)$ est le taux de défaillance de l'élément i et $\lambda(t)$ le taux de défaillance du système, on a :

$$R(t) = \exp\left(-\int_0^t \lambda(u) du\right) = \exp\left(-\int_0^t \sum_{i=1}^n \lambda_i(u) du\right).$$

On en déduit que : $\lambda(u) = \sum_{i=1}^n \lambda_i(u)$.

Si les éléments du système sont réparables, on s'intéresse à la disponibilité. Si on note $A_i(t)$ la disponibilité de l'élément i ($1 \leq i \leq n$) et $A(t)$ la disponibilité du système, on a la relation : $A(t) = \prod_{i=1}^n A_i(t)$. On a établi le résultat suivant :

PROPOSITION 5.5.1 *Pour un système constitué de n éléments en série indépendants et réparables, la fiabilité, le taux de défaillance et la disponibilité du système global sont donnés par :*

$$R(t) = \prod_{i=1}^n R_i(t), \quad \lambda(u) = \sum_{i=1}^n \lambda_i(u), \quad A(t) = \prod_{i=1}^n A_i(t).$$

5.5.2 Éléments en parallèle

Un système de n éléments est dit en parallèle si la panne de tous les éléments est nécessaire pour entraîner la panne du système. On suppose que les temps de défaillance des éléments sont des variables aléatoires indépendantes et on conserve les notations du paragraphe ci-dessus.

Si les éléments ne sont pas réparables, le premier instant de panne du système est donné par : $T = \sup_{1 \leq i \leq n} T_i$. La fiabilité du système global vaut :

$$R(t) = 1 - \mathbf{P}(T \leq t) = 1 - \prod_{i=1}^n \mathbf{P}(T_i \leq t) = 1 - \prod_{i=1}^n (1 - R_i(t)).$$

Il n'y a pas de formule simple donnant le taux de défaillance du système en fonction du taux de défaillance de chaque élément.

Si les éléments du système sont réparables, on peut écrire avec les mêmes notations que pour les éléments en série : $A(t) = 1 - \prod_{i=1}^n (1 - A_i(t))$.

Lorsque les éléments du système sont réparables, le calcul de la fiabilité est beaucoup plus compliqué car si un élément tombe en panne, il peut être réparé avant que les autres éléments tombent en panne et ainsi de suite. Il n'y a pas de formule simple. Lorsque les taux de défaillance et de réparation sont constants on peut utiliser les techniques markoviennes pour faire ce calcul.

Prenons par exemple deux éléments ayant chacun un taux de défaillance égal à λ et un taux de réparation égal à μ . Les temps de fonctionnement et de réparation de ces deux éléments sont des variables aléatoires indépendantes. On a vu (cf. 2.4.3, p. 25) que :

$$A_i(t) = P_t(1, 1) = \frac{\mu}{\lambda + \mu} + \frac{\lambda}{\lambda + \mu} e^{-(\lambda + \mu)t},$$

d'où

$$A(t) = 1 - \left(\frac{\lambda}{\lambda + \mu} \left(1 - e^{-(\lambda + \mu)t} \right) \right)^2.$$

Considérons le processus à (X_t) à valeurs dans l'ensemble $\{0, 1, 2\}$ représentant le nombre d'éléments en état de marche au temps t . Le système fonctionne si X_t vaut 1 ou 2, il est en panne lorsque X_t vaut 0. $(X_t)_{t \geq 0}$ est un processus markovien de sauts de matrice génératrice :

$$A = \begin{pmatrix} -2\mu & 2\mu & 0 \\ \lambda & -(\lambda + \mu) & \mu \\ 0 & 2\lambda & -2\lambda \end{pmatrix}$$

Il serait facile de calculer les caractéristiques de $(X_t)_{t \geq 0}$: probabilité de transition (à l'aide des équations de Kolmogorov) et probabilité invariante. On retrouverait ainsi des valeurs que l'on peut obtenir plus simplement en considérant les deux processus de Markov indépendants qui décrivent l'état de chaque élément.

Pour calculer la fiabilité, on va rendre l'état de panne du système, c'est à dire ici l'état 0, absorbant. On considère le processus Y_t égal à X_t jusqu'à ce que X_t atteigne 0, ensuite Y_t reste absorbé en 0. Ce processus est un processus markovien de sauts de matrice génératrice égale à :

$$\tilde{A} = \begin{pmatrix} 0 & 0 & 0 \\ \lambda & -(\lambda + \mu) & \mu \\ 0 & 2\lambda & -2\lambda \end{pmatrix}$$

Si on note $P_t(i, j)$ les probabilités de transition ce nouveau processus et si au temps 0 les deux éléments sont en état de marche, la fiabilité du système initial

est égale à la disponibilité du système modifié et vaut donc :

$$R(t) = P_t(2,1) + P_t(2,2).$$

Les probabilités de transitions se déduisent des équations $P'_t = P_t \tilde{A}$:

$$\begin{aligned} P'_t(2,1) &= -(\lambda + \mu) P_t(2,1) + 2\lambda P_t(2,2), \\ P'_t(2,2) &= \mu P_t(2,1) - 2\lambda P_t(2,2). \end{aligned}$$

Ce système peut se résoudre en diagonalisant la matrice du système. On peut aussi prendre les transformées de Laplace $P_1^*(s)$ et $P_2^*(s)$ des fonctions $P_t(2,1)$ et $P_t(2,2)$. On trouve :

$$\begin{aligned} sP_1^*(s) &= -(\lambda + \mu) P_1^*(s) + 2\lambda P_2^*(s), \\ sP_2^*(s) - 1 &= \mu P_1^*(s) - 2\lambda P_2^*(s), \end{aligned}$$

d'où

$$P_1^*(s) = \frac{2\lambda}{s^2 + (3\lambda + \mu)s + 2\lambda^2} \quad \text{et} \quad P_2^*(s) = \frac{s + \lambda + \mu}{s^2 + (3\lambda + \mu)s + 2\lambda^2}.$$

Si s_1 et s_2 sont les deux racines du trinôme $s^2 + (3\lambda + \mu)s + 2\lambda^2$, s_1 et s_2 sont réelles, négatives et distinctes :

$$s_1 = \frac{-(3\lambda + \mu) - \sqrt{\lambda^2 + 6\lambda\mu + \mu^2}}{2} \quad \text{et} \quad s_2 = \frac{-(3\lambda + \mu) + \sqrt{\lambda^2 + 6\lambda\mu + \mu^2}}{2}.$$

En inversant les transformées de Laplace (décomposer la fraction en éléments simples et utiliser le dictionnaire d'images de l'annexe) on trouve :

$$R(t) = \frac{s_2}{s_2 - s_1} e^{s_1 t} - \frac{s_1}{s_2 - s_1} e^{s_2 t} \quad (s_1 \text{ et } s_2 < 0).$$

On sait que : MTTF = $\int_0^\infty R(t) dt = R^*(0)$. Donc

$$\text{MTTF} = P_1^*(0) + P_2^*(0) = \frac{3\lambda + \mu}{2\lambda^2}.$$

La fiabilité, comme le MTTF, dépend de l'état du système à l'instant 0. Ici on l'a calculé avec l'état 2 à l'instant initial. Si l'état initial vaut 1, on trouve :

$$\begin{aligned} sP_1^*(s) - 1 &= -(\lambda + \mu) P_1^*(s) + 2\lambda P_2^*(s) \\ sP_2^*(s) &= \mu P_1^*(s) - 2\lambda P_2^*(s) \end{aligned}$$

d'où

$$P_1^*(s) = \frac{2\lambda + s}{s^2 + (3\lambda + \mu)s + 2\lambda^2} \quad \text{et} \quad P_2^*(s) = \frac{\mu}{s^2 + (3\lambda + \mu)s + 2\lambda^2}$$

et

$$\text{MTTF} = P_1^*(0) + P_2^*(0) = \frac{2\lambda + \mu}{2\lambda^2}.$$

Pour calculer le MUT, on va se placer en régime stationnaire pour X_t . Le début d'une période de fonctionnement du système correspond à une transition de l'état 0 à l'état 1. Le MUT est donc égal au MTTF avec état initial 1 :

$$\text{MUT} = \frac{2\lambda + \mu}{2\lambda^2}.$$

D'autre part il est clair que : $\text{MTTR} = \text{MDT} = \frac{1}{2\mu}$. On a alors :

$$\text{MTBF} = \text{MUT} + \text{MDT} = \frac{2\lambda + \mu}{2\lambda^2} + \frac{1}{2\mu} = \frac{(\lambda + \mu)^2}{2\mu\lambda^2}.$$

Chapitre 6

Fiabilité de systèmes complexes

L'étude de la fiabilité d'un système complexe passe par une analyse des conditions dans lesquelles le système tombe en panne. Il existe plusieurs types de représentation de cette analyse que nous exposerons dans le premier paragraphe de ce chapitre. Puis nous étudierons les systèmes markoviens. Enfin nous donnerons quelques pistes pour l'étude de systèmes non markoviens.

6.1 Représentation de la logique d'un système

6.1.1 Diagramme de fiabilité

Pour représenter la logique d'un système, la représentation la plus naturelle est le *diagramme de fiabilité*. Ce diagramme est souvent très proche du schéma fonctionnel du système. Dans cette représentation, les blocs représentant des éléments (matériels ou événements) dont la défaillance entraîne la défaillance du système sont placés en série, ceux dont la défaillance ne provoque la défaillance du système qu'en combinaison avec d'autres blocs sont disposés en parallèle avec ces derniers. Le diagramme de fiabilité est un graphe sans circuit admettant une entrée et une sortie dont les sommets représentent les éléments du système et dont les arcs représentent les relations entre les différents éléments.

Le système fonctionne s'il existe un chemin de succès entre l'entrée et la sortie du diagramme de fiabilité. L'ensemble des chemins de succès représente l'ensemble des états de marche du système.

Dans le chapitre précédent nous avons calculé la disponibilité de certains systèmes simples représentés par des diagrammes de fiabilité en série et en parallèle dont les éléments sont indépendants. On peut essayer de calculer la disponibilité d'un système en remplaçant successivement les éléments en série et en

parallèles par des éléments uniques dont on calcule la disponibilité. Cette technique fonctionne bien pour des systèmes simples mais devient vite pénible pour de gros systèmes. Lorsque les éléments sont très fiables, leur disponibilité est proche de 1. On peut alors faire des calculs approchés avec les indisponibilités qui sont proches de 0.

Ainsi en notant $\bar{A}_i(t)$ l'indisponibilité de l'élément i au temps t , on peut écrire l'indisponibilité $\bar{A}(t)$ du système constitué de n éléments *en série* :

$$\bar{A}(t) = 1 - \prod_{i=1}^n (1 - \bar{A}_i(t)) \approx \sum_{i=1}^n \bar{A}_i(t).$$

Si les éléments du système sont *en parallèle*, on a :

$$\bar{A}(t) = \prod_{i=1}^n \bar{A}_i(t).$$

Ces formules sont très simples, mais il faut faire attention de rester dans un cadre où l'approximation est valide.

6.1.2 Arbre de défaillance

Une des représentations les plus utilisées de la logique d'un système est l'arbre de défaillance (fault tree). Cette méthode naquit en 1962 dans les bureaux d'étude de la compagnie BELL. Une utilisation extensive en a été faite dans les études sur la sûreté des réacteurs nucléaires (rapport Rasmussen ou Wash 1400).

On part d'un événement indésirable unique et bien défini. Dans notre cadre il s'agit du non-fonctionnement du système. Dans le cas d'une étude de sûreté d'un système, il s'agit d'un événement dont les conséquences sont graves. L'arbre de défaillance représente graphiquement les combinaisons d'événements qui conduisent à la réalisation de cet événement indésirable. Il sera formé de niveaux successifs tels que chaque événement est généré par des événements de niveau inférieur par l'intermédiaire de divers opérateurs (ou portes) logiques. Ce processus déductif est poursuivi jusqu'à ce qu'on arrive à des événements indépendants entre eux et probabilisables. Ces événements de base peuvent être des pannes, des erreurs humaines, des conditions extérieures.

Un certain nombre de conventions sont utilisées par les fiabilistes pour dessiner un arbre de défaillance :

- un cercle représente une défaillance de base,
- un rectangle représente un événement intermédiaire,
- un losange représente une défaillance supposée de base qui pourrait être subdivisée en événements de base mais ne le sera pas faute d'intérêt ou de renseignements,

- un double losange représente un événement dont les causes ne sont pas développées mais le seront ultérieurement,
- un opérateur (ou porte) logique ET est représenté par le symbole 
- un opérateur (ou porte) logique OU est représenté par le symbole 

Exemple : on considère le système représenté par l'arbre de défaillance suivant (figure 6.1).

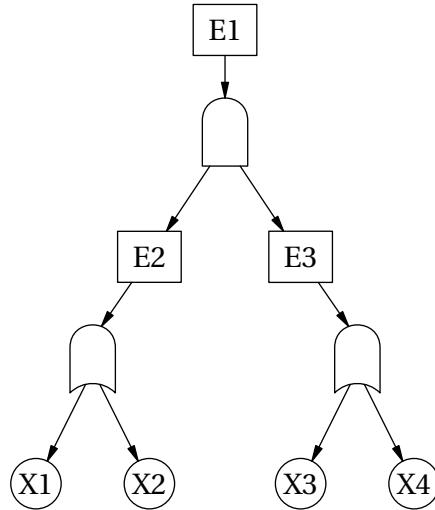


FIGURE 6.1 – Arbre de défaillance

Si p_1, p_2, p_3 et p_4 sont les probabilités des événements X_1, X_2, X_3 et X_4 , la probabilité de E_2 vaut : $e_2 = 1 - (1 - p_1)(1 - p_2)$. Celle de E_3 vaut : $e_3 = 1 - (1 - p_3)(1 - p_4)$. La probabilité de E_1 vaut : $e_1 = e_2 e_3$.

Les probabilités p_i sont en général les probabilités de défaillance d'un élément. Comme plus haut, lorsque ces probabilités sont très petites, on peut écrire :

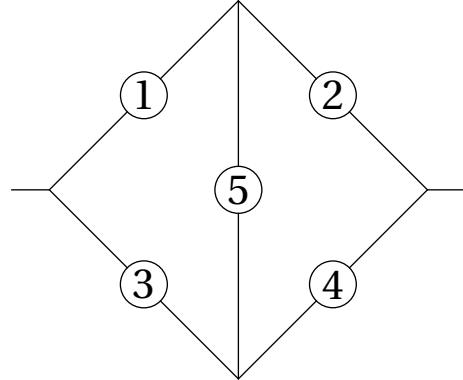
$$e_2 \approx p_1 + p_2 \quad \text{et} \quad e_3 \approx p_3 + p_4.$$

6.1.3 Coupes minimales

Une coupe est un ensemble d'éléments dont la panne entraîne la panne du système. Une coupe minimale est une coupe ne contenant aucun autre coupe. Une coupe est un ensemble d'éléments intersectant chacun des chemins de succès. La recherche des coupes minimales est d'un grand intérêt pour l'étude de la fiabilité et de la disponibilité des systèmes. En effet chaque coupe correspond à un combinaison significative de pannes. L'interprétation des coupes minimales permet de repérer les points faibles du système, les fausses redondances, l'influence

d'un élément donné sur la fiabilité du système.

Exemple : Dans le système représenté par le diagramme de fiabilité ci-contre, il y a quatre coupes minimales : $\{1,3\}$, $\{2,4\}$, $\{3,5,2\}$ et $\{1,5,4\}$.



Les coupes minimales peuvent être déterminées automatiquement, soit à partir du diagramme de fiabilité, soit à partir de l'arbre de défaillance.

Si on note C_i ($1 \leq i \leq m$) les coupes minimales d'un système, on peut écrire l'indisponibilité \bar{A} :

$$\bar{A} = \mathbf{P}\left(\bigcup_{i=1}^m C_i\right).$$

La probabilité $\mathbf{P}(C_i)$ est égale au produit des indisponibilités de chaque élément de la coupe C_i . Si ces quantités sont très petites on peut écrire :

$$\bar{A} \approx \sum_{i=1}^m \mathbf{P}(C_i).$$

À l'aide de la formule de Poincaré il est possible d'obtenir une approximation plus précise ou un encadrement.

Il est intéressant de se rendre compte de l'importance d'un élément particulier dans la disponibilité d'un système. On vient de voir une méthode permettant de calculer l'indisponibilité d'un système. Si on s'intéresse à un élément i donné, cette indisponibilité \bar{A} est une fonction de l'indisponibilité \bar{A}_i de l'élément i . On définit différents nombres qui rendent compte de l'influence de \bar{A}_i sur \bar{A} . En voici deux parmi tous ceux que l'on trouve dans la littérature :

- facteur d'importance marginale : $\frac{\partial \bar{A}}{\partial \bar{A}_i}$,
- facteur d'importance critique : $\frac{\bar{A}_i}{\bar{A}} \frac{\partial \bar{A}}{\partial \bar{A}_i}$.

Le logiciel ARBRE, conçu par N. LIMNIOS de l'U.T.C. de Compiègne, dessine l'arbre de défaillance d'un système, calcule la disponibilité du système en fonction de la disponibilité des éléments de base, calcule des facteurs d'importance.

6.2 Systèmes markoviens

Lorsque les taux de défaillance et de réparation des éléments d'un système sont constants et lorsque les temps de fonctionnement et de réparation de ces éléments sont indépendants, on peut représenter l'évolution du système par un processus markovien de sauts $(X_t)_{t \geq 0}$ à valeurs dans un espace fini formé des états $\{1, 2, \dots, n\}$.

Nous noterons A la matrice génératrice du processus et nous supposerons que le processus est irréductible. Nous supposerons que les états $\{1, 2, \dots, l\}$ correspondent à des états de fonctionnement du système et que les états $\{l + 1, \dots, n\}$ correspondent à des états de panne du système.

On note $P(t)$ la loi de la variable aléatoire X_t , $P^*(s)$ la transformée de Laplace de $P(t)$. Les équations de Kolmogorov pour le processus s'écrivent :

$$\frac{dP}{dt}(t) = P(t)A,$$

ou après transformation de Laplace :

$$sP^*(s) - P(0) = P^*(s)A,$$

d'où :

$$P^*(s) = P(0)(sI - A)^{-1}.$$

On a donc une expression explicite de $P^*(s)$. Il s'agit d'une formule théorique qui n'est pas très utilisable lorsque le nombre d'états est grand. Il faut en effet inverser la matrice $(sI - A)$ et ensuite inverser la transformée de Laplace. Ceci n'est possible à faire analytiquement que dans quelques systèmes simples. Dans un système complexe, on peut utiliser les techniques de résolution numérique des équations différentielles pour trouver directement les solutions de l'équation de Kolmogorov.

On sait que le processus se stabilise sur un état stationnaire décrit par une probabilité π qui satisfait l'équation $\pi A = 0$. Cette probabilité satisfait également la relation :

$$\pi = \lim_{t \rightarrow \infty} P(t) = \lim_{s \rightarrow 0} sP^*(s).$$

Pour calculer la fiabilité du système, on part d'un loi initiale $P(0)$ qui ne charge que les états de fonctionnement. On remplace le processus $(X_t)_{t \geq 0}$ par le processus $(Y_t)_{t \geq 0}$ qui coïncide avec X_t tant que X_t est dans un état de marche et qui reste piégé dans un état de panne dès que X_t atteint un tel état. Ce processus est également un processus markovien de saut. Sa matrice génératrice A' s'obtient

à partir de A en remplaçant les lignes $\{l + 1, \dots, n\}$ de A par des lignes de 0. La fiabilité et le MTTF du système peuvent s'écrire :

$$R(t) = \sum_{i=1}^l \mathbf{P}(Y_t = i) \quad \text{et} \quad \text{MTTF} = \int_0^\infty R(t) dt = R^*(0).$$

On peut écrire les équations de Kolmogorov pour le processus $(Y_t)_{t \geq 0}$ et trouver une expression explicite pour la transformée de Laplace $R^*(s)$ de la fiabilité. La transformée de Laplace du vecteur formé des l premières composantes de la loi de Y_t s'écrit : $P_{1,l}(0)(sI_l - A_{1,l})^{-1}$ où $P_{1,l}(0)$ représente les l premières composantes de $P(0)$, $A_{1,l}$ la matrice carrée d'ordre l correspondant aux l premières lignes et colonnes de A et I_l la matrice identité d'ordre l . La transformée de Laplace de la fiabilité est la somme des l composantes de ce vecteur. On a alors en notant $\mathbf{1}_l$ le vecteur colonne de dimension l formé de 1 :

$$\text{MTTF} = -P_{1,l}(0)(A_{1,l})^{-1} \mathbf{1}_l.$$

On constate évidemment que le MTTF dépend de la loi initiale $P_{1,l}(0)$.

Lorsqu'au temps initial le système est dans un état de panne, on peut faire des calculs semblables pour obtenir la maintenabilité et le MTTR : en notant $P_{l+1,n}(0)$ la loi initiale sur les états de panne, $A_{l+1,n}$ la matrice carrée d'ordre $n - l$ extraite de la matrice A et formée des éléments de A sur les lignes et colonnes d'ordre supérieur ou égal à $l + 1$ et $\mathbf{1}_{n-l}$ le vecteur colonne de dimension $n - l$ formé de 1, on obtient :

$$\text{MTTR} = -P_{l+1,n}(0)(A_{l+1,n})^{-1} \mathbf{1}_{n-l}.$$

Nous allons maintenant chercher les nombres caractéristiques de l'état stationnaire MUT et MDT lorsqu'il n'y a qu'un seul état de panne. Alors nous avons évidemment :

$$\text{MTTR} = \text{MDT} = \frac{1}{-A(n,n)}.$$

Pour identifier le MUT, supposons que nous soyons en régime stationnaire et qu'à l'instant 0 le système soit en panne. Alors, après le premier instant de saut, le processus se trouve dans l'état i ($1 \leq i \leq n - 1$) avec la probabilité $\frac{A(n,i)}{-A(n,n)}$. Le MUT est alors le MTTF avec cette probabilité comme loi initiale :

$$\text{MUT} = \sum_{i=1}^{n-1} \sum_{j=1}^{n-1} \frac{A(n,i)}{A(n,n)} (A_{1,n-1})^{-1}(i,j).$$

On peut évidemment faire le même type de calculs lorsqu'il n'y a qu'un seul état de fonctionnement.

6.3 Systèmes non markoviens

Dans ce paragraphe, nous allons voir comment ramener l'étude de certains systèmes non markoviens à l'étude de systèmes markoviens.

6.3.1 Méthode des états fictifs

Si l'hypothèse « taux de défaillance constant » est souvent raisonnable, il n'en est pas de même de l'hypothèse « taux de réparation constant ». La loi du temps de réparation n'est pas souvent une loi exponentielle. Si cette loi est la convoluée de plusieurs lois exponentielles, on peut introduire des états fictifs pour rendre markovien le processus décrivant le système. Ainsi lorsque le temps de saut de l'état i à l'état j a comme loi la convoluée des lois exponentielles m_1 et m_2 , on introduit un état fictif k entre i et j ; le système passe de i à k au bout d'un temps de loi exponentielle m_1 et passe ensuite de k à j au bout d'un temps indépendant de loi exponentielle m_2 .

On introduit ainsi autant d'états fictifs que nécessaires et on rend le processus markovien. Cette méthode est en particulier possible lorsque les lois de réparation sont des lois d'Erlang (loi de la somme de n variables aléatoires indépendantes de même loi exponentielle). Avec cette méthode en perspective, on peut chercher à estimer les lois de panne et de réparation qui interviennent dans le système comme des convolées de deux lois (ou plus), exponentielles indépendantes de paramètres différents. Cette méthode à l'inconvénient d'introduire des états en plus et de compliquer le graphe du système.

Exemple : Considérons un système formé d'un seul élément dont le taux de défaillance est constant égal à λ et le taux de réparation variable correspondant à un temps de réparation de loi d'Erlang de paramètres μ et 2. On peut introduire un état fictif entre l'état de marche et l'état de panne. Le processus décrivant alors le système est à 3 états : état 1 pour le système en fonctionnement, état 2 pour le système en panne, état 3 pour l'état de panne fictif supplémentaire. Le processus markovien de sauts correspondant admet alors comme matrice :

$$A = \begin{pmatrix} -\lambda & \lambda & 0 \\ 0 & -\mu & \mu \\ \mu & 0 & -\mu \end{pmatrix}$$

Si on suppose que le système est en état de marche à l'instant initial, en écrivant

les équations de Kolmogorov et en prenant la transformée de Laplace, on trouve :

$$\begin{aligned} (s + \lambda) (P_1)^*(s) - \mu (P_3)^*(s) &= 1 \\ -\lambda (P_1)^*(s) + (s + \mu) (P_2)^*(s) &= 0 \\ -\mu (P_2)^*(s) + (s + \mu) (P_3)^*(s) &= 0 \end{aligned}$$

d'où :

$$\begin{aligned} (P_1)^*(s) &= \frac{(s + \mu)^2}{s(s^2 + s(\lambda + 2\mu) + \mu^2 + 2\lambda\mu)}, \\ (P_2)^*(s) &= \frac{\lambda(s + \mu)}{s(s^2 + s(\lambda + 2\mu) + \mu^2 + 2\lambda\mu)}, \\ (P_3)^*(s) &= \frac{\lambda\mu}{s(s^2 + s(\lambda + 2\mu) + \mu^2 + 2\lambda\mu)}. \end{aligned}$$

La transformée de Laplace de la disponibilité vaut $(P_1)^*(s)$ et celle de l'indisponibilité vaut $(P_2)^*(s) + (P_3)^*(s)$. On retrouve la formule déjà vue pour l'indisponibilité limite :

$$A(+\infty) = \lim_{s \rightarrow 0} s (P_1)^*(s) = \frac{\mu}{\mu + 2\lambda}.$$

6.3.2 Méthode des variables complémentaires

Supposons que l'évolution du système soit décrite par un processus de sauts $(X_t)_{t \geq 0}$ à valeurs dans E tel que les temps de séjour dans les différents états du système soient des variables aléatoires non exponentielles. Ce processus n'est pas markovien. Cependant si au moment d'un saut la loi de saut ne dépend que de l'état du système à cet instant et si la loi du temps de séjour dans le nouvel état ne dépend que de cet état on a affaire à un processus semi-markovien. Si on introduit la variable aléatoire U_t représentant le temps passé par le processus dans l'état où il est à l'instant t , le processus $(X_t, U_t)_{t \geq 0}$ est markovien. Mais c'est un processus à valeurs dans $E \times \mathbf{R}_+$. Cet espace n'est plus dénombrable et le processus $(X_t, U_t)_{t \geq 0}$ n'est pas un processus de sauts. C'est un processus de Markov plus compliqué. Il satisfait également des équations de Kolmogorov qui, traduites dans ce cadre, donnent des équations aux dérivées partielles pour la densité du couple (X_t, U_t) .

Exemple : Considérons un élément de taux de défaillance constant λ et de taux de réparation variable égal à $\mu(\cdot)$. Ce système est décrit par le processus de sauts $(X_t)_{t \geq 0}$ qui vaut 1 lorsque le système fonctionne et 0 lorsqu'il est en panne. On note $g(\cdot)$ la densité de la loi du temps de réparation et $P_1(t)$ la probabilité que le

système soit en fonctionnement au temps t . On considère également $p_2(t, x)$ la densité de U_t lorsque le système est en panne au temps t :

$$\forall A \text{ borélien de } \mathbf{R}_+, \quad \mathbf{P}(X_t = 2, U_t \in A) = \int_A p_2(t, x) dx.$$

On peut alors écrire les équations :

$$\begin{aligned} P_1(t + \Delta t) &= P_1(t) (1 - \lambda \Delta t) + \Delta t \int p_2(t, x) \mu(x) dx + o(\Delta t) \\ p_2(t + \Delta t, x + \Delta t) &= p_2(t, x) (1 - \mu(x) \Delta t) + o(\Delta t), \end{aligned}$$

ce qui donne en divisant par Δt en faisant tendre Δt vers 0 :

$$\begin{aligned} \frac{dP_1}{dt}(t) &= -\lambda P_1(t) + \int p_2(t, x) \mu(x) dx \\ \frac{\partial p_2}{\partial t}(t, x) + \frac{\partial p_2}{\partial x}(t, x) &= -\mu(x) p_2(t, x) \end{aligned}$$

avec les conditions limites :

$$P_1(0) = 1, \quad p_2(t, x) = 0 \text{ si } x > t, \quad p_2(t, 0) = \lambda P_1(t).$$

On peut résoudre numériquement ces équations. On peut aussi donner une solution explicite à l'aide des transformées de Laplace. On prend les transformées de Laplace en t dans les équations ci-dessus et on trouve :

$$\begin{aligned} s(P_1)^*(s) - 1 &= -\lambda (P_1)^*(s) + \int (p_2)^*(s, x) \mu(x) dx \\ s(p_2)^*(s, x) + \frac{\partial (p_2)^*}{\partial x}(s, x) &= -\mu(x) (p_2)^*(s, x). \end{aligned}$$

La résolution de l'équation différentielle ci-dessus donne :

$$(p_2)^*(s, x) = \lambda (P_1)^*(s) \exp\left(-sx - \int_0^x \mu(x) dx\right).$$

d'où

$$(P_1)^*(s) = \frac{1}{s + \lambda - \lambda g^*(s)}.$$

On retrouve la formule vue au chapitre précédent. Il reste évidemment à inverser la transformée de Laplace.

Cette méthode s'applique au cas général d'un système décrit par un processus de sauts semi-markovien lorsqu'il n'y a pas de transition possible entre les états de panne et lorsque l'état initial est un état de fonctionnement.

Transformation de Laplace

Le tableau suivant donne les transformées de Laplace

$$\mathcal{L}(f)(s) = F(s) = \int_0^{+\infty} e^{-st} f(t) dt \quad (s \in \mathbf{C})$$

de quelques fonctions f usuelles nulles sur \mathbf{R}_- .

Original $f(t)$	Image $F(s)$	Conditions
e^{-at}	$\frac{1}{s+a}$	$a \in \mathbf{R}$
$t^n e^{-at}$	$\frac{n!}{(s+a)^{n+1}}$	$a \in \mathbf{R}, n \in \mathbf{N}$
t^α	$\frac{\Gamma(\alpha+1)}{s^{\alpha+1}}$	$\alpha \in \mathbf{R}, \alpha > -1$
$\cos \omega t$	$\frac{s}{s^2 + \omega^2}$	$\omega \in \mathbf{R}$
$\sin \omega t$	$\frac{\omega}{s^2 + \omega^2}$	$\omega \in \mathbf{R}$
$f(\lambda t)$	$\frac{1}{\lambda} F\left(\frac{s}{\lambda}\right)$	$\lambda > 0$
$f(t-a)$	$e^{-as} F(s)$	$a > 0$
$e^{-at} f(t)$	$F(s+a)$	$a \in \mathbf{C}$

Rappel : dérivation et convolution

$$\mathcal{L}(f') = s \mathcal{L}(f) - f(0) \quad \text{et} \quad \mathcal{L}(f * g) = \mathcal{L}(f) \cdot \mathcal{L}(g)$$

Bibliographie

Sur les files d'attente

S. ASMUSSEN : *Applied probability and queues*. J. Wiley 1987.

Un livre très complet abordant les techniques markoviennes et les techniques de renouvellement.

A. CERNAULT : *Simulation des systèmes de production*. Cepuades 1988.

Un livre fait par des praticiens de la simulation industrielle.

R. COOPER : *Introduction to queueing theory*. Macmillan 1972.

Une introduction aux modèles de files d'attente très agréable à lire.

E. GELENBE and I. MITRANI : *Analysis and synthesis of computer systems*. Academic Press 1979.

Un livre fait par des informaticiens présentant de nombreuses applications des files d'attente à l'informatique.

E. GELENBE et G. PUJOLLE : *Introduction aux réseaux de files d'attente*. Eyrolles 1985.

Un livre orienté vers les applications à l'informatique.

J. PELLAUMAIL : *Probabilités, statistiques, files d'attente*. Dunod 1986.

Un cours de base en probabilités avec une partie sur les files d'attente.

J. PELLAUMAIL : *Graphes, simulation, L-matrices*. Hermès 1992.

Un livre spécialisé sur les techniques de simulation.

Ph. ROBERT : *Réseaux et files d'attente*. Springer Verlag 2000.

Un livre riche et complet qui va bien au delà du présent cours.

A. RUEGG : *Processus stochastiques*. Presses polytechniques romandes 1989.

Un livre d'initiation très agréable à lire sur les techniques markoviennes en files d'attente et en fiabilité.

Sur la fiabilité

J.-L. BON *Fiabilité des systèmes*. Masson 1995.

Un livre facile d'accès sur les aspects théoriques et pratiques de la fiabilité.

Ch. COCOZZA-THIVENT *Processus stochastiques et fiabilité des systèmes*.

Springer 1997.

Un livre très complet sur les aspects théoriques et pratiques de la fiabilité.

N. LIMNIOS : *Arbres de défaillance*. Hermès 1992.

Un livre intéressant sur les arbres de défaillance.

N.R. MANN, R.E. SCHAFER, N.D. SINGPURWALLA : *Methods for statistical analysis of reliability and life data*. J. Wiley 1974.

Un livre présentant les techniques statistiques d'estimation des différentes lois intervenant dans une étude de fiabilité.

A. PAGES, M. GONDRAN : *Fiabilité des systèmes*. Eyrolles 1980.

Un livre présentant tous les aspects les plus modernes de la fiabilité.

A. VILLEMEUR : *Sûreté de fonctionnement des systèmes industriels*. Eyrolles 1988.

Un livre très complet.