

Modélisation et analyse stochastiques des réseaux de télécommunications

Laurent DECREUSEFOND et Pascal MOYAL

Table des matières

Chapitre 1. Introduction	13
1.1. Trafic, charge, Erlang, etc.	13
1.2. Notations et nomenclature	20
1.3. Lindley et Beneš	21
1.4. Notes et commentaires	28
PREMIÈRE PARTIE. MODÉLISATION À TEMPS DISCRET	33
Chapitre 2. Suites récurrentes stochastiques	35
2.1. Espace canonique	36
2.2. Schéma de Loynes	42
2.3. Couplage	46
2.4. Comparaison des suites récurrentes stochastiques	52
2.5. Notes et commentaires	56
Chapitre 3. Chaînes de Markov	59
3.1. Définition et exemples	59
3.2. Propriété de Markov forte	63
3.3. Classification des états	66
3.4. Mesures et probabilité invariantes	75
3.5. Calcul pratique de la probabilité invariante	90
3.6. Problèmes	92
3.7. Notes et commentaires	95
Chapitre 4. Files d'attente stationnaires	97
4.1. Files à un serveur	98
4.2. La file <i>processor sharing</i>	118
4.3. Files en parallèle	121

6 Modélisation et analyse des réseaux

4.4. La file à S serveurs	131
4.5. Le système à une infinité de serveurs	138
4.6. Les files d'attente avec clients impatientes	142
4.7. Notes et commentaires	160
Chapitre 5. La file M/GI/1	163
5.1. Le nombre de clients dans la file	163
5.2. Formules de Pollacek-Khinchin	166
5.3. Temps de séjour	170
5.4. Distribution de queue du temps d'attente	171
5.5. Périodes d'activité	173
DEUXIÈME PARTIE. MODÉLISATION À TEMPS CONTINU	179
Chapitre 6. Processus de Poisson	181
6.1. Définitions	182
6.2. Propriétés	188
6.3. Analogie discret : le processus de Bernoulli	193
6.4. Simulation du processus de Poisson	196
6.5. Processus de Poisson non homogène	197
6.6. Processus de Cox	201
6.7. Problèmes	202
6.8. Notes et commentaires	203
Chapitre 7. Processus de Markov	205
7.1. Préliminaires	205
7.2. Construction trajectorielle	206
7.3. Semi-groupe markovien et générateur infinitésimal	211
7.4. Problème de martingale	227
7.5. Réversibilité et applications	231
7.6. MMPP	237
7.7. Problèmes	243
7.8. Notes et commentaires	245
Chapitre 8. Systèmes à attente	247
8.1. Formule de Little	247
8.2. File à un serveur	250
8.3. File à plusieurs serveurs	254
8.4. La file <i>processor sharing</i>	261
8.5. La file M/M/ ∞	262
8.6. Le processus des départs	263
8.7. Réseaux de files d'attente	265

8.8. Problèmes	273
8.9. Notes et commentaires	276
Chapitre 9. Modèles à pertes	279
9.1. Généralités	279
9.2. Modèle d'Erlang	282
9.3. La file M/M/1/1 + K	284
9.4. L'effet <i>trunk</i>	287
9.5. Modèle d'Engset	288
9.6. File IPP/M/S/S	289
9.7. Modèles d'Erlang généralisés	292
9.8. Réseaux hiérarchiques	296
9.9. Un modèle avec renoncement	301
9.10. Un centre d'appels avec mise en attente et clients impatientes	307
9.11. Problèmes	309
9.12. Notes et commentaires	310
TROISIÈME PARTIE. MODÉLISATION SPATIALE	313
Chapitre 10. Processus ponctuels spatiaux	315
10.1. Motivation	315
10.2. Géométrie stochastique	316
10.3. Processus de Poisson	317
10.4. Analyse stochastique	332
10.5. Problèmes	342
10.6. Notes et commentaires	344
Annexes	346
A. Cuisine et dépendance	347
A.1. Espace de probabilités, processus	347
A.2. Espérance conditionnelle	357
A.3. Espaces de vecteurs et ordres	362
A.4. Processus à variation finie	366
A.5. Martingales	373
A.6. Transformée de Laplace	389
A.7. Notes et commentaires	390
Bibliographie	391
Index	395

Préface

Dans les télécommunications mobiles, l'ARCEP (Autorité de Régulation des Communications Electroniques et des Postes) publie chaque année une analyse de la qualité des différents réseaux radiomobiles. Pour la voix, les deux critères sont la capacité de démarrer une communication et de la maintenir deux ou cinq minutes ainsi que la qualité audio de la communication. Pour chaque service de données, le temps de transmission et l'intégrité du message (SMS, MMS) sont testés dans différentes situations : milieu urbain, semi-urbain ; pour les piétons, les voitures, les TGV, etc. La figure 1 illustre les quantités mesurées pour un MMS.

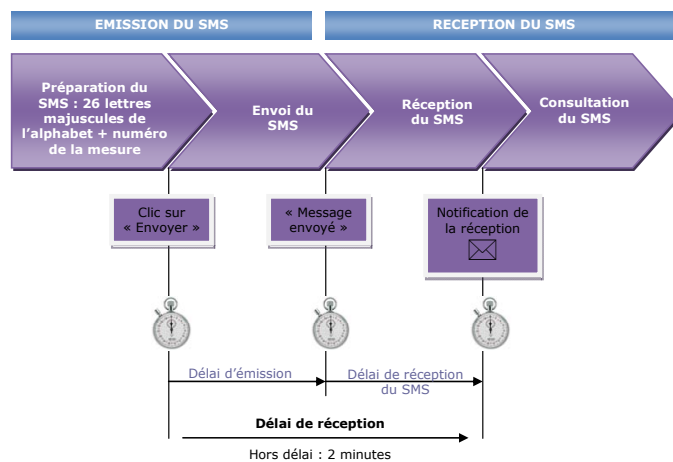


Figure 1 – Procédure de mesure de l'ARCEP pour la qualité d'un MMS. Extrait du rapport 2010 de l'ARCEP [ARC 10]

Les résultats de ces mesures deviennent parfois des arguments commerciaux. *A contrario*, de mauvais résultats nuisent gravement et rapidement à l'image d'un opérateur. Il y a donc un enjeu majeur pour tous les intervenants du domaine.

La satisfaction de certains de ces critères dépend directement du nombre de ressources allouées au réseau, notamment la capacité de ce que l'on appelle les stations de base. L'opérateur doit aller disposer de moyens quantitatifs pour anticiper la demande et son impact sur le dimensionnement de son réseau. Si l'on veut dépasser le stade de la divination, la modélisation s'impose alors. Il s'agit de traduire en équations, ayant certes parfois un aspect kabbalistique, le phénomène que l'on veut étudier. A chaque situation, peuvent correspondre plusieurs modèles selon que l'on s'intéresse à l'échelle microscopique ou macroscopique, au comportement en temps long ou en temps court, etc. Idéalement, le choix ne doit se faire qu'en fonction du but recherché mais il est aussi conditionné par les connaissances techniques et mathématiques des personnes qui construisent le modèle.

Une fois le problème posé, il faut le résoudre : en d'autres termes, si l'on a des nombres en entrée, il faut qu'il y ait des nombres en sortie. Grâce aux progrès de l'informatique, la situation a radicalement changé dans les vingt dernières années. Il est maintenant possible de calculer des quantités qui ne sont plus seulement définies par des formules mathématiques explicites, mais qui peuvent résulter d'algorithmes plus ou moins sophistiqués.

Un modèle est aussi bien souvent un support pour la simulation, cette façon de créer de façon artificielle une simplification de la réalité. Si cette méthode ne donne bien souvent que des résultats approximatifs et fort coûteux en temps de calcul, elle est aussi souvent la seule possible.

Nous avons donc essayé dans cet ouvrage de montrer ce à quoi pouvaient servir les modèles stochastiques dans le cadre des réseaux de télécommunications, tant du point de vue quantitatif que parfois seulement qualitatif. Nous avons voulu varier les approches possibles (temps discret, temps continu, chaînes ou processus de Markov, suites récurrentes, modélisation spatiale) pour permettre au lecteur de poursuivre lui-même ses efforts de modélisation. Nous n'avons pas, loin s'en faut, abordé tous les thèmes et toutes les techniques sur lesquels travaillent les chercheurs en « performances de réseaux », notamment les limites fluides et les mesures de Palm, mais nous espérons que notre lecteur pourra s'emparer de cette riche littérature pour prolonger sa réflexion. Nous avons essayé d'être le plus complet possible dans les prérequis mathématiques. Les démonstrations et les résultats manquants se trouvent aisément dans de nombreux livres qui apparaissent dans les références. Pour insister sur les aspects informatiques et pour aider notamment nos lecteurs étudiants, nous avons explicité aussi souvent que nécessaire les algorithmes à mettre en place pour résoudre tel ou tel problème.

Les langages comme Octave, Scilab ou Scipy/Numpy (disponibles à travers la plateforme SAGE) se prêtent particulièrement bien aux calculs vectoriels qui apparaissent ici et permettent d’instancier les algorithmes décrits en quelques lignes seulement.

Ce livre n’existerait pas sans le concours d’un nombre considérable de personnes. La première ébauche de cet ouvrage est un polycopié de cours de Telecom ParisTech écrit par L. Decreusefond, D. Kofman, H. Korezlioglu et S. Tohmé. L’introduction aux martingales discrètes doit beaucoup à un polycopié de A.S. Üstünel. Nous avons essayé autant que possible de présenter les protocoles réseaux sous-jacents ; le décryptage de normes de plusieurs milliers de pages et la capacité à en extraire la « substantifique moelle » demandent de nombreuses et fines connaissances dans une grande variété de champs disciplinaires, ainsi qu’une infinie patience. Nous tenons donc à remercier C. Rigault et tout particulièrement P. Martins, sans qui nous ne saurions pas ce qu’est le RTC et encore moins l’OFDMA.

Nous tenons à remercier amicalement N. Limnios, qui nous a proposé de nous lancer dans cette aventure de longue haleine, ainsi que nos collègues F. Baccelli, C. Graham, P. Robert et A.S. Üstünel, avec lesquels nous avons beaucoup discuté sur ces sujets depuis plusieurs années. Ce livre n’aurait certainement pas été ce qu’il est sans l’inspiration née de la lecture de leurs ouvrages respectifs. Nos étudiants ou collègues, E. Ferraz, I. Flint, P. Martins, A. Vergne, T.T. Vu ont relu et amendé tout ou partie de cet ouvrage. Qu’ils soient remerciés d’avoir participé à cette tâche souvent ingrate. Les erreurs résiduelles sont nôtres.

Un grand merci à nos compagnes pour nous avoir supportés à certains moments délicats. Merci à Adele pour son aide.

Chapitre 1

Introduction

1.1. Trafic, charge, Erlang, etc.

En électricité, il faut compter les ampères ou les volts ; en météorologie, il faut mesurer la pression ; en télécommunications, il faut compter les erlangs.

Au début, en 1870, était le téléphone. La plupart des concepts et des notations découlent donc de cette époque. Si l'on observe une jonction téléphonique sur une période temps de longueur T , on définit son trafic écoulé, noté ρ , comme le pourcentage du temps pendant lequel cette jonction est occupée (voir figure 1.1), c'est-à-dire que l'on a :

$$\rho = \frac{\sum_i t_i}{T}.$$

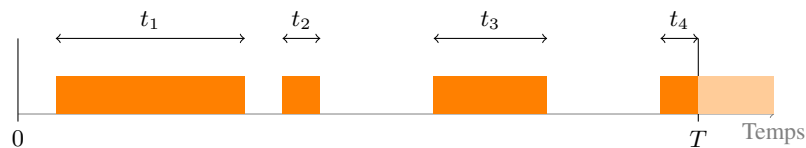


Figure 1.1 – Trafic d'une jonction : rapport du temps d'occupation par le temps total d'observation

A priori, le trafic est une quantité sans dimension puisque c'est le rapport du temps d'occupation par le temps total. Il n'empêche qu'il a tout de même une unité, l'erlang en hommage à celui qui fut, avec Palm, l'un des pionniers de l'évaluation de performances

des réseaux téléphoniques. Ainsi, une charge d'un erlang correspond à une jonction occupée en permanence.

Si l'on observe plusieurs jonctions, le trafic écoulé par ce faisceau est la somme des trafics de chacune des jonctions :

$$\rho_{\text{faisceau}} = \sum_{\text{jonctions}} \rho_{\text{jonction}}.$$

Ce n'est donc plus un pourcentage mais on peut donner à cette quantité une interprétation physique si l'on fait l'hypothèse ergodique. En effet, supposons que le nombre de jonctions soit grand, on peut calculer un taux moyen d'occupation de deux manières différentes : soit en calculant le pourcentage de temps d'occupation d'une jonction particulière sur une grande période de temps ; soit en calculant à un instant donné, le pourcentage de jonctions occupées. En mécanique statistique, l'ergodicité d'un ensemble de molécules de gaz signifie que les moyennes spatiales (moyennes calculées sur l'ensemble des molécules du gaz par exemple) sont égales aux moyennes temporelles (moyennes calculées sur une molécule pendant une grande période de temps). Par analogie, nous supposons dorénavant qu'il en va de même pour le taux d'occupation des jonctions téléphoniques. On a donc :

$$p = \lim_{T \rightarrow \infty} \frac{1}{T} \sum_j t_j = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_n X_n(t), \quad (1.1)$$

où $X_n = 1$ si la jonction n est occupée à l'instant t , $X_n = 0$ sinon. Notons que dans le terme de droite, la valeur de t est quelconque. Cela signifie que nous avons implicitement supposé que le système est dans un état stationnaire, c'est-à-dire que statistiquement son comportement ne change pas avec le temps. Lorsque le nombre de jonctions est grand, il est illusoire de vouloir définir une structure de corrélation entre elles. Il est donc raisonnable de supposer qu'une jonction est libre ou occupée indépendamment, au sens probabiliste, de la situation des autres jonctions. Par conséquent, à un instant donné t , le nombre de jonctions occupées suit une loi binomiale de paramètres N (le nombre total de jonctions) et p (calculé par (1.1)). Le nombre moyen de jonctions occupées est donc, à chaque instant, Np .

Cette relation donne un moyen simple et efficace d'estimer p . Les commutateurs en téléphonie étaient des ordinateurs chargés de gérer les appels téléphoniques en temps réel. Ces prouesses technologiques de l'époque dont la conception a mené quantité d'ingénieurs à la dépression tant leur complexité était grande, avait entre autres fonctions de compter le nombre d'appels en cours à chaque instant. En moyennant ce nombre sur une quinzaine de secondes, on obtient donc une estimation assez précise du nombre moyen d'appels simultanés, soit une estimation de p .

Se pose alors une question : comment choisir T et quand faire les mesures ? Il est en effet clair que le trafic fluctue au cours de la journée en fonction des activités

humaines. Puisque l'on veut dimensionner et garantir un taux d'échec faible, il faut envisager le pire cas et faire les mesures à fort trafic. Depuis des générations, la période d'observation est l'heure et l'on regarde le trafic à l'heure la plus chargée de la journée.

Imaginons un instant que les appels se produisent toutes les $1/\lambda$ secondes et durent exactement $1/\mu$ secondes avec $\mu > \lambda$, voir figure 1.2.

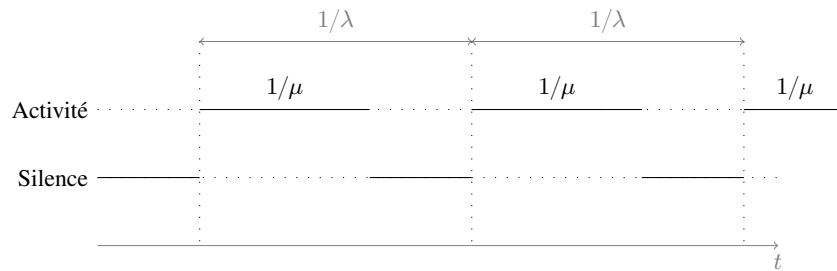


Figure 1.2 – Appels déterministes

Il est évident que le nombre d'appels entre 0 et T est de l'ordre de λT donc le taux d'occupation d'une telle ligne est donné par :

$$\frac{1}{T}(\lambda T \times 1/\mu) = \lambda/\mu.$$

Evidemment, dans la réalité, ni les interarrivées, ni les durées d'appel ne sont déterministes. Imaginons une situation dans laquelle les durées d'occupation et les temps d'inactivité sont indépendants les uns des autres avec une distribution commune à toutes les périodes d'activité, respectivement les périodes de silence, voir figure 1.3. Mathématiquement parlant, soit $(X_n, Y_n, n \geq 1)$ une suite de variables aléatoires indépendantes. Pour tout n , X_n a pour loi \mathbf{P}_X et Y_n pour loi \mathbf{P}_Y . On suppose que ces deux lois ont des moments d'ordre 1 finis et on note :

$$1/\mu = \int y d\mathbf{P}_Y(y), 1/\tau = \int x d\mathbf{P}_X(x), \lambda = \frac{1}{1/\tau + 1/\mu}.$$

On pose :

$$T_0 = 0, T_n = T_{n-1} + X_n + Y_n, T'_n = T_n + X_n,$$

et :

$$X(t) = \begin{cases} 1 & \text{si } T'_n \leq t < T_{n+1} \\ 0 & \text{si } T_n \leq t < T'_n. \end{cases}$$

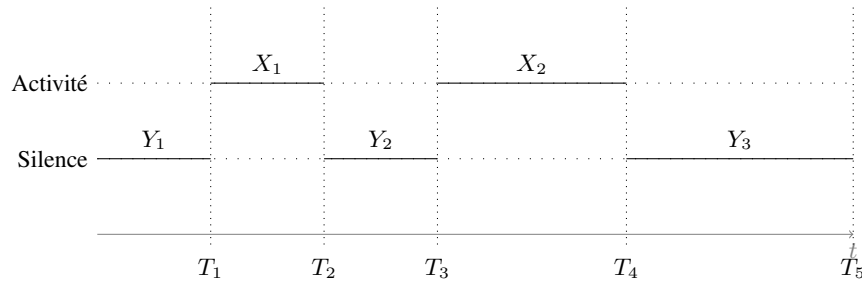


Figure 1.3 – Appels aléatoires

On remarque $\mathbf{E}[T_{n+1} - T_n] = \mathbf{E}[X_n + Y_n] = 1/\lambda$ donc λ apparaît comme le nombre moyen d'arrivées par unité de temps.

La théorie des processus de renouvellement ou la formule de Little (voir chapitre 8) permet de montrer que l'on a la limite suivante :

$$\frac{1}{T} \int_0^T \mathbf{1}_1(X(s)) \, ds \xrightarrow{T \rightarrow \infty} \frac{1}{\mu} \frac{1}{1/\lambda} = \frac{\lambda}{\mu}. \quad (1.2)$$

On a donc montré dans ce cas particulier mais relativement général que :

charge = nombre moyen d'appels par unité de temps \times durée moyenne d'un appel .

Cette formule simple est importante car elle permet de basculer dans le monde de l'internet. L'ARPANET, lointain ancêtre d'Internet, est né dans les années 1970 suite aux travaux faits pour l'armée américaine qui souhaitait un réseau de transmission de données distribué parce que plus résistant contre une attaque ponctuelle. A l'inverse du réseau téléphonique où la ressource, c'est-à-dire la jonction téléphonique, est réservée pendant toute la durée de la communication, les réseaux de données sont dits « non commutés ». L'information est envoyée sous forme de paquets de quelques octets auxquels on ajoute un certain nombre d'identifiants, suivants chacun leur propre chemin dans les méandres du réseau. Dans ce cadre, la notion de jonction n'a plus d'équivalent et la notion de trafic doit être redéfinie. Au vu de ce qui précède, la charge sera définie comme le produit du nombre moyen de requêtes par unité de temps et du temps moyen de traitement, quel que soit le type de réseau considéré.

Les volts et les ampères ne sont rien sans la loi d'Ohm, la météorologie n'est rien sans les équations de la mécanique des fluides. Les erlangs ne servent à rien si l'on ne précise pas la façon dont se produisent les arrivées ou le temps que durent les appels. Comme le prouve la figure 1.4, la charge ne suffit pas à caractériser le nombre de

ressources nécessaires au bon fonctionnement d'un système : pour écouler un erlang, le premier système peut se satisfaire d'une seule jonction, tandis que dans le deuxième, il en faut impérativement deux.

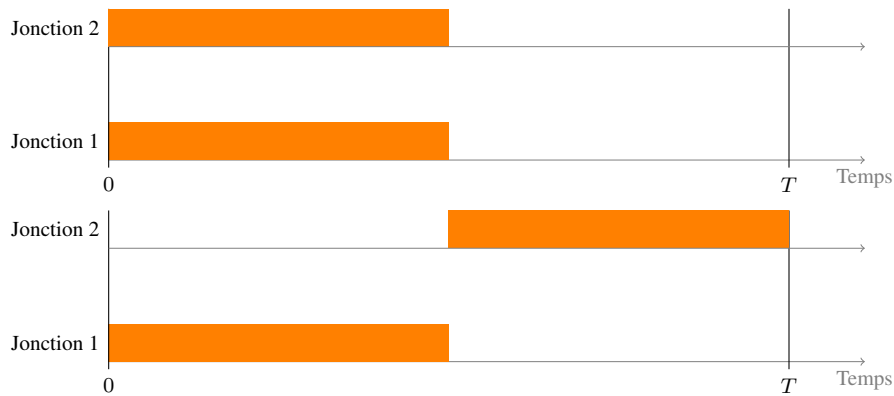


Figure 1.4 – Deux systèmes nécessitant d'écouler un erlang

La situation était simple jusque dans les années 1990. Tant que le système téléphonique était concerné, le processus d'arrivée des appels était modélisé par un processus de Poisson (voir chapitre 6). Cela se justifiait d'une part par les observations statistiques qui le confirmaient et par un raisonnement qualitatif bien connu : chaque abonné au téléphone a une probabilité faible d'appeler à un instant donné mais il y a beaucoup (mathématiquement un nombre infini) d'abonnés. L'approximation d'une loi binomiale par une loi de Poisson justifie, qu'au moins à un instant donné, le nombre d'appels simultanés suive une loi de Poisson. En ce qui concerne la durée des appels, des mesures sur les commutateurs prouvaient que l'on pouvait la considérer comme suivant une loi exponentielle de moyenne trois minutes. Au final, le trafic généré par un abonné était considéré comme valant 0,12 erlang à l'heure chargée.

Pour les réseaux de données, malgré de sérieux doutes quant à sa validité, les paquets étaient toujours supposés arriver selon un processus de Poisson et leur temps de traitement suivre une loi exponentielle dont la moyenne dépendait de la vitesse de traitement des routeurs et de la moyenne de leur longueur. Malheureusement, début 1990, des chercheurs des Bell Labs montrent sur une campagne statistique intensive, que l'on ne peut décemment pas assimiler le trafic dans un réseau haut débit à du trafic poissonnien. En effet, si l'on considère le nombre de paquets qui arrivent pendant cent secondes, dix secondes, ..., dix millisecondes, on constate des comportements similaires à ceux de la figure 1.5.

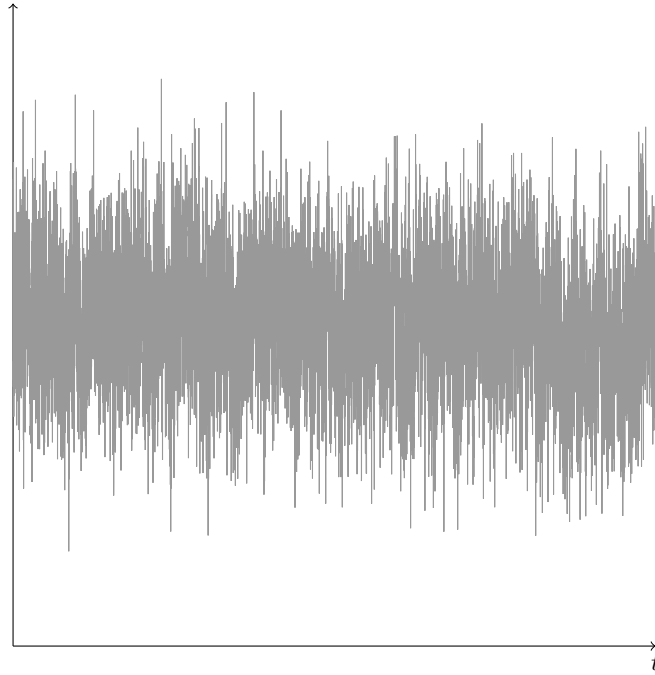


Figure 1.5 – A toutes les échelles de temps observables, après renormalisation, le trafic enregistré par intervalles de temps ressemble à celui-ci

Pour un trafic de type poissonnien, on observe des comportements visuellement similaires à ceux de la figure 1.5 pour de petites échelles de temps mais lorsque l'on agglomère les paquets reçus par période d'une dizaine ou d'une centaine de secondes, on obtient plutôt un graphique du type de la figure 1.6.

Cette invariance du nombre de paquets à grande échelle de temps pour le processus de Poisson s'explique par le théorème A.36. En effet, d'après celui-ci :

$$\frac{1}{\lambda t}(N(t) - \lambda t) \xrightarrow{t \rightarrow \infty} 0, \text{ soit } \frac{N(t)}{t} \xrightarrow{t \rightarrow \infty} \lambda.$$

A la vitesse où sont envoyés les paquets, λ est de l'ordre du millier de paquets par seconde donc un intervalle de temps de cent secondes peut être considéré comme un temps « infini » et on obtient que le nombre de paquets émis par période de cent secondes est quasiment constant. Par conséquent, la situation réelle semble être plus proche de celle d'un système « fractal » : le système conserve la même allure à toutes

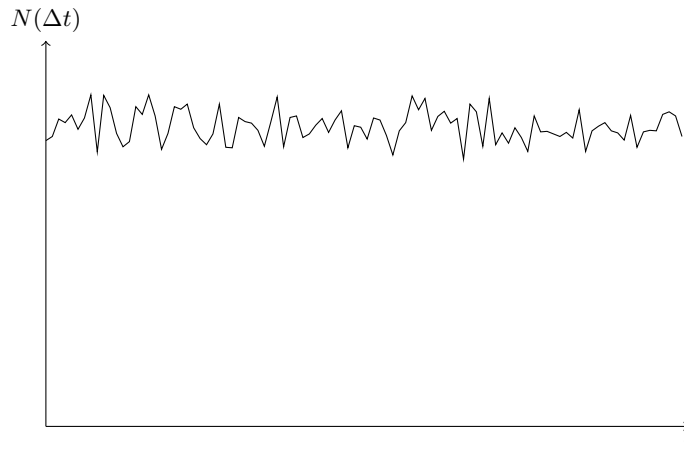


Figure 1.6 – Trafic poissonnien observé sur de grands intervalles de temps

les échelles de temps. Les chercheurs des Bell Labs proposaient même dans leur papier un modèle alternatif, le mouvement brownien fractionnaire.

DÉFINITION 1.1.– Un mouvement brownien fractionnaire d'indice de Hurst H est un processus gaussien centré de covariance donnée par :

$$\mathbf{E}[B_H(t)B_H(s)] = \frac{1}{2}(t^{2H} + s^{2H} - |t - s|^{2H}).$$

Si $H = 1/2$ on retrouve le mouvement brownien ordinaire, si $H > 1/2$ les incréments sont à covariance positive, si $H < 1/2$, ils sont à covariance négative : si le trafic a tendance à augmenter, immédiatement après il aura tendance à diminuer. La figure 1.7 représente des trajectoires de mouvement brownien fractionnaire pour différentes valeurs de H .

A partir de là, toute la communauté académique se mit à bruire autour de cette question : à quoi est dû cet aspect fractal, l'invariance est-elle vraie à toutes les échelles de temps, le modèle doit-il être fractal ou multi fractal et surtout quel est l'impact de cette forme du trafic sur le dimensionnement des files d'attente ? Au bout de dix années de recherche frénétique, on sait à peu près expliquer les raisons de la fractalité mais on ne sait pas la contrôler et pourtant son impact sur le dimensionnement est majeur.

Pour expliquer la fractalité, il suffit de reprendre pour une source un schéma d'émission comme celui de la figure 1.3 : quand $X = 1$ cela signifie que la source émet à son débit maximal, quand $X = 0$, la source n'émet pas. Motivé par les études

statistiques qui prouvent que la longueur des fichiers disponibles sur le *web* a une loi dite à « queue lourde », c'est-à-dire que $\mathbf{P}(X > x) \sim x^{-\alpha}$ pour $\alpha > 0$, par opposition à la loi exponentielle où $\mathbf{P}(X > x) = \exp(-bx)$, la durée de la période d'émission est supposée suivre une loi de Pareto, c'est-à-dire que $d\mathbf{P}_X(x) = cx^{-\alpha} \mathbf{1}_{[K, \infty[}(x)$, de même que la période de silence. Lorsque l'on superpose de nombreuses sources de ce type et que l'on regarde l'état stationnaire de cette superposition, on constate que le processus résultat est justement un mouvement brownien fractionnaire dont l'indice de Hurst dépend des puissances apparaissant dans les lois de Pareto.

Au final, tout cela n'a d'importance que si l'on cherche à raisonner au niveau des paquets. Or les protocoles actuels cherchent plutôt à agglomérer les paquets en flux (pour les prioriser par exemple) et ainsi recréer virtuellement la notion de jonction propre à notre bon vieux téléphone. Dans ces conditions, seuls les instants d'apparition et les durées de vie des flux importent ; or celles-ci sont moins faussement modélisées par un processus de Poisson et des durées de vie indépendantes avec une distribution à queue lourde. Le processus de Poisson a donc encore de beaux jours devant lui...

1.2. Notations et nomenclature

Il est d'usage de noter $T_0 = 0 < T_1 < \dots < T_n$ les instants d'arrivée des clients (paquets, sessions, appels, etc) dans le système « file d'attente ». Les durées $S_n = T_n - T_{n-1}$ sont appelées interarrivées. Le temps de service du n^e client (temps de traitement, durée de communication, etc.) est noté σ_n .

Pour distinguer les différentes files d'attente, on a recours à une nomenclature dite de Kendall. Une file d'attente est un système dynamique à événements discrets décrit par cinq paramètres : le type d'interarrivées, le type des temps de service, le nombre de serveurs, le nombre de ressources total (serveurs plus taille de la salle d'attente), discipline de service. Implicitement, les entrées et les temps de service sont des variables aléatoires indépendantes.

Pour les deux premiers points, on utilise les mêmes abréviations :

- M_λ , pour décrire des interarrivées (ou des durées de service) indépendantes de loi exponentielle de paramètre λ ;
- GI, pour décrire des interarrivées (ou des durées de service) indépendantes de même loi ;
- D_λ , pour décrire des interarrivées (ou des durées de service) déterministes égales à λ ;
- G, pour décrire des interarrivées (ou des durées de service) quelconques.

La discipline de service décrit l'ordre dans lequel sont servis les clients. Les plus connues sont les suivantes :

– FIFO ou FCFS, pour *First In First Out* ou *First Come First Served*. En français, on sert les clients dans l'ordre d'arrivée dans le système ;

– LIFO ou LCFS, pour *Last In First Out*. L'autre traduction est laissée au lecteur. On sert d'abord le dernier client arrivé. Une telle discipline peut être préemptive si le nouveau client interrompt le service en cours ou non préemptive dans le cas contraire. Dans le cas préemptif, on peut distinguer le cas *preemptive resume* où le service d'un client interrompu reprend où il en était, du cas *preemptive non resume* où le service recommence à zéro.

Il est souvent pertinent de faire le tri des clients selon un autre critère que celui de l'instant d'arrivée :

– SRPT ou *Shortest Remaining Processing Time*. On sert d'abord le client qui a le temps résiduel de service le plus faible. Cette discipline peut être préemptive ou pas ;

– EDF ou *Earliest Deadline First*. Chaque client a un marqueur d'impatience et l'on sert en premier le plus pressé.

Une discipline est dite conservative lorsque la charge en entrée est aussi celle en sortie. Evidemment, si les ressources sont finies, aucune discipline ne peut être conservative (sauf dans le cas déterministe avec charge strictement inférieure à 1). Même avec des ressources infinies, une discipline n'est pas forcément conservative : dans la discipline EDF, on peut tout à fait envisager de supprimer tous les clients qui ne sont pas servis avant leur marqueur d'impatience ; les disciplines *preemptive non resume* ne sont pas conservatives non plus puisqu'il y a plus de charge traitée que de charge en entrée. Pour le reste, il ne vous reste plus qu'à en inventer d'autres !

En l'absence d'information sur le nombre de ressources ou la discipline de service, il est entendu que le nombre de ressources est infini et que la discipline de service est la discipline FIFO.

EXEMPLE.– La file M/M/1 est la file où les inter-arrivées sont indépendantes de loi exponentielle, les temps de service indépendants de loi exponentielle et il y a un seul serveur. La salle d'attente est de taille infinie et la discipline de service FIFO.

La file GI/D/S/S+K est une file à S serveurs, K places dans la salle d'attente, des temps de service déterministe, des interarrivées indépendantes identiquement distribuées de loi quelconque.

1.3. Lindley et Beneš

1.3.1. Modèle discret

On s'intéresse souvent au nombre de clients présents dans le système mais la quantité qui contient le plus d'informations est la charge du système, définie à chaque

instant comme étant le temps nécessaire pour que le système se vide en l'absence de nouvelles arrivées. Le serveur travaille à vitesse unité, c'est-à-dire qu'il sert une unité de travail par unité de temps. Par conséquent, la charge diminue à vitesse 1 entre deux arrivées. On construit alors aisément la figure 1.8 qui représente la charge au cours du temps en fonction des arrivées et des temps de service demandés. Ce graphique permet de retrouver les instants de départ dans le cas d'une discipline FIFO donc de représenter l'évolution du nombre de clients dans le système.

DÉFINITION 1.2.– Une période d'activité d'une file d'attente est une période qui commence à l'arrivée d'un client dans un système (serveur plus buffer) vide et se termine à la fin d'un service après lequel le système est de nouveau vide.

DÉFINITION 1.3.– Un cycle est une période de temps qui commence à l'arrivée d'un client dans un système vide et se termine à l'arrivée suivante d'un client dans un système vide. C'est la concaténation d'une période d'activité et du temps d'inactivité, c'est-à-dire le temps écoulé entre le départ du dernier client de la période d'activité et l'arrivée du suivant.

REMARQUE.– Sur la figure 1.8, une période d'activité commence en T_1 et finit en D_4 . Le cycle correspondant commence en T_1 et finit en T_5 .

Il faut noter que tant qu'une politique de service est conservative, la taille d'une période d'activité en est indépendante : pour des salles d'attente de taille infinie, les périodes d'activité ont, par exemple, la même longueur pour la politique FIFO que pour la politique LIFO non préemptive ou *preemptive resume*.

Intéressons nous maintenant à la charge du système juste avant l'arrivée du client n . Si W_{n-1} est la charge du système à l'arrivée du client $n - 1$, celle-ci est augmentée de la charge apportée par ce client, soit σ_{n-1} et diminuée du temps de service écoulé entre les instants d'arrivée T_{n-1} et T_n , soit très exactement S_n . On a donc *a priori* :

$$W_n = W_{n-1} + \sigma_{n-1} - S_n.$$

Toutefois, si $S_n > W_n + \sigma_{n-1}$, l'interarrivée est tellement grande que le système s'est vidé donc $W_n = 0$ et non pas $W_n < 0$. Par conséquent la vraie formule dite de Lindley est donnée par :

$$W_n = \max(W_{n-1} + \sigma_{n-1} - S_n, 0). \quad (1.3)$$

Entre T_n et T_{n+1} , puisque le serveur travaille à vitesse unité, la charge est donnée par :

$$W(t) = \max(W_n + \sigma_n - (t - T_n), 0).$$

Ces deux équations permettent de simuler facilement la charge dans n'importe quel système, peu importe le type des arrivées ou des temps de service ou la discipline de service tant qu'elle est conservative.

Elles permettent aussi d'analyser qualitativement la stabilité du système dans des cas très généraux comme cela sera fait dans le chapitre 4.

1.3.2. Modèle fluide

Le modèle fluide consiste à assimiler une file d'attente à un réservoir de capacité infinie se vidant à vitesse unité et alimenté par un flux quelconque. On peut ainsi facilement obtenir des résultats qualitatifs sur des modèles dont l'étude résiste aux autres approches. D'une part, la méthode ne requiert pas de connaître précisément l'allure du processus d'entrée, d'autre part elle se prête particulièrement bien à l'étude des cas limites : faible et forte charges, superposition de trafics hétérogènes.

On travaille en temps continu et on suppose que tous les processus sont continus à droite avec des limites à gauche. On note :

- $S(t)$, le temps total de service pour les requêtes arrivées jusqu'à l'instant t ;
- $W(t)$, le temps d'attente virtuel d'un client arrivant à l'instant t , c'est-à-dire le temps que ce client doit attendre avant de commencer à être servi ;
- $X(t) = S(t) - t$.

Comme le système n'a pas de pertes, on a :

$$W(t) = X(t) - (t - \int_0^t \mathbf{1}_{\{0\}}(W(s)) \, ds). \quad (1.4)$$

En effet, $W(t)$ représente ce qui reste à l'instant t , $X(t)$ ce qui est rentré jusqu'à cet instant, le dernier terme représente le temps où le serveur a été actif donc la quantité de travail servi jusqu'à t puisqu'il travaille à vitesse unité. En utilisant la théorie de la réflexion, nous allons nous attacher à montrer une formulation équivalente de cette équation.

THÉORÈME 1.1 (Equation de Beneš).– Avec les notations précédentes, on obtient en toute généralité l'identité suivante, pour $x \geq 0$:

$$\mathbf{P}(W(t) < x) = \mathbf{P}(X(t) < x) - \frac{\partial}{\partial x} \int_0^t \mathbf{P}(X(t) - X(u) < x \mid W(u) = 0) \mathbf{P}(W(u) = 0) \, du, \quad (1.5)$$

et pour $-t \leq x \leq 0$:

$$\mathbf{P}(X(t) < x) = \frac{\partial}{\partial x} \int_0^{t+x} \mathbf{P}(X(t) - X(u) < x \mid W(u) = 0) \mathbf{P}(W(u) = 0) du.$$

DÉFINITION 1.4.– Soit $(X(t))$, $t \geq 0$) un processus continu à gauche dont les sauts sont positifs, le couple (W, L) résout l'équation de réflexion associé à X si :

$$- W(t) = X(t) + L(t), \forall t \geq 0;$$

$$- W(t) \geq 0, \quad \forall t \geq 0;$$

- L est un processus croissant, continu à gauche, nul en zéro tel que la mesure $dL(s)(\omega)$ ne charge que l'ensemble $\{s : W(s)(\omega) = 0\}$, c'est-à-dire que L ne croît qu'aux instants où W s'annule.

THÉORÈME 1.2.– Le problème de réflexion associé à X possède une solution unique donnée par :

$$L(t) = \sup_{s \leq t} X(s)^-, \quad W(t) = X(t) + \sup_{s \leq t} X(s)^-$$

où $x^+ = \max(x, 0)$ et $x^- = \max(-x, 0)$.

Démonstration. Si (W, L) et (\tilde{W}, \tilde{L}) sont deux solutions :

$$\begin{aligned} (W(t) - \tilde{W}(t))^2 &= (L(t) - \tilde{L}(t))^2 = 2 \int_0^t (L(s) - \tilde{L}(s)) d(L - \tilde{L})(s) \\ &= 2 \int_0^t (W(s) - \tilde{W}(s)) d(L - \tilde{L})(s) \\ &= -2 \int_0^t (W(s) d\tilde{L}(s) + \tilde{W}(s) dL(s)) \leq 0. \end{aligned}$$

où l'on a successivement utilisé : la relation entre W , X et L (respectivement \tilde{W} , X et \tilde{L}); la dérivabilité presque partout du processus croissant $s \mapsto L(\omega, s)$; la formule d'intégration par parties (A.12) et la relation « $W(s)dL(s)(\omega) = 0$ ». La positivité de la dernière intégrale découle de la positivité de W et \tilde{W} , ainsi que de celle de dL et $d\tilde{L}$ en tant que mesures. Par conséquent, $W = \tilde{W}$ et l'unicité en résulte.

Il suffit maintenant de vérifier que le processus $\sup_{s \leq t} X(s)^-$ convient pour L . Il est clair qu'ainsi défini, L est un processus croissant, positif, nul en 0. D'autre part :

$$W(t) = X(t) + L(t) = X(t)^+ - X(t)^- + \sup_{s \leq t} X(s)^- \geq 0.$$

Soit T_0 un point de croissance de L , alors pour tout h positif, il existe t_h tel que $L_{t_0-h} \leq X_{t_h}^- \leq L_{t_0}$. Quand h tend vers 0, par continuité à gauche $L_{t_0} = X_{t_0}^- = -X_{t_0}$ donc $X_{t_0} + L_{t_0} = W_{t_0} = 0$. \square

COROLLAIRE 1.3.– Avec les notations précédentes, on a l'identité suivante :

$$\exp\left(-\lambda \int_0^t \mathbf{1}_{\{0\}}(W(s)) \, ds\right) = 1 - \lambda \int_0^t e^{\lambda X(s)} \mathbf{1}_{\{0\}}(W(s)) \, ds \quad (1.6)$$

Démonstration. De la relation $f(t) - f(0) = \int_0^t f'(u) \, du$, on déduit :

$$\begin{aligned} \exp\left(-\lambda \int_0^t \mathbf{1}_{\{0\}}(W(s)) \, ds\right) &= 1 - \lambda \int_0^t e^{-\lambda \int_0^s \mathbf{1}_{\{0\}}(W(u)) \, du} \mathbf{1}_{\{0\}}(W(s)) \, ds \\ &= 1 - \lambda \int_0^t e^{-\lambda(W(s)-X(s))} \mathbf{1}_{\{0\}}(W(s)) \, ds \\ &= 1 - \lambda \int_0^t e^{\lambda X(s)} \mathbf{1}_{\{0\}}(W(s)) \, ds. \end{aligned}$$

D'où le résultat. \square

Nous sommes maintenant en mesure de démontrer une version intermédiaire, aussi intéressante en soi, de l'équation de Beneš.

THÉORÈME 1.4.– Avec les notations précédentes, on a l'identité suivante :

$$\begin{aligned} \mathbf{E}[f(W(t))] &= \mathbf{E}[f(X(t))] + \mathbf{E}\left[\int_0^t f'(X(t) - X(u)) \mathbf{1}_{\{0\}}(W(u)) \, du\right] \\ &= \mathbf{E}[f(X(t))] + \int_0^t \mathbf{E}[f'(X(t) - X(u)) \mid W(u) = 0] \mathbf{P}(W(u) = 0) \, du. \quad (1.7) \end{aligned}$$

Démonstration. En multipliant les deux termes de la relation (1.6) par $e^{-\lambda X(t)}$ et en prenant les espérances, il vient :

$$\mathbf{E}\left[e^{-\lambda W(t)}\right] = \mathbf{E}\left[e^{-\lambda X(t)}\right] - \lambda \int_0^t \mathbf{E}\left[e^{-\lambda(X(t)-X(u))} \mathbf{1}_{\{0\}}(W(u)) \, du\right].$$

Plus généralement, pour toute fonction deux fois dérivable et bornée, on obtient la relation (1.7). \square

Maintenant nous pouvons donner une idée de la démonstration de l'équation de Beneš (1.5).

REMARQUE.— $\mathbf{P}(X(t) - X(u) < x \mid W(u) = 0) = 0$ pour $t + x \leq u \leq t$, puisque sur $(W(u) = 0)$:

$$X(t) - X(u) = W(t) - L(t) + L(u) \geq -(L(t) - L(u)),$$

or $L(t) \leq t$ donc $X(t) - X(u) \geq -(t - u)$. Donc $X(t) - X(u)$ ne peut pas être plus petit que x si x lui-même est inférieur à $u - t$, soit $u \geq t + x$.

Démonstration de l'équation de Beneš. En multipliant les termes de (1.5) par $e^{-\lambda x}$ puis en intégrant de $-t$ à $+\infty$, il apparaît que l'équation (1.5) est équivalente à :

$$\begin{aligned} \int_{-t}^{+\infty} e^{-\lambda x} \mathbf{P}(W(t) < x) dx &= \int_{-t}^{+\infty} e^{-\lambda x} \mathbf{P}(X(t) < x) dx \\ - \int_{-t}^{+\infty} e^{-\lambda x} \frac{\partial}{\partial x} \int_0^{t+x} \mathbf{P}(X(t) - X(u) < x \mid W(u) = 0) \mathbf{P}(W(u) = 0) du dx, \end{aligned}$$

en vertu de la remarque ci-dessus. En utilisant l'identité :

$$\int_0^{+\infty} e^{-\lambda x} \mathbf{P}(X < x) dx = \frac{1}{\lambda} \mathbf{E}[e^{-\lambda X}],$$

et une intégration par parties, on en tire :

$$\begin{aligned}
\mathbf{E} \left[e^{-\lambda W(t)} \right] &= \mathbf{E} \left[e^{-\lambda X(t)} \right] \\
&\quad - \lambda \frac{\partial}{\partial x} \int_0^t \mathbf{E} \left[e^{-\lambda(X(t)-X(u))} \mid W(u) = 0 \right] \mathbf{E} \left[\mathbf{1}_{\{0\}}(W(u)) \right] du \\
&= \mathbf{E} \left[e^{-\lambda X(t)} \right] \\
&\quad - \lambda \left[e^{-\lambda x} \int_0^{t+x} \mathbf{P}(X(t) - X(u) < x \mid W(u) = 0) \mathbf{P}(W(u) = 0) du \right]_{x=-t}^{x=+\infty} \\
&+ \lambda^2 \int_{-t}^{+\infty} e^{-\lambda x} \int_0^{t+x} \mathbf{P}(X(t) - X(u) < x \mid W(u) = 0) \mathbf{P}(W(u) = 0) du dx \\
&= \mathbf{E} \left[e^{-\lambda X(t)} \right] \\
&\quad + \lambda^2 \int_0^{+\infty} \mathbf{P}(W(u) = 0) \int_{u-t}^{+\infty} \mathbf{P}(X(t) - X(u) < x \mid W(u) = 0) dx du \\
&= \mathbf{E} \left[e^{-\lambda X(t)} \right] \\
&\quad + \lambda^2 \int_0^{+\infty} \mathbf{P}(W(u) = 0) \int_0^{+\infty} \mathbf{P}(X(t) - X(u) < x \mid W(u) = 0) dx du \\
&= \mathbf{E} \left[e^{-\lambda X(t)} \right] \\
&\quad + \lambda \int_0^{+\infty} \mathbf{P}(W(u) = 0) \int_0^{+\infty} \mathbf{E} \left[e^{-\lambda(X(t) - X(u))} \mid W(u) = 0 \right] dx du,
\end{aligned}$$

toujours d'après la remarque initiale. Le résultat découle alors du théorème 1.4 appliqué à la fonction exponentielle. \square

Au-delà de son intérêt trajectorien évident, la représentation sous forme de processus réfléchi est à la base de nombreux théorèmes de convergence. C'est aussi grâce à cette formule que l'on peut étudier l'importance de la mémoire longue sur le délai de transmission. Les autres méthodes d'origine markovienne ne pouvant traiter cette situation. Il est aussi une classe particulière de processus pour laquelle on peut déduire des résultats complets.

DÉFINITION 1.5.— Un processus X est dit à accroissements indépendants si et seulement si pour tout $0 \leq t_1 < \dots < t_n$, les variables aléatoires $X(t_1)$, $X(t_2) - X(t_1)$, \dots , $X(t_n) - X(t_{n-1})$ sont indépendantes. Le processus est dit à accroissements homogènes si pour tout couple (t, s) de réels positifs, la loi de $X(t+s) - X(t)$ est celle de $X(s)$.

THÉORÈME 1.5.– Si $X(t) = S(t) - t$ est un processus à accroissements indépendants homogènes et (W, L) est la solution du problème de réflexion associé à X , on a :

$$\sigma \int_0^{+\infty} e^{-\sigma t} \mathbf{E} \left[e^{-\lambda W(t) - \mu L(t)} \right] dt = \Psi_\sigma^-(-\lambda) \Psi_\sigma^+(-\mu) \quad (1.8)$$

où :

– Ψ la fonction de Lévy-Khintchine de X définie par $\mathbf{E} [e^{sX(t)}] = e^{t\psi(s)}$ dont on sait qu'elle peut s'écrire :

$$\Psi(s) = as + \frac{1}{2} \sigma_0^2 s^2 + \int_{|x| < 1} (e^{sx} - 1 - sx) d\Pi(x) + \int_{|x| \geq 1} (e^{sx} - 1) d\Pi(x)$$

pour une mesure Π intégrant $x^2 \wedge 1$;

– η est une fonction de Lévy-Khintchine telle que $\psi(\eta(s)) = s$;

– Ψ_σ^+ et Ψ_σ^- sont définies par :

$$\Psi_\sigma^+(s) = \frac{1}{1 - \frac{s}{\eta(s)}} \text{ et } \Psi_\sigma^-(s) = \left(1 - \frac{s}{\eta(s)}\right) \left(\frac{\sigma}{\sigma - \Psi(s)}\right).$$

Ce résultat est complémentaire des résultats à venir sur la file M/GI/1 (voir chapitre 5) puisqu'une telle file se représente par un modèle fluide en considérant :

$$S(t) = \sum_{T_n \leq t} Y_n$$

où $(T_n, n \geq 1)$ est la suite des instants d'arrivées distribués selon un processus de Poisson et $(Y_n, n \geq 1)$ une suite de variables aléatoires indépendantes et identiques distribuées.

1.4. Notes et commentaires

Pour plus de détails sur la conception des réseaux téléphoniques, leur histoire et leurs développements ultérieurs, on pourra consulter [RIG 98]. Le papier original des Bells Labs a pour référence [LEL 94], les études qui montrent que les fichiers ont une longueur qui suit une loi de Pareto sont disponibles dans [CRO 96]. L'explication mathématique de l'autosimilarité vient de [SHE 97]. Le premier à avoir étudié l'impact de l'autosimilarité pour les réseaux est I. Norros dans [NOR 94]. De nombreux résultats tant sur l'estimation statistique, que sur l'utilisation en réseaux ou des aspects plus mathématiques sur les processus à mémoire longue se trouvent dans [DOU 02].

Résumé

– La charge s'appelle aussi le trafic. Elle est définie comme le nombre moyen d'appels par unité de temps multiplié par le temps moyen de traitement d'un appel. Son unité est l'Erlang.

– La nomenclature de Kendall permet de décrire les différentes files d'attente. La file $M/M/1$ est la file où les inter-arrivées et les temps de service sont indépendants et suivent des lois exponentielles. Il n'y a qu'un serveur, la file d'attente est infinie et la politique de service, FIFO.

– On sait étudier qualitativement les files $M/M/*/FIFO$ et dans une moindre mesure les files $M/GI/1$ et $GI/M/1$. Pour les autres (autres lois d'interarrivées ou de temps de service, autres disciplines), on n'a souvent que des résultats partiels ou asymptotiques.

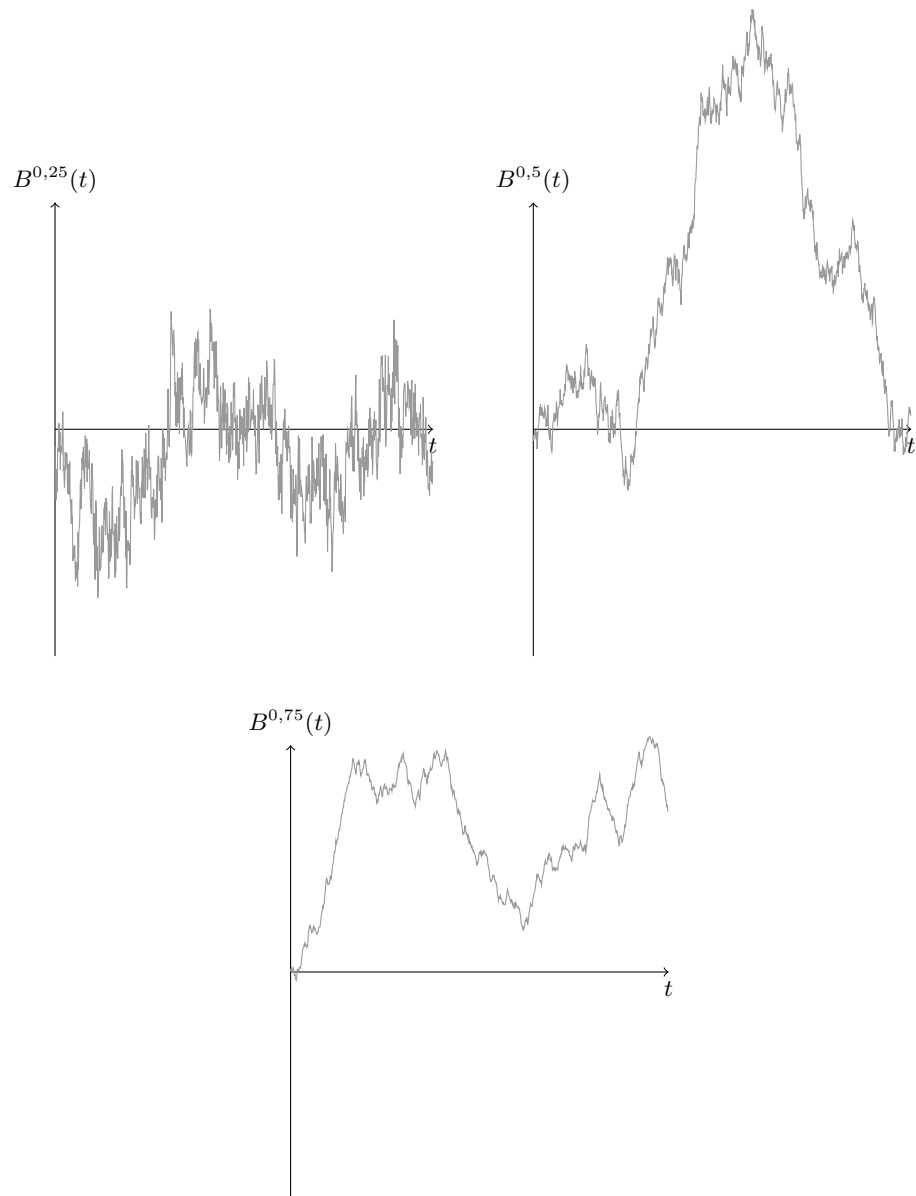


Figure 1.7 – Mouvement brownien fractionnaire pour différentes valeurs de H : de gauche à droite $H = 0,2$; $H = 0,5$; $H = 0,75$; plus H est faible plus les trajectoires sont irrégulières

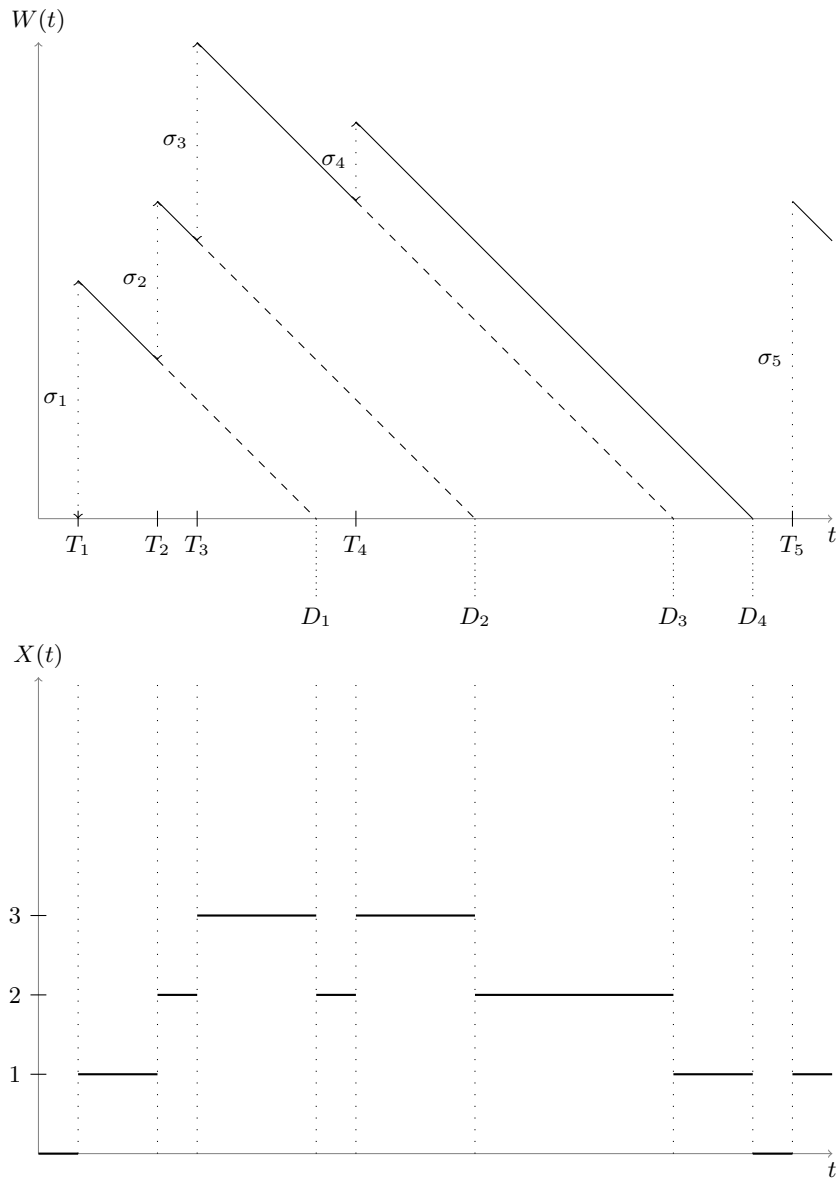


Figure 1.8 – L'évolution de la charge du système en fonction du temps

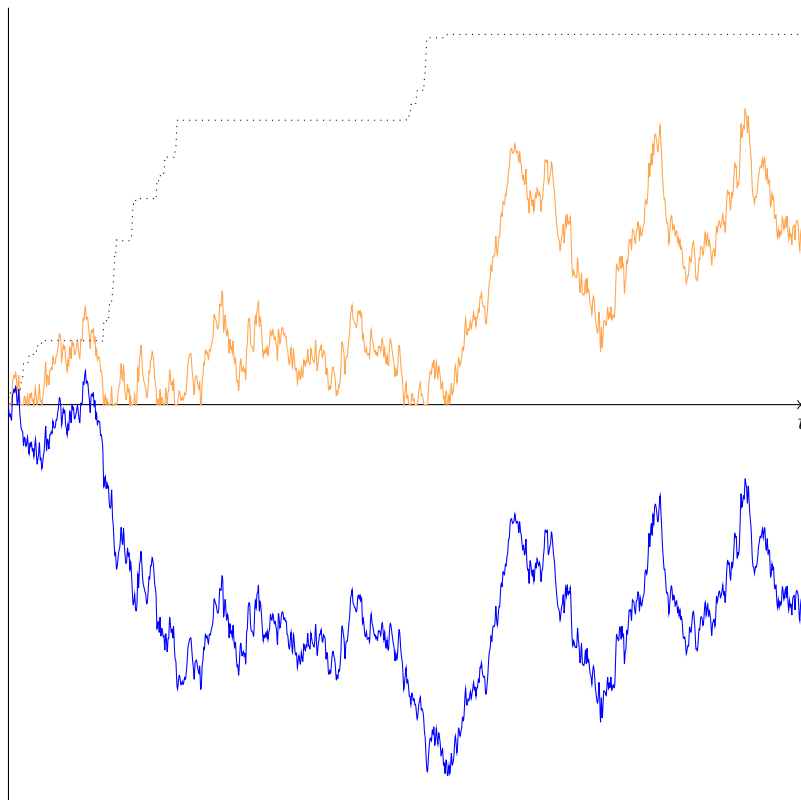


Figure 1.9 – Un exemple de processus réfléchi : en foncé, le processus d'entrée X ; en pointillé, le processus L ; en clair, le processus W

PREMIÈRE PARTIE

Modélisation à temps discret

Chapitre 2

Suites récurrentes stochastiques

La modélisation des systèmes dynamiques déterministes à temps discret repose sur les suites récurrentes, c'est-à-dire les suites de la forme $u_{n+1} = f(u_n)$. On s'intéresse principalement à la convergence quand n tend vers l'infini et à la valeur de la limite qui, à supposer que f soit continue, est nécessairement solution de l'équation $l = f(l)$.

L'objet de ce chapitre est de développer les outils permettant de répondre aux mêmes questions pour les suites récurrentes stochastiques.

Considérons par exemple une file G/G/1 (qui sera traitée exhaustivement en section 4.1). En notant $(\xi_n, n \in \mathbf{N})$ la suite des temps interarrivées et $(\sigma_n, n \in \mathbf{N})$ la suite des temps de service, la charge de travail W_{n+1} du serveur à l'arrivée du $n + 1^{\text{e}}$ client se déduit de la charge à l'arrivée du n^{e} client par l'équation de Lindley :

$$W_{n+1} = (W_n + \sigma_n - \xi_n)^+ . \quad (2.1)$$

Si les deux suites sont indépendantes (cas GI/GI/1), la suite $(W_n, n \in \mathbf{N})$ est une chaîne de Markov à valeurs dans l'espace non dénombrable \mathbf{R}^+ . Son analyse est impossible avec les outils du chapitre 3 puisque nous nous y limitons aux chaînes de Markov à espace d'états au plus dénombrable.

Evidemment, il ne saurait y avoir de convergence ponctuelle des suites de $W_n(\omega)$ pour tout ω vers la même limite, mais on peut espérer une convergence « en loi », c'est-à-dire $\mathbf{P}(W_n \in [a, b]) \xrightarrow[n \rightarrow \infty]{} \mathbf{P}(W_\infty \in [a, b])$ pour une variable aléatoire W_∞ donc il nous faut déterminer la loi. On dit alors que la suite converge vers son état stationnaire. Il est alors remarquable qu'en choisissant bien l'espace de probabilités, on peut écrire une équation déterministe, analogue à l'équation $l = f(l)$, dont la loi stationnaire est

solution. Plus généralement, l'étude asymptotique repose essentiellement sur celle des propriétés de la fonction de récurrence (monotonie, continuité, etc.), sur des critères de comparaison avec des suites de référence et sur la résolution, dans un cadre stochastique, d'une équation limite de type point fixe (voir (2.7)).

On l'aura compris, ce chapitre est purement théorique mais introduit les outils nécessaires à l'étude de la stabilité des files d'attente, sous les hypothèses les plus générales possibles.

2.1. Espace canonique

La notion de stationnarité renvoie à celle d'invariance dans le temps, c'est-à-dire qu'un décalage dans le temps ne modifie pas l'état des choses. Si l'idée se comprend aisément, sa formalisation opacifie rapidement le concept de base. Considérons dans un premier temps, l'ensemble des suites d'éléments d'un ensemble F , noté comme il se doit $F^{\mathbf{N}}$. L'opérateur de décalage θ est alors défini par :

$$\theta : \begin{cases} F^{\mathbf{N}} & \longrightarrow F^{\mathbf{N}} \\ (\omega_n, n \geq 0) & \longmapsto (\omega_{n+1}, n \geq 0) = (\omega_n, n \geq 1). \end{cases}$$

Défini de cette façon, cet opérateur a le défaut de ne pas être bijectif : si l'on se donne une suite $\beta = (\beta_n, n \geq 0)$, toutes les suites obtenues par concaténation d'un élément quelconque de F et de β ont β comme image par θ . Pour pallier ce problème, il est d'usage de travailler avec des suites indexées par \mathbf{Z} et non plus par \mathbf{N} .

⚡ Mathématiquement, cela ne change presque rien, l'espace d'indexation reste dénombrable. Philosophiquement, cela revient à dire qu'il n'y a plus d'origine des temps...

L'opérateur de décalage est donc défini sur $F^{\mathbf{Z}}$ par la formule :

$$\theta(\omega_n, n \in \mathbf{Z}) = (\omega_{n+1}, n \in \mathbf{Z}).$$

Ainsi défini, il est bijectif !

Supposons maintenant que F soit un espace polonais, de sorte que $F^{\mathbf{Z}}$ soit polonais et puisse donc être muni de la tribu de ses boréliens $\mathfrak{B}(F^{\mathbf{Z}})$. Notre espace canonique dans ce chapitre sera $\Omega = (F^{\mathbf{Z}}, \mathfrak{B}(F^{\mathbf{Z}}))$. Pour $n \in \mathbf{Z}$, on note X_n , la « n^e » application coordonnée :

$$X_n : \begin{cases} \Omega = F^{\mathbf{Z}} & \longrightarrow F \\ \omega = (\omega_n, n \in \mathbf{Z}) & \longmapsto \omega_n. \end{cases}$$

Notons au passage l'identité intéressante suivante, pour tout $k \in \mathbf{Z}$:

$$X_k = X_0 \circ \theta^k. \quad (2.2)$$

DÉFINITION 2.1.— Soit $(E, \mathcal{E}, \mathbf{P})$ un espace probabilisé et ψ une application mesurable de (E, \mathcal{E}) dans (F, \mathcal{F}) , on note $\psi^*\mathbf{P}$ la mesure image de \mathbf{P} par ψ , c'est-à-dire :

$$\forall A \in \mathcal{F}, (\psi^*\mathbf{P})(A) = \mathbf{P}(\psi^{-1}(A)),$$

où $\psi^{-1}(A) = \{x \in E, \psi(x) \in A\}$.

DÉFINITION 2.2.— Une probabilité \mathbf{P} sur Ω est dite stationnaire lorsque pour tout $A \in \mathfrak{B}(F^{\mathbf{Z}})$:

$$\mathbf{P}(A) = \mathbf{P}(\theta^{-1}A).$$

De manière équivalente, cela revient à dire que $\theta^*\mathbf{P} = \mathbf{P}$.

En particulier, si l'on considère les événements de la forme :

$$A = (X_{k_1} \in A_1, \dots, X_{k_n} \in A_n),$$

on déduit que :

$$\mathbf{P}(X_{k_1} \in A_1, \dots, X_{k_n} \in A_n) = \mathbf{P}(X_{k_1-1} \in A_1, \dots, X_{k_n-1} \in A_n). \quad (2.3)$$

Une suite de variables aléatoires satisfaisant (2.3) pour tout n , tout k_1, \dots, k_n et tout $A_1, \dots, A_n \in \mathfrak{B}(F)$, sera dite stationnaire.

⚡ Si l'espace canonique $\Omega = F^{\mathbf{Z}}$ est muni d'une probabilité stationnaire, la suite des applications coordonnées est une suite de variables aléatoires stationnaire.

REMARQUE.— Réciproquement, si la suite $(\alpha_n, n \in \mathbf{Z})$, définie sur un espace probabilisé $(\tilde{\Omega}, \tilde{\mathbf{P}})$, est une suite stationnaire, on peut considérer \mathbf{P}_α sa loi sur $F^{\mathbf{Z}}$, c'est-à-dire la mesure image de $\tilde{\mathbf{P}}$ par l'application :

$$\begin{cases} \tilde{\Omega} & \longrightarrow F^{\mathbf{Z}} \\ \tilde{\omega} & \longmapsto (\alpha_n(\tilde{\omega}), n \in \mathbf{Z}). \end{cases}$$

Il est alors automatique que \mathbf{P}_α est une mesure stationnaire sur $F^{\mathbf{Z}}$. En effet, il suffit de vérifier (2.3) pour $\mathbf{P} = \mathbf{P}_\alpha$. Or, par définition d'une mesure image, on a :

$$\mathbf{P}_\alpha(X_{k_1} \in A_1, \dots, X_{k_n} \in A_n) = \tilde{\mathbf{P}}(\alpha_{k_1} \in A_1, \dots, \alpha_{k_n} \in A_n),$$

d'où le résultat.

Dorénavant, toute suite stationnaire sera vue comme une probabilité stationnaire sur $F^{\mathbf{Z}}$.

REMARQUE.— Construire une suite stationnaire n'est pas chose aisée. L'exemple le plus simple est celui des suites de variables aléatoires indépendantes et identiquement distribuées puisque dans ce cas :

$$\tilde{\mathbf{P}}(\alpha_{k_1} \in A_1, \dots, \alpha_{k_n} \in A_n) = \prod_{j=1}^n \mathbf{P}_{\alpha_0}(A_j),$$

et que cette quantité ne dépend pas de (k_1, \dots, k_n) .

Une autre construction de suites stationnaires peut être obtenue à partir des chaînes de Markov (voir chapitre 3) irréductibles récurrentes positives. Soit \tilde{X} une chaîne de Markov sur E , d'opérateur de transition Q et de probabilité invariante π . On note $\tilde{\mathbf{P}}$ la loi de \tilde{X} sur $E^{\mathbf{N}}$ lorsque $\text{Loi}(X_0) = \pi$. On sait d'après le lemme de Kolmogorov que pour définir une probabilité sur $E^{\mathbf{Z}}$, il suffit d'en définir les lois fini-dimensionnelles. Or en posant pour tout n -uplet d'entiers relatifs $k_1 < k_2 < \dots < k_n$, tout $A_1, \dots, A_n \subset E$:

$$\mathbf{P}(X_{k_1} \in A_1, \dots, X_{k_n} \in A_n) = \tilde{\mathbf{P}}(\tilde{X}_0 \in A_1, \dots, \tilde{X}_{k_n - k_1} \in A_n),$$

on définit bien les marginales fini-dimensionnelles d'une mesure de probabilité unique sur $E^{\mathbf{Z}}$. La stationnarité de \mathbf{P} ainsi définie est évidente.

On définit le quadruplet $\mathfrak{D} = (\Omega = F^{\mathbf{Z}}, \mathcal{F} = \mathfrak{B}(F^{\mathbf{Z}}), \mathbf{P}, \theta)$.

DÉFINITION 2.3.— Une mesure de probabilité \mathbf{P} sur $F^{\mathbf{Z}}$ est ergodique lorsque :

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \Phi \circ \theta^i = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \Phi \circ \theta^{-i} = \mathbf{E}_{\mathbf{P}}[\Phi], \mathbf{P} \text{ — presque sûrement,} \quad (2.4)$$

pour toute fonction $\Phi \in L^1(\mathbf{P})$. Le cas échéant, le quadruplet \mathfrak{D} sera dit ergodique.

Une suite stationnaire $(\alpha_n, n \in \mathbf{Z})$ est dite ergodique si sa loi induit une mesure ergodique sur $F^{\mathbf{Z}}$.

EXEMPLE 2.1.— Soient a et b deux réels. La suite aléatoire $(\alpha_n, n \in \mathbf{Z})$ à valeurs dans (a, b) valant a, b, a, b, \dots avec probabilité $1/2$ et b, a, b, a, \dots avec probabilité $1/2$ est stationnaire (on a $\mathbf{P}(\alpha_n = a) = \mathbf{P}(\alpha_n = b)$ pour tout n) et ergodique (la condition (2.4) est clairement vérifiée).

EXEMPLE 2.2.— La suite $(\beta_n, n \in \mathbf{Z})$ valant a, \dots, a ou b, \dots, b avec probabilité $1/2$ pour chaque possibilité, est également stationnaire. En revanche, il est facile de vérifier qu'elle n'est pas ergodique.

LEMME 2.1.— Soit \mathfrak{D} un quadruplet ergodique. Tout événement $A \in \mathfrak{B}(F^{\mathbf{Z}})$ tel que $A = \theta^{-1}A$ est trivial : $\mathbf{P}(A) = 0$ ou 1 .

Démonstration. Pour tout entier $n \in \mathbf{N}$, on définit $\mathcal{F}_n^0 = \sigma\{X_k, k \leq n\}$ et $W_n = \mathbf{E}[\mathbf{1}_A | \mathcal{F}_n^0]$. La suite W est de manière évidente une martingale uniformément bornée donc convergente presque sûrement et dans L^1 vers $\mathbf{1}_A$. Admettons temporairement que $W_n = W_0 \circ \theta^n$. Si une suite $(u_n, n \in \mathbf{N})$ converge vers une limite l alors ces moyennes de Césaro convergent vers la même limite donc :

$$\frac{1}{n} \sum_{k=1}^n W_k \xrightarrow[n \rightarrow \infty]{} \mathbf{1}_A.$$

D'autre part :

$$\frac{1}{n} \sum_{k=1}^n W_k = \frac{1}{n} \sum_{k=1}^n W_0 \circ \theta^k \xrightarrow[n \rightarrow \infty]{} \mathbf{E}[W_0],$$

d'après l'hypothèse (2.4). Il s'ensuit que la variable aléatoire $\mathbf{1}_A$ est constante d'où le résultat.

Il nous reste à prouver que $W_n = W_0 \circ \theta^n$ ou de manière équivalente, $\mathbf{E}[\mathbf{1}_A | \mathcal{F}_n^0] = \mathbf{E}[\mathbf{1}_A | \mathcal{F}_{n-1}^0] \circ \theta$. Remarquons d'abord que si ϕ est \mathcal{F}_n^0 -mesurable alors $\phi \circ \theta^{-1}$ est \mathcal{F}_{n-1}^0 -mesurable. De plus, $\theta^{-1}A = A$ est équivalent à $\mathbf{1}_A \circ \theta = \mathbf{1}_A$. Par définition de W_n , pour toute ϕ bornée et \mathcal{F}_n^0 -mesurable, on a :

$$\begin{aligned} \mathbf{E}[\phi W_n] &= \mathbf{E}[\phi \mathbf{1}_A] \\ &= \mathbf{E}[\phi \circ \theta^{-1} \circ \theta \mathbf{1}_A \circ \theta] \\ &= \mathbf{E}[\phi \circ \theta^{-1} \mathbf{1}_A] \\ &= \mathbf{E}[\phi \circ \theta^{-1} W_{n-1}] \\ &= \mathbf{E}[\phi \circ \theta^{-1} W_{n-1} \circ \theta \circ \theta^{-1}] \\ &= \mathbf{E}[\phi W_{n-1} \circ \theta], \end{aligned}$$

où l'on a utilisé deux fois l'invariance de \mathbf{P}_X par θ , en d'autres termes la stationnarité de \mathbf{P}_X . On en déduit l'identité recherchée par identification. \square

Le résultat suivant sera fréquemment utilisé par la suite.

LEMME 2.2 (Lemme ergodique).— Soit Y , une variable aléatoire définie sur le quadruplet stationnaire ergodique \mathfrak{D} , à valeurs réelles, \mathbf{P} -presque sûrement positive, telle que $Y \circ \theta - Y$ est intégrable. Alors, $\mathbf{E}[Y \circ \theta - Y] = 0$.

Démonstration. Pour tout $n \in \mathbf{N}$, la v.a. $Y \wedge n$ est intégrable et donc :

$$\mathbf{E}[(Y \wedge n) \circ \theta - Y \wedge n] = \mathbf{E}[(Y \wedge n) \circ \theta] - \mathbf{E}[Y \wedge n] = 0. \quad (2.5)$$

La suite $((Y \wedge n) \circ \theta - Y \wedge n, n \in \mathbf{N})$ converge \mathbf{P} -presque sûrement vers $Y \circ \theta - Y$, de plus il est aisé de voir que pour tout n :

$$|(Y \wedge n) \circ \theta - Y \wedge n| = |(Y \circ \theta) \wedge n - Y \wedge n| \leq |Y \circ \theta - Y|.$$

Le théorème de convergence dominée implique que :

$$\mathbf{E}[(Y \wedge n) \circ \theta - Y \wedge n] \xrightarrow[n \rightarrow \infty]{} \mathbf{E}[Y \circ \theta - Y],$$

ce qui conclut la preuve. \square

Il est important de noter (et c'est sa raison d'être) que ce dernier résultat est vrai lorsque Y n'est pas supposée intégrable.

DÉFINITION 2.4.— Une suite récurrente stochastique (SRS en abrégé) $(W_n, n \in \mathbf{N})$ à valeurs dans E polonais, est définie sur un quadruplet stationnaire ergodique par une variable aléatoire Y dans E , une application mesurable φ de $E \times F$ dans E et les relations :

$$W_0 = Y \text{ et } W_{n+1} = \varphi(W_n, X_n) \text{ pour } n \geq 1. \quad (2.6)$$

On dira que la SRS $(W_n, n \in \mathbf{N})$ est portée par φ et issue de Y , et l'on notera souvent $(W_n^Y, n \in \mathbf{N})$ pour souligner la dépendance en la condition initiale Y .

EXEMPLE 2.3.— La suite $(W_n, n \in \mathbf{N})$ des charges de travail de la file $G/G/1$ évoquée en introduction de ce chapitre a une telle forme : on prend pour tout n , $X_n = (\sigma_n, \xi_n)$, $E = \mathbf{R}^+$, $F = \mathbf{R}^+ \times \mathbf{R}^+$ et :

$$\varphi : \begin{cases} E \times F & \longrightarrow E \\ (x, (y, z)) & \longmapsto (x + y - z)^+. \end{cases}$$

Il y a donc deux sources d'aléa dans l'évolution de W : la condition initiale Y et le « stimulus » représenté par la suite X . L'espace de probabilités sur lequel est défini M est donc :

$$(E \times F^{\mathbf{Z}}, \mathfrak{B}(E) \otimes \mathfrak{B}(F^{\mathbf{Z}})).$$

En tant que variable aléatoire, W est à valeurs dans $E^{\mathbf{N}}$. La loi de la suite W définit une probabilité sur $E^{\mathbf{N}}$. Cet espace est aussi muni d'un décalage θ_E défini de manière analogue à celui sur $F^{\mathbf{Z}}$, que nous noterons temporairement θ_F . Le décalage θ_E n'est pas bijectif mais nous n'aurons pas besoin de cette propriété pour l'instant. Les définitions de stationnarité et d'ergodicité demeurent valides à l'identique.

Le stimulus est donné par le modèle, nous ne pouvons donc qu'agir sur la condition initiale. La question est de savoir si l'on peut choisir Y comme une fonction du stimulus de sorte que la loi de W soit stationnaire. Une condition suffisante est fournie par le théorème suivant. Elle transforme une identité en loi en une identité trajectorielle que l'on espère plus facile à prouver.

THÉORÈME 2.3.— S'il existe une variable aléatoire Y telle que :

$$Y \circ \theta_F = \varphi(Y, X_0), \mathbf{P}_X \text{ — presque sûrement,} \quad (2.7)$$

alors la SRS W définie par (2.6) admet une probabilité stationnaire.

Démonstration. Moyennant quelques notations, le résultat est totalement évident. Introduisons les applications :

$$Y : \begin{cases} F^{\mathbf{Z}} & \longrightarrow F^{\mathbf{Z}} \times E \\ \omega & \longmapsto (\omega, Y(\omega)) \end{cases} \quad \text{et } W : \begin{cases} F^{\mathbf{Z}} \times E & \longrightarrow E^{\mathbf{N}} \\ (\omega, \eta) & \longmapsto (\eta, \varphi(\eta, \omega), \dots) \end{cases} .$$

On a alors le diagramme suivant :

$$\begin{array}{ccc} F^{\mathbf{Z}} & \xrightarrow{W \circ Y} & E^{\mathbf{N}} \\ \theta_F \downarrow & & \downarrow \theta_E \\ F^{\mathbf{Z}} & \xrightarrow{W \circ Y} & E^{\mathbf{N}} \end{array}$$

La n^{e} composante de $W \circ Y \circ \theta_F(\omega)$ est $W_n(Y(\theta_F \omega), \theta_F \omega)$. D'autre part, la n^{e} composante de $\theta_E \circ M \circ Y(\omega)$ est $W_{n+1}(Y(\omega), \omega)$. En particulier pour $n = 0$, on a d'une part $Y \circ \theta_F(\omega)$ et d'autre part $\varphi(Y(\omega), \omega_0)$, donc l'équation (2.7) signifie que ces deux quantités sont égales. Par récurrence, on en déduit que c'est aussi le cas pour toutes les composantes donc $W \circ Y \circ \theta_F = \theta_E \circ M \circ Y$. En termes savants, on dit que « le diagramme est commutatif ».

On en déduit que les mesures images de \mathbf{P}_X par ces deux applications sont les mêmes. Notons \mathbf{P} la loi de W , c'est-à-dire la mesure image de \mathbf{P}_X par $W \circ Y$. On a d'une part :

$$(\theta_E \circ W \circ Y)^* \mathbf{P}_X = \theta_E^* \mathbf{P}$$

et d'autre part, comme $\theta_F^* \mathbf{P}_X = \mathbf{P}_X$, on a :

$$(W \circ Y \circ \theta_F)^* \mathbf{P}_X = (W \circ Y)^* \mathbf{P}_X = \mathbf{P}.$$

On a donc prouvé que $\theta_E^* \mathbf{P} = \mathbf{P}$, ce qui selon la définition 2.2 signifie que \mathbf{P} est stationnaire. \square

La question centrale de la stationnarité de la SRS revient donc à la résolution de l'équation presque sûre (2.7). Nous proposons dans la suite de ce chapitre, deux méthodes qui permettent de conclure dans de nombreux cas.

2.2. Schéma de Loynes

Nous considérons ici le cas où l'espace d'état E est muni d'un ordre partiel \leq et d'un point minimal $\mathbf{0}$ tels que $\mathbf{0} \leq x$ pour tout $x \in E$. On suppose qu'il existe sur E une métrique d_E telle que toutes les suites croissantes et majorées pour \leq convergent dans \bar{E} , l'adhérence de E .

DÉFINITION 2.5.– Une fonction $\varphi : E \times F^{\mathbf{Z}} \rightarrow E$ est dite \leq -croissante lorsque :

$$\eta \leq \eta' \implies \varphi(\eta, \omega) \leq \varphi(\eta', \omega), \text{ d } \mathbf{P}_X(\omega) \text{ – presque sûrement.}$$

Elle est dite continue par rapport à sa première variable lorsque pour \mathbf{P}_X presque tout ω , la fonction $(\eta \mapsto \varphi(\eta, \omega))$ est continue pour la métrique d_E .

THÉORÈME 2.4 (Théorème de Loynes).– Si φ est \leq -croissante et continue, l'équation (2.7) admet une solution M_∞ à valeurs dans l'adhérence \bar{E} de E .

Démonstration. Souvenons-nous que nous supposons connu le stimulus par son quadruplet \mathcal{D} , dont l'élément générique est noté ω . On cherche Y , variable aléatoire à valeurs dans E , satisfaisant (2.7). On va obtenir Y comme limite d'une suite presque sûrement convergente.

Pour ce faire, on considère la suite de Loynes $(M_n, n \in \mathbf{N})$ définie par :

$$M_0(\omega) = \mathbf{0} \text{ et } M_{n+1}(\omega) = \varphi(M_n \circ \theta^{-1}(\omega), \theta^{-1}\omega), \forall n \geq 1. \quad (2.8)$$

Par définition de $\mathbf{0}$, on a $M_0 = \mathbf{0} \leq M_1$ et si l'on suppose que pour un certain $n > 1$, $M_{n-1} \leq M_n$ presque sûrement, d'après la croissance de φ , \mathbf{P}_X -presque sûrement, on a l'identité suivante :

$$M_n(\omega) = \varphi(M_{n-1}(\theta^{-1}\omega), \theta^{-1}\omega) \leq \varphi(M_n(\theta^{-1}\omega), \theta^{-1}\omega) = M_{n+1}(\omega).$$

Par conséquent, la suite $(M_n, n \in \mathbf{N})$ est presque sûrement croissante. Compte tenu des hypothèses sur les suites croissantes de E , elle converge donc presque sûrement vers la variable aléatoire $M_\infty = \sup_{n \in \mathbf{N}} M_n$, qui est à valeurs dans \bar{E} . Par continuité de φ , la relation deuxième relation de (2.8) implique que :

$$M_\infty(\omega) = \varphi(M_\infty \circ \theta^{-1}(\omega), \theta^{-1}\omega)$$

et comme θ est bijectif, on en déduit que M_∞ est solution de (2.7). \square

REMARQUE.– La suite M s'interprète en fait assez aisément. Soit $(W_n^0, n \in \mathbf{N})$, la SRS issue de 0 et portée par φ . Il est aisé de vérifier que pour tout $n \in \mathbf{N}$, on a :

$$M_n = W_n^0 \circ \theta^{-n}, \text{ presque sûrement.}$$

En effet, cette relation est vraie pour $n = 0$, et si elle l'est au rang n , alors presque sûrement, on a :

$$\begin{aligned} M_{n+1}(\omega) &= \varphi(M_n(\theta^{-1}\omega), \theta^{-1}\omega) \\ &= \varphi(W_n^0 \circ \theta^{-n} \circ \theta^{-1}\omega, \theta^n \circ \theta^{-(n+1)}\omega) \\ &= W_{n+1}^0(\theta^{-(n+1)}\omega). \end{aligned}$$

De manière concrète, M_n est la valeur à l'instant 0 de la suite W^0 , en partant de 0 à l'instant $-n$ et en utilisant comme stimulus les valeurs de $X_{-n}, X_{-n+1}, \dots, X_0$. On appelle pour cette raison la construction de Loynes un schéma de récurrence arrière. On remarquera au passage l'aisance apportée par la construction sur $F^{\mathbf{Z}}$ et non $F^{\mathbf{N}}$ du stimulus. L'idée sous-jacente est qu'en faisant partir la suite de $-\infty$, on aura atteint l'état stationnaire à l'instant 0.

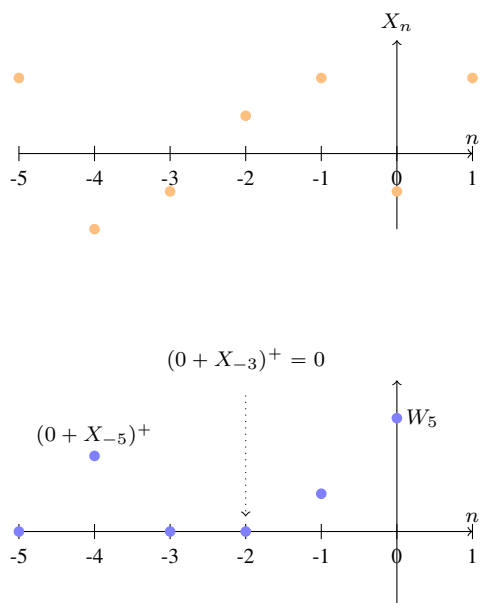


Figure 2.1 – Schéma de récurrence arrière pour $\varphi(x, z) = (x + z)^+$

EXEMPLE 2.4. – L'exemple 2.3 rentre tout à fait dans ce cadre puisque la fonction $(x \mapsto (x + z)^+)$ est évidemment continue et croissante pour tout $z \in \mathbf{R}$. Il existe donc une variable aléatoire Y solution de (2.7) mais on ne sait pas *a priori* si sa distribution est « propre », c'est-à-dire si $\mathbf{P}(Y = +\infty) = 0$. Ce sera l'un des objets d'étude du chapitre 4.

Nous pouvons maintenant répondre à la question de la convergence faible d'une SRS issue de l'état minimal.

COROLLAIRE 2.5.– Sous les hypothèses du théorème 2.4, la SRS $(W_n^0, n \in \mathbf{N})$ issue de $\mathbf{0}$ et portée par φ converge en loi vers M_∞ .

Démonstration. Soit F continue bornée de E dans \mathbf{R} , comme \mathbf{P}_X est invariante par θ , on a :

$$\mathbf{E}[F(W_n^0)] = \mathbf{E}[F(W_n^0 \circ \theta^{-n})] = \mathbf{E}[F(M_n)] \xrightarrow{n \rightarrow \infty} \mathbf{E}[F(M_\infty)],$$

d'où le résultat. \square

Le résultat suivant sera primordial dans les applications aux files d'attente.

THÉORÈME 2.6.– Sous les hypothèses du théorème 2.4, la solution M_∞ construite par le schéma de Loynes est \leq -minimale parmi les solutions de (2.7).

Démonstration. Soit Y , une solution de (2.7). On a $M_0 = \mathbf{0} \leq Y$, presque sûrement et $M_n \leq Y$, presque sûrement implique que l'on a :

$$M_{n+1}(\omega) \leq \varphi(Y \circ \theta^{-1}(\omega), \theta^{-1}\omega) = Y(\omega), \mathbf{P}_X - \text{presque sûrement,}$$

soit $M_{n+1} \leq Y$, presque sûrement. L'inégalité est conservée par passage à la limite presque sûre donc $M_\infty \leq Y$. \square

On peut appliquer les résultats précédents à $E = \mathbf{R}^+$, totalement ordonné par « \leq » et de point minimal 0. Dans ce cadre, le théorème ergodique de Birkhoff peut être vu comme une application fondamentale du résultat de Loynes.

THÉORÈME 2.7 (Théorème ergodique de Birkhoff).– Pour toute variable aléatoire réelle $Y \in \mathbf{L}^1(\mathbf{P})$, on a l'identité suivante :

$$\mathbf{E}[Y] = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n Y \circ \theta^{-i}, \mathbf{P} - \text{presque sûrement.}$$

Démonstration. Soit $\varepsilon > 0$. On définit la variable aléatoire :

$$Y^\varepsilon = Y - \mathbf{E}[Y] - \varepsilon$$

et l'application aléatoire de \mathbf{R}^+ dans lui même :

$$\varphi^\varepsilon : x \mapsto (x + Y^\varepsilon)^+ = x - x \wedge (-Y^\varepsilon).$$

La fonction φ^ε est presque sûrement croissante et continue, donc la suite de Loynes de condition initiale ε portée par $\varphi^\varepsilon(M_n^\varepsilon, n \in \mathbf{N})$ est presque sûrement croissante dans \mathbf{R}^+ et converge presque sûrement en vertu du théorème 2.4 vers la variable aléatoire M_∞^ε , qui vérifie :

$$M_\infty^\varepsilon \circ \theta = \varphi^\varepsilon(M_\infty^\varepsilon) = (M_\infty^\varepsilon + Y^\varepsilon)^+. \quad (2.9)$$

Une récurrence immédiate montre que les $(M_n^\varepsilon, n \in \mathbf{N})$ sont intégrables. Pour tout $n \in \mathbf{N}$, on a donc l'identité :

$$\begin{aligned} 0 \geq \mathbf{E}[M_n^\varepsilon] - \mathbf{E}[M_{n+1}^\varepsilon] &= \mathbf{E}[M_n^\varepsilon - M_{n+1}^\varepsilon \circ \theta] \\ &= \mathbf{E}[M_n^\varepsilon - \varphi^\varepsilon(M_n^\varepsilon)] = \mathbf{E}[M_n^\varepsilon \wedge (-Y^\varepsilon)]. \end{aligned}$$

Par convergence dominée, il vient :

$$\mathbf{E}[M_\infty^\varepsilon \wedge (-Y^\varepsilon)] \leq 0.$$

En vertu de (2.9), l'évènement $\{M_\infty^\varepsilon = +\infty\}$ est θ -invariant et est donc de probabilité 0 ou 1. Mais $M_\infty^\varepsilon = \infty$ presque sûrement impliquerait que $\mathbf{E}[-Y^\varepsilon] \leq 0$, ce qui est absurde. Par conséquent, M_∞^ε est presque sûrement fini. Soit maintenant l'application aléatoire de \mathbf{R}^+ dans lui-même, définie par $\tilde{\varphi}^\varepsilon : x \mapsto x + Y^\varepsilon$, et $(\tilde{M}_n^\varepsilon, n \in \mathbf{N})$, la suite de Loynes associée. Notons que par construction, $\tilde{M}_n^\varepsilon = \sum_{i=1}^n Y^\varepsilon \circ \theta^{-i}$ (en comprenant que $\sum_{i=1}^0 = 0$). De plus, comme $\tilde{\varphi}^\varepsilon(x) \leq \varphi^\varepsilon(x)$ pour tout x , il est facile de vérifier par récurrence que $\tilde{M}_n^\varepsilon \leq M_n^\varepsilon$ presque sûrement pour tout $n \in \mathbf{N}$. En particulier, on a $\tilde{M}_n^\varepsilon \leq M_\infty^\varepsilon$ presque sûrement pour tout $n \in \mathbf{N}$, ce qui revient à dire que :

$$\frac{1}{n} \sum_{i=1}^n Y \circ \theta^{-i} \leq \frac{1}{n} M_\infty^\varepsilon + \mathbf{E}[Y] + \varepsilon.$$

Ceci est vrai pour tout $\varepsilon > 0$, on a par conséquent :

$$\limsup_{n \rightarrow +\infty} \frac{1}{n} \sum_{i=1}^n Y \circ \theta^{-i} \leq \mathbf{E}[Y], \mathbf{P} - \text{presque sûrement.}$$

Cette dernière relation est également vérifiée par la variable aléatoire intégrable $-Y$, on a donc aussi :

$$\liminf_{n \rightarrow +\infty} \frac{1}{n} \sum_{i=1}^n Y \circ \theta^{-i} \geq \mathbf{E}[Y], \mathbf{P} - \text{presque sûrement,}$$

ce qui conclut la preuve. \square

REMARQUE.— Le quadruplet $(\Omega, \mathcal{F}, \mathbf{P}, \theta)$ est stationnaire ergodique si et seulement si $(\Omega, \mathcal{F}, \mathbf{P}, \theta^{-1})$ l'est aussi. On peut donc remplacer l'énoncé du théorème 2.7 par :

$$\mathbf{E}[Y] = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n Y \circ \theta^i, \mathbf{P} - \text{presque sûrement.}$$

2.3. Couplage

L'idée de couplage est centrale dans l'étude asymptotique des SRS. Il est en effet possible d'énoncer des conditions sous lesquelles les trajectoires de deux SRS (ou éventuellement celles des schémas de récurrence arrière correspondants) coïncident à partir d'un certain point. Ces propriétés impliquent naturellement, en particulier, des propriétés plus classiques de convergence pour les suites aléatoires, comme la convergence en loi.

Nous n'énonçons ici que les résultats qui nous seront utiles dans les applications aux files d'attente, sous leur forme la plus simple.

Dans un second temps, nous développons la théorie des événements de renouvellement de Borovkov qui donne des conditions suffisantes de couplage, et même de couplage arrière fort. Les résultats de Borovkov permettent par ailleurs de résoudre dans de nombreux cas l'équation (2.7), même lorsque les conditions de continuité et monotonie du théorème 2.4 ne sont pas satisfaites. En particulier, on pourra également traiter sous cet angle la question délicate du comportement transitoire en fonction des conditions initiales. Dans toute la suite, $\mathfrak{D} = (\Omega, \mathcal{F}, \mathbf{P}, \theta)$ désigne un système dynamique stationnaire ergodique.

2.3.1. Définition

Nous commençons par définir les différents types de couplage.

DÉFINITION 2.6.– Soient $(W_n, n \in \mathbf{N})$ et $(Y_n, n \in \mathbf{N})$ deux suites aléatoires définies sur $(\Omega, \mathcal{F}, \mathbf{P}, \theta)$.

– On dit que $(W_n, n \in \mathbf{N})$ et $(Y_n, n \in \mathbf{N})$ couplent si :

$$\mathbf{P}(W_n = Y_n; \forall n \geq N) \xrightarrow{N \rightarrow \infty} 1. \quad (2.10)$$

– On dit qu'il y a couplage arrière fort entre $(W_n, n \in \mathbf{N})$ et $(Y_n, n \in \mathbf{N})$ si :

$$\mathbf{P}(W_n \circ \theta^{-n} = Y_n \circ \theta^{-n}; \forall n \geq N) \xrightarrow{N \rightarrow \infty} 1. \quad (2.11)$$

Dans la suite, nous noterons :

$$\begin{aligned} \tau_F &= \inf \{N \in \mathbf{N}; W_n = Y_n, \forall n \geq N\}, \\ \tau_B &= \inf \{N \in \mathbf{N}; W_n \circ \theta^{-n} = Y_n \circ \theta^{-n}, \forall n \geq N\}, \end{aligned}$$

les indices (aléatoires) de couplage respectifs vers l'avant (*forward*) et vers l'arrière (*backward*) des deux suites, en comprenant bien sûr comme infinie ces variables

aléatoires lorsque l'ensemble de droite est vide. On voit alors facilement que le couplage vers l'avant et le couplage arrière fort de $(W_n, n \in \mathbf{N})$ avec $(Y_n, n \in \mathbf{N})$ admettent respectivement les définitions équivalentes suivantes :

$$\mathbf{P}(\tau_F \geq N) \xrightarrow{N \rightarrow \infty} 0, \quad (2.12)$$

$$\mathbf{P}(\tau_B \geq N) \xrightarrow{N \rightarrow \infty} 0. \quad (2.13)$$

On remarque déjà un lien immédiat entre couplage et convergence en loi.

THÉORÈME 2.8.— Si la suite $(W_n, n \in \mathbf{N})$ à valeurs dans E couple avec la suite stationnaire $(Y \circ \theta^n, n \in \mathbf{N})$ alors $(W_n, n \in \mathbf{N})$ converge en loi vers Y .

Démonstration. Soit G , une fonction continue et bornée : $E \rightarrow \mathbf{R}$, et soit $\|G\|_\infty$ son supremum. Pour tout entier N , on a :

$$\begin{aligned} \left| \mathbf{E}[G(W_N)] - \mathbf{E}[G(Y)] \right| &= \left| \mathbf{E}[G(W_N)] - \mathbf{E}[G(Y) \circ \theta^N] \right| \\ &\leq \mathbf{E} \left[\left| G(W_N) - G(Y \circ \theta^N) \right| \right] \\ &= \mathbf{E} \left[\left| G(W_N) - G(Y \circ \theta^N) \right| \mathbf{1}_{\tau_F > N} \right] \\ &\leq 2 \|G\|_\infty \mathbf{P}(\tau_F > N) \end{aligned}$$

et la quantité de droite tend vers 0 par hypothèse. \square

Notons :

$$\tau_f = \begin{cases} \infty & \text{si } W_n \neq Y_n \forall n \in \mathbf{N}, \\ \inf \{n \in \mathbf{N}, W_n = Y_n\} & \text{sinon,} \end{cases}$$

le premier instant où les deux suites $(W_n, n \in \mathbf{N})$ et $(Y_n, n \in \mathbf{N})$ coïncident.

REMARQUE.— On peut remarquer que deux SRS portées par la même fonction de récurrence φ (nous les notons alors $(W_n^Y, n \in \mathbf{N})$ et $(W_n^Z, n \in \mathbf{N})$, seules leurs variables aléatoires initiales les différencient éventuellement) couplent dès l'indice τ_f . En effet, presque sûrement $W_n^Y(\omega) = W_n^Z(\omega)$ implique que :

$$W_{n+1}^Y(\omega) = \varphi(W_n^Y(\omega), \theta^n \omega) = \varphi(W_n^Z(\omega), \theta^n \omega) = W_{n+1}^Z(\omega).$$

Par conséquent, pour tout $n \in \mathbf{N}$, $\{\tau_f \leq n\} \subset \{\tau_F \leq n\}$, soit :

$$\tau_F = \tau_f \text{ presque sûrement.} \quad (2.14)$$

En revanche, les deux suites $(W^Y \circ \theta^{-n}, n \in \mathbf{N})$ et $(W^Z \circ \theta^{-n}, n \in \mathbf{N})$ peuvent coïncider pour un certain indice, sans pour autant qu'il y ait couplage arrière fort, puisque pour tout ω tel que $W_n^Y(\theta^{-n}\omega) = W_n^Z(\theta^{-n}\omega)$, on a :

$$\begin{aligned} W_{n+1}^Y(\theta^{-(n+1)}\omega) &= \varphi(W_n^Y(\theta^{-(n+1)}\omega), \theta^n(\theta^{-(n+1)}\omega)) \\ &= \varphi(W_n^Y(\theta^{-(n+1)}\omega), \theta^{-1}\omega) \\ W_{n+1}^Z(\theta^{-(n+1)}\omega) &= \varphi(W_n^Z(\theta^{-(n+1)}\omega), \theta^{-1}\omega) \end{aligned}$$

et ces deux quantités ne sont pas égales en général.

Dans le cas des SRS, le lien entre les différents types de couplage est établi dans les théorèmes suivants.

THÉORÈME 2.9.– Soient Z et Y , deux variables aléatoires à valeurs dans E et la SRS $(W_n^Z, n \in \mathbf{N})$, une SRS issue de Z et portée par φ . S'il y a couplage arrière fort entre $(W_n^Z, n \in \mathbf{N})$ et la suite stationnaire $(Y \circ \theta^n, n \in \mathbf{N})$ alors ces deux suites couplent, $(W_n^Z, n \in \mathbf{N})$ converge en loi vers Y et Y est une solution de (2.7).

Démonstration. Tout d'abord, pour tout $\omega \in \theta^{-1}(\{\tau_B < \infty\})$ (c'est-à-dire tel que $\tau_B(\theta\omega) < \infty$), pour tout $n \geq \tau_B(\theta\omega)$, on a :

$$\begin{aligned} Y \circ \theta(\omega) &= W_{n+1}^Z \circ \theta^{-(n+1)}(\theta\omega) \\ &= \varphi(W_n^Z \circ \theta^{-n}(\omega), \theta^n \circ \theta^{-n}\omega) \\ &= \varphi(Y(\omega), \omega). \end{aligned}$$

L'événement $\theta^{-1}(\tau_B < \infty)$ est de probabilité 1 par hypothèse, on a donc bien :

$$Y \circ \theta(\omega) = \varphi(Y(\omega), \omega), \text{ presque sûrement.} \quad (2.15)$$

Par ailleurs, pour tout $N \in \mathbf{N}$, on a :

$$\begin{aligned} \mathbf{P}(\tau_B \leq N) &= \mathbf{P}(W_n^Z \circ \theta^{-n} = Y, \forall n \geq N) \\ &= \mathbf{P}\left(\theta^{-N} \left\{ W_n^Z \circ \theta^{-n} = Y, \forall n \geq N \right\}\right) \\ &= \mathbf{P}(W_n^Z \circ \theta^{N-n} = Y \circ \theta^N, \forall n \geq N) \\ &\leq \mathbf{P}(W_N^Z = Y \circ \theta^N), \\ &\leq \mathbf{P}(\tau_f \leq N). \end{aligned} \quad (2.16)$$

Mais d'après (2.15), $(Y \circ \theta^n, n \in \mathbf{N}) \equiv (W_n^Y, n \in \mathbf{N})$, la SRS issue de Y et portée par φ , et donc on est dans le cas de la remarque 2.3.1. D'après (2.14), $\tau_f = \tau_F$ presque sûrement dans ce cas et donc (2.16) implique :

$$\mathbf{P}(\tau_B > N) \geq \mathbf{P}(\tau_F > N).$$

Le terme de droite tend donc vers 0 quand N tend vers l'infini comme le terme de gauche, ce qui montre la propriété de couplage. Finalement, la convergence en loi découle du théorème 2.8. \square

THÉORÈME 2.10.— Soit $(W_n^0, n \in \mathbf{N})$ une SRS à valeurs dans E issue de $\mathbf{0}$ et portée par φ , une application presque sûrement \leq -croissante et continue. Soit W , une solution de l'équation (2.7) correspondant à φ , à valeurs dans E . Les suites $(W_n^0, n \in \mathbf{N})$ et $(W \circ \theta^n, n \in \mathbf{N})$ admettent un couplage avant si et seulement si elles admettent un couplage arrière fort.

Démonstration. Nous allons en fait montrer que τ_F et τ_B ont ici la même loi. Notons comme précédemment, $(M_n, n \in \mathbf{N}) = (W_n^0 \circ \theta^{-n}, n \in \mathbf{N})$ la suite de Loynes correspondante, et rappelons que cette suite croît presque sûrement vers $M_\infty \leq W$. En particulier, presque sûrement $M_N(\omega) = W(\omega)$ implique que $M_n(\omega) = M_\infty(\omega) = W(\omega)$. Pour tout entier $N \in \mathbf{N}$, on a donc :

$$\begin{aligned} \mathbf{P}(\tau_F \leq N) &= \mathbf{P}(W_N^0 = W \circ \theta^N) \\ &= \mathbf{P}(M_N = W) \\ &= \mathbf{P}(M_n = W, \forall n \geq N) \\ &= \mathbf{P}(W_n^0 \circ \theta^{-n} = W, \forall n \geq N) \\ &= \mathbf{P}(\tau_N \leq N). \end{aligned}$$

Le résultat est démontré. \square

2.3.2. Événements de renouvellement

La théorie des événements de renouvellement donne des critères simples pour le couplage arrière fort d'une SRS avec une solution de (2.7). Dans toute cette sous-section, $(W_n, n \in \mathbf{N})$ désigne une SRS définie sur $\mathfrak{D} = (\Omega, \mathcal{F}, \mathbf{P}, \theta)$, à valeurs dans E et portée par φ .

DÉFINITION 2.7.— Soit N un entier strictement positif. On dit que la suite d'événements mesurables $(\mathfrak{A}_n, n \in \mathbf{N})$ est une suite d'événements de renouvellement de longueur N pour $(W_n, n \in \mathbf{N})$ si et seulement s'il existe une variable aléatoire α

définie sur Ω et à valeurs dans un espace mesurable F , et une application déterministe $\Phi : F^N \rightarrow E$ telle que pour tout $n \geq N$, sur \mathfrak{A}_{n-N} :

$$W_n = \Phi (\alpha \circ \theta^{n-N}, \alpha \circ \theta^{n-2}, \dots, \alpha \circ \theta^{n-1}).$$

La propriété d'admettre une suite d'événements de renouvellement est donc une propriété d'« oublier » : sur \mathfrak{A}_{n-N} , W_n ne dépend que d'une liste de N valeurs de la suite stationnaire $(\alpha \circ \theta^n, n \in \mathbf{N})$. De manière imagée, si un événement donné se produit N pas de temps dans le passé, W_n ne dépend de son passé que jusqu'à l'instant $n - N$, pas au-delà.

EXEMPLE 2.5.— On suppose que $(W_n, n \in \mathbf{N})$ est une SRS portée par φ . Soit $x \in E$ et pour tout n , $\mathfrak{A}_n = \{W_{n-1} = x\}$. Il est facile de voir qu'alors, $(\mathfrak{A}_n, n \in \mathbf{N})$ est une suite d'événements de renouvellement de longueur 1. En effet, pour tout $n \geq 1$, sur \mathfrak{A}_{n-1} , $W_n = \varphi \circ \theta^{n-1}(x)$, et l'on retrouve la définition en prenant $F = E$, Φ l'identité sur E et $\alpha = \varphi(x)$. Dans les applications (chapitre 4), on considèrera essentiellement ce type d'événements de renouvellement dans le cas particulier $x = 0$.

DÉFINITION 2.8.— Une suite d'événements $(\mathfrak{A}_n, n \in \mathbf{N})$ est dite θ -compatible, si pour tout $n \geq 0$, $\mathfrak{A}_n = \theta^{-n}\mathfrak{A}_0$. On peut alors définir la suite $(\mathfrak{A}_n, n \in \mathbf{Z})$ en posant $\mathfrak{A}_{-n} = \theta^n\mathfrak{A}_0$ pour tout $n \geq 0$.

EXEMPLE 2.6.— Soit β , une variable aléatoire définie sur Ω et à valeurs dans (E, \mathcal{E}) . Alors, pour tout $\mathcal{B} \in \mathcal{E}$, la suite d'événements définie par $\mathfrak{A}_n = \{\beta \circ \theta^n \in \mathcal{B}\}$ pour tout n est θ -compatible, puisque pour tout n , $\omega \in \mathfrak{A}_n$ revient à $\beta \circ \theta^n \omega \in \mathcal{B}$, soit $\beta \circ \theta^{n+1}(\theta^{-1}\omega) \in \mathcal{B}$, ou encore $\omega \in \theta\mathfrak{A}_{n+1}$. La suite des images réciproques d'un ensemble mesurable par une suite stationnaire est donc logiquement une suite θ -compatible.

Le théorème suivant est dû à Borovkov.

THÉORÈME 2.11.— Soit \mathcal{Z} , une famille de variables aléatoires à valeurs dans E . On suppose que toutes les suites $(W_n^Z, n \in \mathbf{N})$, $Z \in \mathcal{Z}$, admettent la même suite d'événements de renouvellement $(\mathfrak{A}_n, n \in \mathbf{N})$, de même longueur N , de même variable aléatoire associée α et de même application associée Φ , et que la suite $(\mathfrak{A}_n, n \in \mathbf{N})$ est θ -compatible et telle que $\mathbf{P}(\mathfrak{A}_0) > 0$. Il existe alors une variable aléatoire finie W telle que pour tout $Z \in \mathcal{Z}$, il y a couplage arrière fort entre les suites $(W_n^Z, n \in \mathbf{N})$ et $(W \circ \theta^n, n \in \mathbf{N})$.

Démonstration. Soit $n \geq N$, $i \in \llbracket 0, \dots, n - N \rrbracket$ et $k \geq 0$. Par θ -compatibilité :

$$\mathfrak{A}_{-N-i} = \bigcap_{k \geq 0} \theta^{n+k} \mathfrak{A}_{n+k-N-i}$$

et on a donc pour toute variable aléatoire $Z \in \mathcal{Z}$, pour tout $\omega \in \mathfrak{A}_{-N-i}$ et pour tout $k \geq 0$:

$$\begin{aligned} \theta^{-(n+k)}\omega &\in \mathfrak{A}_{n+k-N-i} \\ \iff W_{n+k-i}^Z(\theta^{-(n+k)}\omega) &= \Phi\left(\alpha \circ \theta^{n+k-i-N}(\theta^{-(n+k)}\omega), \dots, \alpha \circ \theta^{n+k-i-1}(\theta^{-(n+k)}\omega)\right) \\ \iff W_{n+k-i}^Z \circ \theta^{-(n+k)}(\omega) &= \Phi\left(\alpha \circ \theta^{i-N}(\omega), \dots, \alpha \circ \theta^{i-1}(\omega)\right), \end{aligned}$$

une variable aléatoire qui ne dépend ni de Z , ni de k . Ainsi, pour tout couple de conditions initiales Z et $Z' \in \mathcal{Z}$, sur \mathfrak{A}_{-N-i} , on a :

$$W_{n+k-i}^{Z'} \circ \theta^{-(n+k)} = W_{n-i}^Z \circ \theta^{-n}, \quad \forall k \geq 0.$$

Par conséquent :

$$\begin{aligned} W_{n+1+k-i}^{Z'} \circ \theta^{-(n+k)} &= \varphi \circ \theta^{n+k-i} \circ \theta^{-(n+k)} \left(W_{n+k-i}^{Z'} \circ \theta^{-(n+k)} \right) \\ &= \varphi \circ \theta^{n-i} \circ \theta^{-n} \left(W_{n-i}^Z \circ \theta^{-n} \right) \\ &= W_{n+1-i}^Z \circ \theta^{-n}, \end{aligned}$$

ce qui implique par une récurrence immédiate jusqu'au rang i que :

$$W_{n+k}^{Z'} \circ \theta^{-(n+k)} = W_{n+i+k-i}^{Z'} \circ \theta^{-(n+k)} = W_{n+i-i}^Z \circ \theta^{-n} = W_n^Z \circ \theta^{-n}.$$

Ainsi, sur \mathfrak{A}_{-N-i} les suites $(W_n^Z \circ \theta^{-n}, n \in \mathbf{N})$ et $(W_n^{Z'} \circ \theta^{-n}, n \in \mathbf{N})$ sont constantes et confondues après n . Ceci est vrai pour tout $i \in \{0, \dots, n-N\}$, donc en notant pour tout $Z \in \mathcal{Z}$, τ_B^Z l'instant de couplage arrière fort de la suite $(W_n^Z, n \in \mathbf{N})$, on a :

$$\begin{aligned} \bigcup_{i=0}^{n-N} \mathfrak{A}_{-i-N} \\ \subseteq \mathcal{B}_n = \left\{ \tau_B^Z \leq n, \forall Z \in \mathcal{Z} \right\} \cap \left\{ W_n^Z \circ \theta^{-n} = W_n^{Z'} \circ \theta^{-n}, \forall Z, Z' \in \mathcal{Z} \right\}. \end{aligned}$$

Ceci implique que :

$$\mathfrak{A} = \bigcup_{j=N}^{+\infty} \mathfrak{A}_{-j} = \bigcup_{n=N}^{+\infty} \bigcup_{i=N}^n \mathfrak{A}_{-i} \subseteq \bigcup_{n=N}^{+\infty} \mathcal{B}_n.$$

Mais $\mathfrak{A} \subseteq \theta\mathfrak{A}$, alors que $\mathfrak{A}_{-N} \subseteq \mathfrak{A}$ a une probabilité strictement positive par hypothèse. L'événement \mathfrak{A} est donc de probabilité 1. Donc l'événement $\bigcup_{n=N}^{+\infty} \mathcal{B}_n$ est presque sûr, ce qui signifie que toutes les suites $(W_n^Z, n \in \mathbf{N})$, $Z \in \mathcal{Z}$ couplent (au sens du couplage arrière fort), et avec la même variable aléatoire puisque les suites $(W_n^Z \circ \theta^{-n}, n \in \mathbf{N})$, $Z \in \mathcal{Z}$ sont confondues à partir d'un certain rang. Le théorème est démontré. \square

Un corollaire immédiat, mais crucial de ce théorème est la donnée d'une condition suffisante pour l'existence d'une solution à l'équation stationnaire (2.7).

COROLLAIRE 2.12.— S'il existe un ensemble de conditions \mathcal{Z} non vide satisfaisant les hypothèses du théorème 2.11, l'équation (2.7) admet une solution à valeurs dans E .

Démonstration. Pour tout $Z \in \mathcal{Z}$, il y a couplage arrière fort pour $(W_n^Z, n \in \mathbf{N})$ d'après le théorème 2.11. Ceci implique le résultat en vertu de l'assertion (ii) du théorème 2.9. \square

Le deuxième corollaire suivant est un critère pour l'unicité d'une solution à (2.7).

COROLLAIRE 2.13.— Si l'ensemble :

$$\mathcal{Z} = (\text{solutions de (2.7) à valeurs dans } E)$$

est non vide et vérifie le théorème 2.11, il est réduit à un singleton.

Démonstration. Pour tout couple de solutions $Z, Z' \in \mathcal{Z}$, les deux suites stationnaires $(W_n^Z, n \in \mathbf{N}) = (Z \circ \theta^n, n \in \mathbf{N})$ et $(W_n^{Z'}, n \in \mathbf{N}) = (Z' \circ \theta^n, n \in \mathbf{N})$ couplent pour le couplage arrière fort avec la même suite stationnaire $(W \circ \theta^n, n \in \mathbf{N})$. On a donc naturellement $Z = Z' = W$, \mathbf{P} -presque sûrement. \square

2.4. Comparaison des suites récurrentes stochastiques

Nous donnons dans cette dernière section deux résultats remarquables de comparaison pour les SRS, que nous appliquerons aux modèles de files d'attente dans le chapitre 4. Dans toute cette section, les espaces euclidiens \mathbf{R}^K , $K \geq 1$ sont munis de l'ordre partiel $<$ défini dans l'appendice A.

DÉFINITION 2.9.— Soient W et Y deux variables aléatoires à valeurs dans \mathbf{R}^k , définies éventuellement sur deux espaces de probabilité différents Ω et $\hat{\Omega}$. On dit que Y majore W stochastiquement ou pour l'ordre fort, et on note $W \leq_{\text{st}} Y$, si pour toute fonction $<$ -croissante $F : \mathbf{R}^k \rightarrow \mathbf{R}$ telle que les intégrales suivantes existent :

$$\mathbf{E}[F(W)] \leq \hat{\mathbf{E}}[F(Y)].$$

On peut remarquer en particulier que si W et Y sont des variables aléatoires réelles :

$$W \leq_{\text{st}} Y \iff \mathbf{P}(W \leq x) \geq \hat{\mathbf{P}}(Y \leq x) \text{ pour tout } x \in \mathbf{R}. \quad (2.17)$$

Le théorème suivant est le résultat fondamental de la théorie de la comparaison stochastique.

THÉORÈME 2.14 (Théorème de Strassen).— Soient W et Y deux variables aléatoires à valeurs dans \mathbf{R}^k . Alors $W \leq_{\text{st}} Y$ si et seulement si il existe un espace de probabilité sur lequel sont définies deux variables aléatoires \tilde{W} et \tilde{Y} , de mêmes lois respectives que W et Y sur \mathbf{R}^d et telles que :

$$\tilde{W} < \tilde{Y} \text{ presque sûrement.}$$

THÉORÈME 2.15.— Soient α et $\bar{\alpha}$ deux variables aléatoires définies sur le quadruplet stationnaire ergodique $(\Omega, \mathcal{F}, \mathbf{P}, \theta)$, à valeurs dans \mathbf{R}^m et intégrables. Soit f , une application déterministe : $\mathbf{R}^d \times \mathbf{R}^m \rightarrow \mathbf{R}^d$. On note $(W_n, n \in \mathbf{N})$ et $(\bar{W}_n, n \in \mathbf{N})$ les deux SRS à valeurs dans \mathbf{R}^d et issues de 0, portées respectivement par les applications aléatoires :

$$x \mapsto f(x, \alpha) \text{ et } x \mapsto f(x, \bar{\alpha}).$$

Si $\bar{\alpha} \leq_{\text{st}} \alpha$ et que f est croissante en ses deux arguments alors $\bar{W}_n \leq_{\text{st}} W_n$ pour tout $n \in \mathbf{N}$.

Démonstration. D'après le théorème de Strassen, il existe un espace de probabilité sur lequel sont définies les suites $(\alpha_n, n \in \mathbf{N})$ et $(\bar{\alpha}_n, n \in \mathbf{N})$, telles que l'on ait les égalités en loi :

$$\alpha_n \stackrel{\mathcal{L}}{=} \alpha \circ \theta^n \text{ et } \bar{\alpha}_n \stackrel{\mathcal{L}}{=} \bar{\alpha} \circ \theta^n, \text{ pour tout } n \in \mathbf{N},$$

et telles que :

$$\bar{\alpha}_n < \alpha_n \text{ presque sûrement pour tout } n \in \mathbf{N}.$$

Soient $(Y_n, n \in \mathbf{N})$ et $(\bar{Y}_n, n \in \mathbf{N})$, les deux SRS définies sur ce nouvel espace, issues de 0 et portées respectivement par les suites d'applications $(f(\cdot, \alpha_n), n \in \mathbf{N})$ et $(f(\cdot, \bar{\alpha}_n), n \in \mathbf{N})$. Montrons par récurrence sur n que :

$$\bar{Y}_n < Y_n, \text{ presque sûrement pour tout } n \in \mathbf{N}.$$

On a $Y_0 = \bar{Y}_0 = 0$ presque sûrement et si $\bar{Y}_n < Y_n$ pour un certain $n \in \mathbf{N}$, alors par monotonie de f , on a presque sûrement :

$$\begin{aligned}\bar{Y}_{n+1} &= f(\bar{Y}_n, \bar{\alpha}_n) \\ &< f(Y_n, \bar{\alpha}_n) \\ &< f(Y_n, \alpha_n) \\ &= Y_{n+1}.\end{aligned}$$

Par conséquent, $\bar{Y}_n < Y_n$ presque sûrement pour tout $n \in \mathbf{N}$ sur le nouvel espace de probabilité. Comme \bar{Y}_n (respectivement Y_n) a clairement même loi que \bar{W}_n (respectivement W_n) pour tout n , la réciproque du théorème de Strassen permet de conclure. \square

Supposons maintenant, de plus, que l'application f est continue en sa première variable. D'après le théorème de Loynes, les suites de Loynes $(W_n \circ \theta^{-n}, n \in \mathbf{N})$ et $(\bar{W}_n \circ \theta^{-n}, n \in \mathbf{N})$ convergent presque sûrement vers les versions stationnaires minimales respectives W_∞ et \bar{W}_∞ des deux SRS.

COROLLAIRE 2.16.— Sous les hypothèses du théorème 2.15, si le théorème de Loynes s'applique aux deux SRS et si les solutions minimales W_∞ et \bar{W}_∞ sont presque sûrement finies, elles vérifient :

$$\bar{W}_\infty \leq_{\text{st}} W_\infty.$$

Démonstration. D'après le théorème 2.15, pour toute fonction $F : E \rightarrow \mathbf{R}$ croissante telle que les espérances suivantes existent, on a :

$$\mathbf{E}[F(\bar{W}_n) \circ \theta^{-n}] = \mathbf{E}[F(\bar{W}_n)] \leq \mathbf{E}[F(W_n)] = \mathbf{E}[F(W_n) \circ \theta^{-n}],$$

par θ -invariance. On conclut par convergence monotone. \square

Le théorème suivant est le cas particulier (qui nous sera utile sous cette forme au chapitre suivant) d'un résultat plus général, qui décline la propriété de comparaison du théorème 2.15 pour un autre ordre stochastique impliquant les fonctions-test convexes, en appliquant un corollaire du théorème de Strassen pour cet ordre. Dans la suite, si Y est une variable aléatoire à valeurs dans \mathbf{R}^k , $k \geq 1$ et \mathcal{A} est une tribu, on note classiquement :

$$\mathbf{E}[Y | \mathcal{A}] = (\mathbf{E}[Y(1) | \mathcal{A}], \mathbf{E}[Y(2) | \mathcal{A}], \dots).$$

THÉORÈME 2.17.— On suppose que l'application f est $<$ -croissante en son premier argument, et convexe de $\mathbf{R}^d \times \mathbf{R}^m$ dans \mathbf{R}^d . Supposons en outre qu'il existe une filtration $(\mathcal{F}_n, n \in \mathbf{N})$ telle que pour tout $n \in \mathbf{N}$ et pour tout $i \in \llbracket 0, n \rrbracket$, on ait :

$$\bar{\alpha} \circ \theta^i = \mathbf{E}[\alpha \circ \theta^i | \mathcal{F}_n].$$

Sous ces hypothèses, pour toute fonction $F : \mathbf{R}^d \rightarrow \mathbf{R}$, \prec -croissante et convexe, telle que pour tout entier n , $F(\bar{W}_n) \in L^1$ et $F(W_n) \in L^1$, on a :

$$\mathbf{E} [F(\bar{W}_n)] \leq \mathbf{E} [F(W_n)] \text{ pour tout } n \in \mathbf{N}. \quad (2.18)$$

Démonstration. Fixons $n \in \mathbf{N}$. Montrons par récurrence sur $i \in \llbracket 0, n \rrbracket$ la relation :

$$\bar{W}_i < \mathbf{E} [W_i | \mathcal{F}_n] \text{ presque sûrement pour tout } i \in \llbracket 0, n \rrbracket, \quad (2.19)$$

qui est bien sûr vérifiée pour $i = 0$. En supposant qu'elle est vérifiée pour un certain $i \in \llbracket 0, n \rrbracket$, l'inégalité de Jensen nous donne l'identité suivante, valable presque sûrement :

$$\begin{aligned} \mathbf{E} [W_{i+1} | \mathcal{F}_n] &= \mathbf{E} [f(W_i, \alpha \circ \theta^i) | \mathcal{F}_n] \\ &> f(\mathbf{E} [W_i | \mathcal{F}_n], \mathbf{E} [\alpha \circ \theta^i | \mathcal{F}_n]) \\ &> f(\bar{W}_i, \mathbf{E} [\alpha \circ \theta^i | \mathcal{F}_n]) \\ &= f(\bar{W}_i, \bar{\alpha} \circ \theta^i) \\ &= \bar{W}_{i+1}, \end{aligned}$$

et (2.19) est démontré. En particulier, on a :

$$\bar{W}_n < \mathbf{E} [W_n | \mathcal{F}_n] \text{ presque sûrement.}$$

Compte-tenu de l'inégalité de Jensen, pour toute fonction $F : \mathbf{R}^d \rightarrow \mathbf{R}$ croissante et convexe, si les intégrales sont bien définies, on a :

$$\begin{aligned} F(\bar{W}_n) &\leq F(\mathbf{E} [W_n | \mathcal{F}_n]) \\ &\leq \mathbf{E} [F(W_n) | \mathcal{F}_n], \end{aligned}$$

et l'on conclut en passant aux espérances. \square

Comme pour le corollaire 2.16, on en déduit le résultat suivant.

COROLLAIRE 2.18.— Sous les hypothèses du théorème 2.17, si de plus le théorème de Loynes s'applique aux deux SRS et si les solutions minimales \bar{W}_∞ et W_∞ sont presque sûrement finies, elles vérifient :

$$\mathbf{E} [F(\bar{W}_\infty)] \leq \mathbf{E} [F(W_\infty)],$$

pour toute fonction $F : \mathbf{R}^d \rightarrow \mathbf{R}$ croissante et convexe, telle que les espérances sont bien définies.

2.5. Notes et commentaires

Le théorème de Loynes a été introduit dans [LOY 62] dans le cas particulier de la file $G/G/1$. Il a été généralisé sous la forme présentée ici, par exemple dans [BAC 02] et [NEV 84].

Pour plus de renseignements sur la notion de couplage, le lecteur pourra consulter [BRA 90] et [THO 00].

La théorie des événements de renouvellement est due à Borovkov et Foss. Elle est introduite dans [BOR 84], puis développée dans [BOR 92, BOR 94] et [BOR 98].

Pour un tour plus complet sur les comparaisons de suites récurrentes stochastiques, nous renvoyons le lecteur vers les références [BAC 02] et [STO 83]. La construction que nous utilisons ici est due à Baccelli et Makowski [BAC 89], nous n'en donnons qu'une version simplifiée des résultats.

Résumé

– Une suite récurrente stochastique est de la forme $X_{n+1} = f(X_n, \alpha_n)$, où (α_n) est une suite équadistribuée et ergodique.

– L'existence d'une loi stationnaire pour (X_n) revient à celle d'une variable aléatoire X résolvant sur l'espace canonique de (α_n) l'équation presque sûre :

$$X \circ \theta = f(X, \alpha),$$

où θ est l'opérateur bijectif de décalage.

– Le schéma de récurrence arrière de Loynes garantit l'existence d'une solution X (éventuellement infinie) si f est presque sûrement croissante et continue en sa première variable.

– La théorie des événements de renouvellement de Borovkov permet d'obtenir des conditions d'existence et d'unicité d'une solution X , et de couplage de la SRS avec sa version stationnaire.

– Le théorème de Strassen permet de comparer les solutions stationnaires d'une SRS en fonction de l'ordre relatif des suites qui la dirigent.

Chapitre 3

Chaînes de Markov

Lorsque l'on veut rendre compte de l'évolution temporelle d'un système, il faut décrire la façon dont le futur dépend du présent ou du passé. Parmi les principaux exemples, citons celui des équations différentielles (en temps continu) et celui des suites récurrentes (en temps discret). Dans le cas aléatoire à temps discret, l'analogie des suites récurrentes a été traité dans le chapitre 2 dans le cas stationnaire. Un cas particulier, sous des hypothèses supplémentaires d'indépendance, est celui des chaînes de Markov. Derrière une description apparemment simple se cache un outil mathématique redoutable d'efficacité pour les applications et riche de nombreuses propriétés.

On rappelle qu'il est conseillé de lire le petit paragraphe sur les notations de la remarque A.1.1.

3.1. Définition et exemples

Considérons donc une suite de variables aléatoires $X = (X_n, n \geq 0)$ à valeurs dans E fini ou dénombrable et indexée par les entiers, et la filtration $\mathcal{F}_n = \sigma\{X_j, 0 \leq j \leq n\}$ engendrée par cette suite.

Les trajectoires sont des éléments de $E^{\mathbb{N}}$, c'est-à-dire des suites d'éléments de E . Le décalage (voir l'annexe A) est alors défini par :

$$\begin{aligned} \theta : E^{\mathbb{N}} &\longrightarrow E^{\mathbb{N}} \\ (x_0, x_1, \dots) &\longmapsto (x_1, x_2, \dots). \end{aligned}$$

Le décalage est donc la restriction à $E^{\mathbb{N}}$, non bijective, du flot bijectif défini sur $E^{\mathbb{Z}}$ et introduit en section 2.1. Comme pour le flot, on aura souvent besoin du n^{e} itéré

de θ , noté θ^n et défini par :

$$\begin{aligned} \theta^n : E^{\mathbf{N}} &\longrightarrow E^{\mathbf{N}} \\ (x_0, x_1, \dots) &\longmapsto (x_n, x_{n+1}, x_{n+2}, \dots). \end{aligned}$$

Par la suite, on identifiera θ et θ^1 .

DÉFINITION 3.1.– La suite X est une chaîne de Markov lorsque pour tout $n \leq m$, la tribu \mathcal{F}_n est indépendante de la tribu $\sigma(X_m)$ conditionnellement à $\sigma(X_n)$. En d'autres termes, pour toute fonction F et G bornées, on a :

$$\begin{aligned} \mathbf{E}[F(X_0, \dots, X_n) G(X_m) | X_n] \\ = \mathbf{E}[F(X_0, \dots, X_n) | X_n] \mathbf{E}[G(X_m) | X_n]. \end{aligned} \quad (3.1)$$

En vertu du théorème A.12, on sait que cette propriété est équivalente à l'indépendance du passé et du futur conditionnellement au présent, ce qui s'exprime par :

$$\mathbf{E}[F(X_0, \dots, X_n) G \circ \theta^n | \mathcal{F}_n] = F(X_0, \dots, X_n) \mathbf{E}[G \circ \theta^n | X_n]. \quad (3.2)$$

En particulier, pour $G = \mathbf{1}_{\{y\}}(X_1)$, pour tout entier n , on obtient :

$$\mathbf{P}(X_{n+1} = y | \mathcal{F}_n) = \mathbf{P}(X_{n+1} = y | X_n).$$

DÉFINITION 3.2.– La chaîne de Markov X est dite homogène lorsque :

$$\mathbf{P}(X_{n+1} = y | X_n = x)$$

ne dépend pas de n mais seulement de x et y . On notera cette quantité $p(x, y)$ et on appelle $P = (\mathbf{P}(X_1 = y | X_0 = x), x, y \in E)$ l'opérateur de transition de X . Si E est de cardinal fini, P s'identifie à une matrice qui a autant de lignes et de colonnes que d'éléments dans E .

EXEMPLE 3.1.– Un rat se déplace dans le labyrinthe à sept cases représenté dans la figure 3.1. Il passe d'une case à l'autre uniformément suivant les possibilités qui lui sont offertes, c'est-à-dire que lorsqu'il y a deux (respectivement trois) sorties dans la case où il se trouve, il va dans chacune des cases possibles avec une probabilité d'un demi (respectivement d'un tiers). Son évolution est sans mémoire : chaque changement ne dépend que de la situation courante, pas du passé. On appelle X_n la position du rat après son n^{e} mouvement, X_0 est sa position initiale.

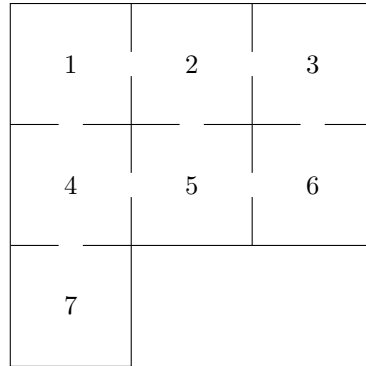


Figure 3.1 – Le labyrinthe

Ici, $E = \{1, 2, 3, 4, 5, 6, 7\}$ et la matrice de transition se déduit aisément de la figure 3.1 :

$$P = \begin{pmatrix} 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 \\ \frac{1}{3} & 0 & \frac{1}{3} & 0 & \frac{1}{3} & 0 & 0 \\ 0 & \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} & 0 \\ \frac{1}{3} & 0 & 0 & 0 & \frac{1}{3} & 0 & \frac{1}{3} \\ 0 & \frac{1}{3} & 0 & \frac{1}{3} & 0 & \frac{1}{3} & 0 \\ 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{pmatrix}$$

EXEMPLE 3.2.– Partant d'un score vierge, on lance deux dés non pipés. Si la somme est différente de 7, on ajoute cette somme au score courant et on rejoue. Sinon le score s'annule et la partie s'arrête. On appelle X_n le score après le n^e lancer. Il faut distinguer ici deux états 0 si l'on veut que X soit une chaîne homogène. En effet, on peut quitter le 0 du départ mais on ne peut pas quitter le 0 consécutif à un 7. On prend donc espace d'états, $E = \mathbf{N} \cup \{\delta\}$, δ est ce qu'il est convenu d'appeler un point cimetièrè. Les transitions sont alors données pour tout $i \neq \delta$ par :

$$\begin{aligned} \mathbf{P}(X_{n+1} = i + 2 \mid X_n = i) &= \mathbf{P}(X_{n+1} = i + 12 \mid X_n = i) &= 1/36 \\ \mathbf{P}(X_{n+1} = i + 3 \mid X_n = i) &= \mathbf{P}(X_{n+1} = i + 11 \mid X_n = i) &= 2/36 \\ \mathbf{P}(X_{n+1} = i + 4 \mid X_n = i) &= \mathbf{P}(X_{n+1} = i + 10 \mid X_n = i) &= 3/36 \\ \mathbf{P}(X_{n+1} = i + 5 \mid X_n = i) &= \mathbf{P}(X_{n+1} = i + 9 \mid X_n = i) &= 4/36 \\ \mathbf{P}(X_{n+1} = i + 6 \mid X_n = i) &= \mathbf{P}(X_{n+1} = i + 8 \mid X_n = i) &= 5/36 \\ \mathbf{P}(X_{n+1} = \delta \mid X_n = i) &= 1/6 \\ \mathbf{P}(X_{n+1} = \delta \mid X_n = \delta) &= 1. \end{aligned}$$

La définition même d'une chaîne de Markov implique que toute son évolution est déterminée par la loi de la position initiale que l'on notera ν et l'opérateur de transition P .

THÉORÈME 3.1.– Pour tout n , la loi jointe de (X_0, \dots, X_n) est déterminée par la loi de X_0 et P à partir de la formule suivante :

$$\mathbf{P}(X_0 = x_0, \dots, X_n = x_n) = \nu(\{x_0\}) \prod_{l=0}^{n-1} p(x_l, x_{l+1}),$$

pour tout n et tout x_0, \dots, x_n dans E .

REMARQUE.– Par la suite, nous noterons \mathbf{P}_ν la loi d'une chaîne de Markov avec loi initiale ν . Par abus de notation, \mathbf{P}_x représentera la loi de la chaîne si le point de départ est fixe, égal à $x \in E$. Comme E est au plus dénombrable, on peut numéroter les états : utiliser une injection qui existe entre E et \mathbf{N} . On peut alors supposer que $E \subset \mathbf{N}$. On retrouve alors le formalisme des vecteurs et matrices même si l'on est amené à manipuler de tels objets avec un nombre infini de composantes... On considère souvent le « vecteur » π_n défini par $\pi_n(x) = \mathbf{P}(X_n = x)$ pour $x \in E \subset \mathbf{N}$. Il est usuel de le considérer comme un vecteur ligne. Pour tout n , la relation :

$$\begin{aligned} \mathbf{P}(X_{n+1} = j) &= \sum_{i \in E} \mathbf{P}(X_{n+1} = j | X_n = i) \mathbf{P}(X_n = i) \\ &= \sum_{i \in E} \mathbf{P}(X_n = i) p(i, j), \quad j \in E, \end{aligned}$$

se lit en notation matricielle :

$$\pi_{n+1} = \pi_n \cdot P \text{ soit } \pi_n = \pi_0 \cdot P^n \quad (3.3)$$

où P^n est la n^{e} puissance de P . En particulier, si π_0 n'est composé que de 0 sauf un 1 en i^{e} position (ce qui revient à travailler sous \mathbf{P}_i) alors pour tout $j \in E$, on a :

$$\mathbf{P}_i(X_n = j) = p^{(n)}(i, j)$$

où $p^{(n)}(i, j)$ est le terme en i^{e} ligne et j^{e} colonne de P^n .

Comme $P^{n+m} = P^n P^m$, on déduit de (3.3) l'équation de Chapman-Kolmogorov :

$$p^{(n+m)}(x, y) = \sum_{z \in E} p^{(n)}(x, z) p^{(m)}(z, y), \quad (3.4)$$

valable pour tout n, m , toute condition initiale et tout état final. On l'a écrite ici sous forme « intrinsèque », c'est-à-dire sans tenir compte de l'injection mentionnée plus haut.

3.1.1. Simulation

Rappelons d'abord comment simuler une loi ν sur un ensemble au plus dénombrable E . Les états sont numérotés grâce à une bijection ϕ entre E et un sous-ensemble de \mathbb{N} . On pose ensuite :

$$r_0 = \nu(\{\phi^{-1}(0)\}) \text{ et } r_n = \sum_{j=0}^n \nu(\{\phi^{-1}(j)\}) = \nu(\phi^{-1}(\{0, \dots, n\})).$$

Algorithme 3.1 . Réalisation d'une variable aléatoire de loi ν

Données : r_0, r_1, \dots
Résultat : un élément de E choisi selon la loi ν
 $x \leftarrow$ réalisation d'une loi uniforme sur $[0, 1]$;
 $n \leftarrow 0$;
tant que $x > r_n$ **faire**
 | $n \leftarrow n + 1$
fin
retourner $\phi^{-1}(n)$

Dans une chaîne de Markov, lorsque l'on est à l'état x , on passe à l'état y avec probabilité $p(x, y)$. Pour passer d'une étape à l'autre, il suffit d'appliquer l'algorithme précédent à la loi $\mu = (p(x, y), y \in E)$.

Algorithme 3.2 . Simulation d'une trajectoire d'une chaîne de Markov (ν, P)

Données : ν, P, N
Résultat : une trajectoire à N pas de la chaîne de Markov (ν, P)
Choisir x_0 condition initiale selon ν ;
pour compteur $\leftarrow 1$ **à** N **faire**
 | Choisir x_{compteur} selon la loi $(p(x_{\text{compteur}-1}, y), y \in E)$;
fin
retourner x_0, x_1, \dots, x_N

3.2. Propriété de Markov forte

Pour T temps d'arrêt, sur $(T < \infty)$, on définit θ^T par :

$$\theta^T(\omega) = (\omega_{T(\omega)}, \omega_{T(\omega)+1}, \dots).$$

Pour $x \in E$, on définit :

$$\tau_x^1 = \begin{cases} \infty & \text{si } X_n \neq x \text{ pour tout } n > 0; \\ \inf\{n > 0, X_n = x\} & \text{sinon.} \end{cases}$$

$$\tau_x^k = \begin{cases} \infty & \text{si } \tau_x^{k-1} = \infty \text{ ou } X_n \neq x \text{ pour tout } i > \tau_x^{k-1} \\ \inf\{n > \tau_x^{k-1}, X_n = x\} & \text{sinon.} \end{cases}$$

Pour tout k , τ_x^k est l'instant du k^{e} passage de la chaîne X à l'état x .

LEMME 3.2.— Pour x fixé dans E , sur l'événement $\{\tau_x^1 < \infty\}$, on a :

$$\tau_x^k = \tau_x^{k-1} + \tau_x^1 \circ \theta^{\tau_x^{k-1}}. \quad (3.5)$$

Démonstration. Si $\tau_x^{k-1} = \infty$ alors on a ∞ des deux côtés de l'égalité. Sinon, le résultat est immédiat à partir du moment où l'on se persuade que $\theta^{\tau_x^{k-1}}(\omega)$ représente la partie de la trajectoire postérieure à la $(k-1)^{\text{e}}$ visite à l'état x . Par conséquent, la première visite après la $(k-1)^{\text{e}}$ (si elle existe) est la k^{e} visite depuis le début. \square

THÉORÈME 3.3.— Soit T un temps d'arrêt presque sûrement fini et $F : \Omega \rightarrow \mathbf{R}^+$ une variable aléatoire intégrable. On a l'identité suivante :

$$\mathbf{E}[F \circ \theta^T | \mathcal{F}_T] = \mathbf{E}[F | X_0 = X_T]. \quad (3.6)$$

Pour calculer le terme de droite, on calcule $\mathbf{E}[F | X_0 = x] = \phi(x)$ et l'on a :

$$\mathbf{E}[F | X_0 = X_T] = \phi(X_T).$$

Démonstration. Pour $A \in \mathcal{F}_T$, puisque $A \cap \{T = n\} \in \mathcal{F}_n$ et en utilisant (3.2) et les propriétés de l'espérance conditionnelle, il vient :

$$\begin{aligned} \mathbf{E}[F \circ \theta^T \cdot \mathbf{1}_A] &= \sum_{n=0}^{\infty} \mathbf{E}[F \circ \theta^n \cdot \mathbf{1}_{A \cap \{T=n\}}] \\ &= \sum_{n=0}^{\infty} \mathbf{E}[\mathbf{E}[F \circ \theta^n | \mathcal{F}_n] \mathbf{1}_{A \cap \{T=n\}}] \\ &= \sum_{n=0}^{\infty} \mathbf{E}[\mathbf{E}[F | X_0 = X_n] \mathbf{1}_{A \cap \{T=n\}}] \\ &= \mathbf{E}[\mathbf{E}[F | X_0 = X_T] \mathbf{1}_A] \end{aligned}$$

L'égalité est vraie par linéarité pour toutes les fonctions étagées et donc pour toutes les fonctions positives. \square

REMARQUE.— Soit la chaîne de Markov à deux états 0 et 1, et de matrice de transition $p_{0,0} = 0,9, p_{1,1} = 1$. Soit $T = \sup\{n \geq 1, X_n = 0\}$. Sous \mathbf{P}_0 , $T = Y - 1$ où Y suit une loi géométrique de paramètre 0,1. Il s'ensuit T est presque sûrement fini. Or, $\mathbf{P}_0(X_{T+1} = 1 | X_T = 0) = 1$, qui est différent de $\mathbf{P}_0(X_{n+1} = 1 | X_n = 0) = 0,1$.

Cet exemple illustre que l'on ne peut pas supprimer l'hypothèse « T temps d'arrêt » dans la propriété de Markov forte. Il est clair qu'ici T n'est pas un temps d'arrêt car savoir si T est inférieur à n nécessite de connaître la trajectoire après l'instant n pour être sûr que l'on ne repassera pas par 0 après n .

EXEMPLE (Suite de 3.1).— Supposons qu'il y ait un bout de fromage en case 3 et une batterie en case 7. On veut calculer la probabilité que le rat puisse manger avant d'être électrocuté. Introduisons :

$$\tau_3 = \inf\{n \geq 0, X_n = 3\} \text{ et } \tau_7 = \inf\{n \geq 0, X_n = 7\}.$$

On pose $u_i = \mathbf{P}_i(\tau_3 < \tau_7)$. Il est clair que $u_3 = 1$ et que $u_7 = 0$. Pour $i \notin \{3; 7\}$,

$$u_i = \sum_{j=1}^7 \mathbf{P}_i(\tau_3 < \tau_7 | X_1 = j) \mathbf{P}_i(X_1 = j).$$

Puisque i est différent de 3 et 7, l'événement $(\tau_3 < \tau_7)$ est \mathbf{P}_i presque sûrement égal à A_1 où :

$$\begin{aligned} A_1 &= \{\omega, \exists i \geq l \text{ tel que } \omega_i = 3 \text{ et } \omega_j \in \{1, 2, 4, 5, 6\} \text{ pour tout } l \leq j < i\} \\ &= \{\text{postérieurement à l'instant } l, \text{ on atteint 3 avant 7}\}. \end{aligned}$$

Comme $\mathbf{1}_{A_1} = \mathbf{1}_{A_0} \circ \theta$, on a :

$$\mathbf{P}_i(\tau_3 < \tau_7 | X_1 = j) = \mathbf{P}_j(\tau_3 < \tau_7).$$

Compte tenu du fait que $\mathbf{P}_i(X_1 = j) = p(i, j)$ on voit que les u_i sont solutions du système linéaire :

$$u_3 = 1, u_7 = 0, u_i = \sum_{j=1}^6 p(i, j) u_j \text{ pour } i \notin \{3; 7\}.$$

La résolution de ce système donne $u_1 = 7/12, u_2 = 3/4, u_4 = 5/12, u_5 = 2/3, u_6 = 5/6$.

Sans fromage et batterie, calculons maintenant le temps moyen d'atteinte de la case 3. Posons $v_i = \mathbf{E}_i[\tau_3]$. Il est clair $v_3 = 0$. Par ailleurs, pour $i \neq 3$, on a :

$$\mathbf{E}_i[\tau_3] = \sum_{j=1}^7 \mathbf{E}_i[\tau_3 | X_1 = j] p(i, j).$$

Si l'on a comme trajectoire $\omega = (1, 2, 5, 2, 5, 6, 3, \dots)$ alors $\tau_3(\omega) = 6$ mais $\tau_3(\theta\omega) = 5$. Plus généralement, on a, conditionnellement à $X_0 \neq 3$, $\tau_3 = \tau_3 \circ \theta + 1$. Par conséquent, on a :

$$v_i = \sum_{j=1}^7 (\mathbf{E}_i [\tau_3 \circ \theta | X_1 = j] + 1)p(i, j) = \sum_{j=1}^7 p(i, j)v_j + 1,$$

d'après la relation (3.2). Les v_i sont donc les solutions d'un système linéaire à six équations et six inconnues qu'il ne vous reste plus, cher lecteur, qu'à résoudre.

3.3. Classification des états

On note N_x le nombre de visites à l'état x après le départ :

$$N_x = \sum_{n=1}^{\infty} \mathbf{1}_{\{X_n=x\}}.$$

LEMME 3.4.– Pour tout k , les deux événements $\{N_x \geq k\}$ et $\{\tau_x^k < \infty\}$ coïncident.

Démonstration. $N_x \geq k$ signifie qu'il y a eu plus de k visites à l'état x , ce qui est très exactement équivalent à dire que $\tau_x^k < \infty$. \square

DÉFINITION 3.3.– Un état x est dit récurrent lorsque $\mathbf{P}_x(\tau_x^1 < \infty) = 1$. Sinon, x est dit transient. La chaîne X est dite récurrente (respectivement transiente) si tous ses états sont récurrents (respectivement transients).

LEMME 3.5.– Pour tout couple (x, y) de E , on a :

$$\mathbf{P}_y(\tau_x^k < \infty) = \mathbf{P}_x(\tau_x^1 < \infty)^{k-1} \mathbf{P}_y(\tau_x^1 < \infty). \quad (3.7)$$

En particulier, si $x = y$, $\mathbf{P}_x(\tau_x^k < \infty) = \mathbf{P}_x(\tau_x^1 < \infty)^k$. Par ailleurs, on a :

$$\mathbf{E}_y[N_x] = \frac{\mathbf{P}_y(\tau_x^1 < \infty)}{1 - \mathbf{P}_x(\tau_x^1 < \infty)} = \sum_{n \geq 1} p^{(n)}(y, x). \quad (3.8)$$

▣ Prenant comme instant présent celui de la k^{e} visite à l'état x , d'après la propriété de Markov forte, le passé et le futur conditionnellement à cette visite sont indépendants. Par conséquent, sachant que l'on a déjà visité k fois l'état x , la probabilité que l'on revienne en x une $(k+1)^{\text{e}}$ fois est la même que lors de la première visite en x on revienne au moins une fois. De plus, ces deux événements sont indépendants.

Démonstration. Pour $k > 2$, d'après (3.5) et (3.6), on a :

$$\begin{aligned}
 \mathbf{P}_y(\tau_x^k < \infty) &= \mathbf{P}_y(\tau_x^{k-1} < \infty, \tau_x^1 \circ \theta^{\tau_x^{k-1}} < \infty) \\
 &= \mathbf{E}_y \left[\mathbf{1}_{\{\tau_x^{k-1} < \infty\}} \mathbf{P}_y(\tau_x^1 \circ \theta^{\tau_x^{k-1}} < \infty \mid \mathcal{F}_{\tau_x^{k-1}}) \right] \\
 &= \mathbf{E}_y \left[\mathbf{1}_{\{\tau_x^{k-1} < \infty\}} \mathbf{P}_y(\tau_x^1 < \infty \mid X_0 = X_{\tau_x^{k-1}}) \right] \\
 &= \mathbf{E}_y \left[\mathbf{1}_{\{\tau_x^{k-1} < \infty\}} \mathbf{P}_y(\tau_x^1 < \infty \mid X_0 = x) \right] \\
 &= \mathbf{E}_y \left[\mathbf{1}_{\{\tau_x^{k-1} < \infty\}} \mathbf{P}_y(\tau_x^1 < \infty) \right] \\
 &= \mathbf{P}_y(\tau_x^{k-1} < \infty) \mathbf{P}_y(\tau_x^1 < \infty),
 \end{aligned}$$

et on retrouve (3.7) par récurrence.

Maintenant, d'après le théorème de Fubini, on peut écrire :

$$\mathbf{E}_y [N_x] = \sum_{k \geq 1} \mathbf{P}_y(N_x \geq k) = \sum_{k \geq 1} \mathbf{P}_y(\tau_x^k < \infty).$$

La première égalité de (3.8) en découle. Toujours d'après le théorème de Fubini et (3.4), on a :

$$\mathbf{E}_y [N_x] = \mathbf{E}_y \left[\sum_{n=1}^{\infty} \mathbf{1}_{\{X_n=x\}} \right] = \sum_{n=1}^{\infty} \mathbf{E}_y [\mathbf{1}_{\{X_n=x\}}] = \sum_{n=1}^{\infty} p^{(n)}(y, x),$$

d'où le résultat. \square

Le théorème suivant permet de donner les différentes caractérisations de la récurrence et de la transience.

THÉORÈME 3.6.— Soit x un état fixé.

- 1) Les assertions suivantes sont équivalentes :
 - a) x est récurrent,
 - b) $\mathbf{P}_x(N_x = \infty) = 1$,
 - c) $\mathbf{E}_x [N_x] = \infty$.
- 2) Les assertions suivantes sont équivalentes :
 - a) x est transient,
 - b) $\mathbf{P}_x(N_x < \infty) = 1$,
 - c) $\mathbf{E}_x [N_x] < \infty$.

Démonstration. Montrons d'abord que $a \Rightarrow b$. D'après (3.4) et le lemme 3.5 :

$$\mathbf{P}_x(N_x > k) = \mathbf{P}_x(\tau_x^k < \infty) = \mathbf{P}_x(\tau_x < \infty)^k, \quad (3.9)$$

et d'après le théorème de convergence monotone :

$$\mathbf{P}_x(N_x = \infty) = \lim_{k \rightarrow \infty} \mathbf{P}_x(N_x > k). \quad (3.10)$$

La récurrence de x signifie $\mathbf{P}_x(\tau_x < \infty) = 1$ et implique donc $N_x = \infty$, \mathbf{P}_x presque sûrement. Par conséquent, x récurrent implique que $\mathbf{P}_x(\tau_x^1 < \infty) = 1$. Par le même raisonnement, x transient implique $\mathbf{P}_x(\tau_x^1 < \infty) < 1$.

$b \Rightarrow c$. Immédiat dans le cas de x récurrent. Pour l'autre cas, utilisons la relation :

$$\mathbf{E}_x[N_x] = \sum_{k \geq 0} \mathbf{P}_x(\tau_x < \infty)^k. \quad (3.11)$$

Comme N_x est fini presque sûrement, d'après (3.10), $\mathbf{P}_x(N_x > k)$ tend vers 0 quand k tend vers l'infini. D'après (3.9) ceci implique que $\mathbf{P}_x(\tau_x < \infty) < 1$ donc que la série converge.

$c \Rightarrow a$. Dans les deux cas, la relation (3.11) permet de conclure. \square

DÉFINITION 3.4.— On dit qu'un état x conduit à un état y et on le note $x \longrightarrow y$, s'il existe un entier strictement positif m tel que $p^{(m)}(x, y) > 0$. Ce qui revient à dire que $\mathbf{P}_x(\tau_y^1 < \infty) = 1$.

THÉORÈME 3.7.— Si x est un état récurrent et $x \longrightarrow y$, alors $y \longrightarrow x$ et y est récurrent.

☞ Partant de x on sait que l'on va presque sûrement en y , si de y il y a un risque de ne pas revenir en x on va finir par effectivement ne pas y revenir ; on ne fera donc qu'un nombre fini de visites à x ce qui est incompatible avec l'hypothèse de récurrence. De plus, si de y on est presque sûr de revenir en x et que l'on passe une infinité de fois par x on passera vraisemblablement une infinité de fois en y aussi.

Démonstration. Montrons par l'absurde que y conduit à x en écrivant que la probabilité de ne jamais revenir en x en étant parti de x est supérieure à la probabilité de la même chose mais en passant une fois par y :

$$\begin{aligned} \mathbf{P}_x(\tau_x = \infty) &\geq \mathbf{P}_x(\tau_x \circ \theta^{\tau_y} = \infty, \tau_y < \infty) \\ &= \mathbf{P}_x(\tau_y < \infty) \mathbf{P}_y(\tau_x = \infty), \end{aligned}$$

d'après la propriété de Markov forte. Si y ne conduit pas à x , cette quantité est strictement positive ce qui est en contradiction avec la récurrence de x . De même, on a :

$$\begin{aligned} \mathbf{P}_y(\tau_y < \infty) &\geq \mathbf{P}_y(\tau_y \circ \theta^{\tau_x} < \infty, \tau_x < \infty) \\ &= \mathbf{P}_y(\tau_x < \infty)\mathbf{P}_x(\tau_y < \infty) = 1, \end{aligned}$$

donc y est récurrent. \square

THÉORÈME 3.8.– La relation \longrightarrow restreinte aux états récurrents est une relation d'équivalence.

Démonstration. La réflexivité, c'est-à-dire $x \longrightarrow x$, est induite par la définition même d'un état récurrent. La symétrie, c'est-à-dire $x \longrightarrow y \implies y \longrightarrow x$, découle du théorème 3.7. Soit x, y et z trois états de E tels que $x \longrightarrow y$ et $y \longrightarrow z$. Par définition, il existe deux entiers positifs que nous appelons r et s tels que $p^{(r)}(x, y) > 0$ et $p^{(s)}(y, z) > 0$. L'équation de Chapman-Kolmogorov implique que :

$$p^{(r+s)}(x, z) = \sum_{\ell \in E} p^{(r)}(x, \ell)p^{(s)}(\ell, z).$$

Tous les termes de cette somme sont positifs et il existe au moins un terme strictement positif : $p^{(r)}(x, y)p^{(s)}(y, z)$. Nous avons donc trouvé un entier positif, $r + s$, tel que $p^{(r+s)}(x, z) > 0$, d'où le résultat. \square

L'ensemble des points récurrents peut donc être partitionné en classes d'équivalence. Par définition, un état appartenant à une classe conduit à tous les autres états de cette classe et ne conduit à aucun état récurrent appartenant à une autre classe ni à un état transient. En revanche, un état transient peut conduire aussi bien à un état transient qu'à un état récurrent.

DÉFINITION 3.5.– Un sous-ensemble F de E est dit fermé, si pour tout x et y :

$$(x \in F \text{ et } x \longrightarrow y) \implies y \in F.$$

Autrement dit, $\sum_{y \in F} p(x, y) = 1$ pour tout $x \in F$.

THÉORÈME 3.9.– Tout ensemble fermé de cardinal fini contient au moins un point récurrent.

Démonstration. Soit F un ensemble fermé, si tous ses états sont transients, on a :

$$\mathbf{E}_y [N_x] = \mathbf{P}_y(\tau_x^1 < \infty)\mathbf{E}_x [N_x] < \infty$$

pour tout couple (x, y) de F . Comme F est de cardinal fini, $\sum_{x \in F} \mathbf{E}_y [N_x] < \infty$. Or, on a la suite d'identités :

$$\begin{aligned} \sum_{x \in F} \mathbf{E}_y [N_x] &= \sum_{x \in F} \mathbf{E}_y \left[\sum_{n \geq 0} \mathbf{1}_{\{X_n = x\}} \right] = \sum_{n \geq 0} \mathbf{E}_y \left[\sum_{x \in F} \mathbf{1}_{\{X_n = x\}} \right] \\ &= \sum_{n \geq 0} 1 = \infty, \end{aligned}$$

puisque F est fermé. On a abouti à une absurdité, il existe donc au moins un point récurrent. \square

EXEMPLE 3.3.– Il est souvent simple d'avoir une représentation graphique de la matrice de transition d'une chaîne de Markov. Pour ce faire, on construit un graphe orienté dont les sommets correspondent aux états. L'arête (orientée) x, y a comme poids la probabilité de transition de x vers y . Si cette probabilité est nulle, l'arête n'existe pas. Considérons la chaîne de Markov de matrice de transition :

$$P = \begin{pmatrix} 0.8 & 0.1 & 0 & 0.1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & \frac{1}{1000} & \frac{999}{1000} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1/3 & 1/3 & 1/3 \end{pmatrix}.$$

La représentation graphique est alors celle de la figure 3.2.

Il est clair que les ensembles $\{2, 3\}$ et $\{4, 5, 6\}$ sont des ensembles fermés. Comme ils sont tous deux de cardinal fini, ils contiennent tous deux au moins un point récurrent. A l'intérieur de chacun d'entre eux, les points conduisent tous les uns aux autres, donc ils sont tous récurrents. Si l'on quitte le point 1 pour le point 2 ou 4, on n'est sûr de ne pas y revenir donc la probabilité de ne pas revenir en 1 est $2 * 1/10 > 0$ donc 1 est transient.

Remarquons que même si l'ensemble E est fermé, le théorème 3.9 n'induit pas de contradiction quant au statut de l'état 1. En effet, on sait qu'il existe au moins un point récurrent dans E mais on ne sait pas lequel et on ne peut, *a priori*, rien dire de plus puisque tous les états ne communiquent pas entre eux.

EXEMPLE (Suite de l'exemple 3.1).– Tous les états conduisent les uns aux autres donc le seul sous-ensemble fermé est E lui-même. Comme il est de cardinal fini, il existe au moins un état récurrent donc ils sont tous récurrents.

EXEMPLE (Suite de l'exemple 3.2).– Tant qu'il n'atteint pas δ , le score ne peut que croître donc tous les états de \mathbf{N} sont transients. δ est lui récurrent.

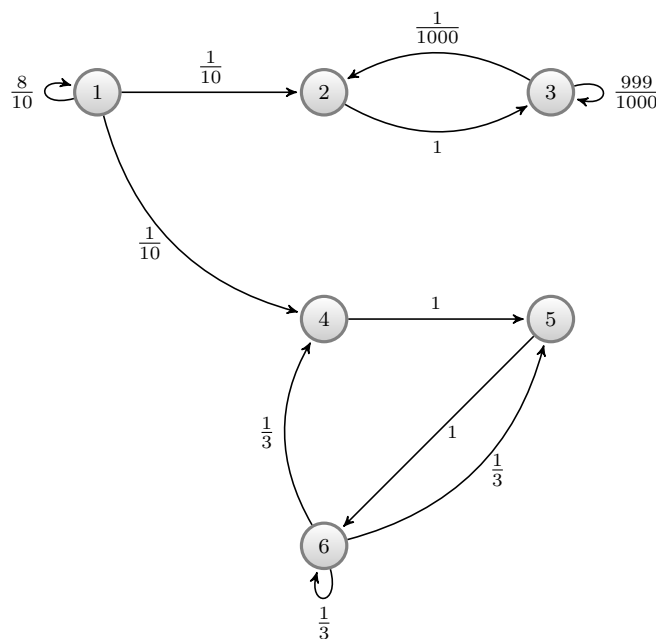


Figure 3.2 – Représentation graphique d’une chaîne de Markov

EXEMPLE.– Un jeu de N cartes est mélangé en le coupant en deux parts qui sont ensuite interverties. Chaque mélange du jeu est représenté par une permutation de $\{1, \dots, N\}$. Si $N = 3$ et que le mélange est représenté par $(3, 2, 1)$, cela signifie que la carte 3 est en position 1, la carte 2 en position 2 et la carte 1 en position 3. On note X_n l’état du paquet de cartes après la n^{e} opération de mélange. L’espace d’états est donc le groupe des permutations de $\{1, \dots, N\}$ dans lui-même, noté \mathfrak{S}_N . Si $X_0 = (3, 2, 1)$ et que la coupe se fait entre la première et la deuxième carte, on a $X_1 = (2, 1, 3)$. En d’autres termes, on a juste fait une permutation circulaire sur les cartes mais on n’a pas changé leur ordre relatif. Pour définir les probabilités de transition, considérons l’ensemble à N éléments :

$$E_1 = \left\{ \sigma \in \mathfrak{S}_N, \exists k \in \{1, \dots, N\}, \sigma = (k+1, k+2, \dots, N, 1, \dots, k) \right\}.$$

Lorsque l’on coupe le paquet au niveau de la k^{e} carte, on applique le cycle $(k+1, k+2, \dots, N, 1, \dots, k)$ à la situation courante. Comme le choix de l’endroit de la coupe est supposé être uniforme sur $\{1, \dots, N\}$, on a :

$$\mathbf{P}(X_1 = \tau \mid X_0 = \sigma) = \frac{1}{N} \text{ si } \tau\sigma^{-1} \in E_1.$$

Les classes d'équivalence de la relation \longrightarrow sont celles de la relation $\sigma \mathfrak{R} \tau \equiv \tau \sigma^{-1} \in E_1$. En d'autres termes, σ conduit à τ si et seulement s'il existe $\rho \in E_1$ tel que $\tau = \rho \sigma$. Il y a donc $(n-1)!$ classes d'équivalence de cardinal n chacune. Toutes ces classes forment des ensembles fermés de cardinal fini qui contiennent donc toutes au moins un point récurrent. Comme à l'intérieur de ces classes les états communiquent tous entre eux, ils sont tous récurrents. La chaîne est donc récurrente.

Lorsque l'espace d'état est infini, on ne peut pas appliquer le théorème 3.9. On introduit alors la notion suivante.

DÉFINITION 3.6.— Une chaîne de Markov est dite irréductible lorsque tous les états conduisent les uns aux autres. En particulier, le plus petit sous-espace fermé est E lui-même et tous les états ont même nature.

REMARQUE.— Si le nombre d'états transients est fini, comme l'on ne passe qu'un nombre fini de fois en chacun d'eux, la chaîne de Markov sera inexorablement contrainte à aller dans une classe de récurrence et à y rester. Remarquons qu'en vertu du lemme 3.7, une classe de récurrence est forcément un sous-ensemble irréductible. Si le nombre d'états transients est infini, le raisonnement précédent ne s'applique plus automatiquement mais les cas dans lesquels on n'atterrit pas automatiquement dans un sous-ensemble fermé irréductible sont hors de notre propos. Pour ce qui nous intéresse (le comportement asymptotique des chaînes de Markov), il n'y a donc pas de perte de généralité à supposer que les chaînes de Markov étudiées sont irréductibles.

Lorsque x est récurrent, on sait que partant de x on reviendra nécessairement en x en un temps fini mais *quid* du temps moyen de retour en x ?

DÉFINITION 3.7.— Un état récurrent x est dit récurrent positif, si $\mathbf{E}_x [\tau_x^1] < \infty$; récurrent nul, si $\mathbf{E}_x [\tau_x^1] = \infty$.

La chaîne X est alors dite récurrente positive (respectivement récurrente nulle) si tous ses états sont récurrents positifs (respectivement récurrents nuls).

La construction suivante est utilisée plusieurs fois par la suite.

DÉFINITION 3.8.— Soit X , une chaîne de Markov irréductible et récurrente sur E et F , un sous-ensemble de E . On note :

$$\tau_F^1 = \inf \{ n \geq 1, X_n \in F \} \text{ et } \tau_F^{k+1} = \tau_F^k + \tau_F^1 \circ \theta^{\tau_F^k},$$

les instants de visite successifs de la chaîne à l'ensemble F . On considère la suite aléatoire X^F , définie par $X_n^F = X_{\tau_F^n}$, $n \in \mathbf{N}$. On vérifie facilement que X^F est une chaîne de Markov sur F , appelée chaîne de Markov restreinte à F .

THÉORÈME 3.10.— Soit X une chaîne de Markov irréductible et F un sous-ensemble fini de E . Si pour tout $x \in F$, $\mathbf{E}_x [\tau_F^1] < \infty$ alors X est récurrente positive.

Démonstration. Soit pour tout $x \in F$, $\sigma_x = \inf\{n > 0, X_n^F = x\}$ et pour tout $k \in \mathbf{N}^*$, $Y_k = \tau_F^k - \tau_F^{k-1}$. Comme F est fini, X^F est récurrente positive donc $\mathbf{E}_x [\sigma_x] < \infty$ pour tout $x \in F$. Il nous faut prouver que $\mathbf{E}_x [\tau_x] < \infty$. Par construction des variables Y_k , on a :

$$\begin{aligned} \mathbf{E}_x [\tau_x] &= \mathbf{E}_x \left[\sum_{k=1}^{\sigma_x} Y_k \right] \\ &= \sum_{n \geq 1} \mathbf{E}_x \left[\sum_{k=1}^{\sigma_x} Y_k \mathbf{1}_{\{\sigma_x = n\}} \right] \\ &= \sum_{k=1}^{\infty} \mathbf{E}_x \left[Y_k \sum_{n \geq k} \mathbf{1}_{\{\sigma_x = n\}} \right] \\ &= \sum_{k=1}^{\infty} \mathbf{E}_x [Y_k \mathbf{1}_{\{\sigma_x \geq k\}}]. \end{aligned}$$

En utilisant la propriété de Markov forte, on obtient :

$$\begin{aligned} \mathbf{E}_x [Y_k \mathbf{1}_{\{\sigma_x \geq k\}}] &= \mathbf{E}_x \left[\mathbf{E}_x [Y_k | \mathcal{F}_{\tau_{F}^{k-1}}] \mathbf{1}_{\{\sigma_x \geq k\}} \right] \\ &= \mathbf{E}_x \left[\mathbf{E}_{X_{\tau_{F}^{k-1}}} [Y_1] \mathbf{1}_{\{\sigma_x \geq k\}} \right] \\ &\leq \sup_{y \in F} \mathbf{E}_y [Y_1] \mathbf{P}_x(\sigma_x \geq k). \end{aligned}$$

Le *supremum* est fini par hypothèse puisque F est fini. On obtient donc :

$$\mathbf{E}_x [\tau_x] \leq c \sum_{k=1}^{\infty} \mathbf{P}_x(\sigma_x \geq k) = c \mathbf{E}_x [\sigma_x].$$

En vertu de la remarque initiale, cela prouve la positive récurrence de X . □

LEMME 3.11.— Soit X une chaîne de Markov et $h : E \times E \rightarrow \mathbf{R}$, bornée. Pour tout entier n , on a :

$$\mathbf{E} [h(X_n, X_{n+1}) | \mathcal{F}_n] = P(h(X_n, \cdot))(X_n) = \sum_{y \in E} p(X_n, y) h(X_n, y). \quad (3.12)$$

Démonstration. Comme h est bornée, seul reste le calcul de l'espérance conditionnelle. D'après la propriété de Markov :

$$\mathbf{E}[h(X_n, X_{n+1}) | \mathcal{F}_n] = \mathbf{E}[h(X_n, X_{n+1}) | X_n].$$

Soit maintenant $\phi : E \rightarrow \mathbf{R}$ bornée. On a :

$$\begin{aligned} \mathbf{E}[h(X_n, X_{n+1})\phi(X_n)] &= \int \phi(x) \int h(x, y) d\mathbf{P}_{X_{n+1}|X_n=x}(y) d\mathbf{P}_{X_n}(x) \\ &= \int \phi(x) \sum_{y \in E} h(x, y)p(x, y) d\mathbf{P}_{X_n}(x) \\ &= \mathbf{E}[\phi(X_n)P(h(X_n, \cdot))(X_n)]. \end{aligned}$$

Comme la précédente équation est vraie pour toute fonction ϕ , on en déduit (3.12). \square

THÉORÈME 3.12 (Critère de Foster).— Soit E_0 une partie finie de E . S'il existe une fonction $h : E \rightarrow \mathbf{R}$ telle que l'ensemble $\{x \in E, h(x) < K\}$ soit fini pour tout entier K et que :

$$h(y) \geq \mathbf{E}_y[h(X_1)] \text{ pour tout } y \in E_0^c,$$

alors X est récurrente.

Démonstration. En particulier, h est minorée donc quitte à rajouter une constante, on peut supposer $h \geq 0$. Soit le temps d'arrêt $\tau = \inf\{n, X_n \in E_0\}$ et Y définie par $Y_n = h(X_n) \mathbf{1}_{\{\tau > n\}}$. Montrons que Y est une sur-martingale positive dès que $X_0 \in E_0^c$. Soit $x \in E_0^c$, on a :

$$\begin{aligned} \mathbf{E}_x[h(X_{n+1}) \mathbf{1}_{\{\tau > n+1\}} | \mathcal{F}_n] &= \mathbf{1}_{\{\tau > n+1\}} \mathbf{E}_{X_n}[h(X_{n+1})] \\ &\leq \mathbf{1}_{\{\tau > n\}} h(X_n) = Y_n, \end{aligned}$$

car sur $(\tau > n + 1)$, X_n n'appartient pas à E_0 . Y converge donc presque sûrement vers une variable aléatoire Y_∞ .

Supposons que X soit transiente. Soit $x \notin E_0$, pour tout entier K , l'ensemble $\{x, h(x) < K\}$ est fini donc n'est visité qu'un nombre fini de fois par X donc X n'est pas bornée. Comme Y_∞ est finie, nécessairement τ est fini presque sûrement. Ce qui revient à dire que pour $x \notin E_0$, $\mathbf{P}_x(\tau < \infty) = 1$. Partant de E_0^c , on arrive forcément dans E_0 . Soit on reste dans E_0 pour toujours et comme E_0 est fini, E_0 est récurrent et par irréductibilité la chaîne l'est. Soit la chaîne quitte E_0 mais en vertu de ce que l'on vient de démontrer, elle y reviendra. Le nombre de visites à E_0 est donc infini, ce qui implique encore une fois que E_0 est récurrent donc la chaîne l'est aussi. \square

3.4. Mesures et probabilité invariantes

DÉFINITION 3.9.— Soit E un ensemble dénombrable et P une opérateur de transition sur $E \times E$. Une mesure positive finie ν sur E est dite invariante par rapport à P , si et seulement si :

$$\nu = \nu P \text{ c'est-à-dire } \nu(y) = \sum_{x \in E} \nu(x)p(x, y) \text{ pour tout } y \in E. \quad (3.13)$$

Si de plus $\sum \nu(x) = 1$, ν est une probabilité invariante.

REMARQUE.— Si $\pi_0 = \nu$ alors $\pi_n = \pi_0 P^n = \pi_0$.

THÉORÈME 3.13.— Soit x un état récurrent, alors la mesure ν définie par :

$$\nu(y) = \mathbf{E}_x \left[\sum_{n=0}^{\tau_x^1 - 1} \mathbf{1}_{\{X_n = y\}} \right] = \sum_{n=0}^{\infty} \mathbf{P}_x(X_n = y, \tau_x^1 > n)$$

est une mesure invariante.

Démonstration. Montrons d'abord l'égalité des deux expressions de ν . Puisque x est récurrent, τ_x^1 est presque sûrement fini donc $\cup_{n \geq 1} \{\tau_x^1 = n\}$ est une partition de Ω . En déconditionnant sur toutes les valeurs possibles de τ_x^1 et en utilisant le théorème de Fubini, on obtient :

$$\begin{aligned} \mathbf{E}_x \left[\sum_{n=0}^{\tau_x^1 - 1} \mathbf{1}_{\{X_n = y\}} \right] &= \sum_{\ell=1}^{\infty} \mathbf{E}_x \left[\sum_{n=0}^{\ell-1} \mathbf{1}_{\{X_n = y\}} \mathbf{1}_{\{\tau_x^1 = \ell\}} \right] \\ &= \sum_{n=0}^{\infty} \mathbf{E}_x \left[\sum_{\ell > n} \mathbf{1}_{\{\tau_x^1 = \ell\}} \mathbf{1}_{\{X_n = y\}} \right] \\ &= \sum_{n=0}^{\infty} \mathbf{E}_x \left[\mathbf{1}_{\{\tau_x^1 > n\}} \mathbf{1}_{\{X_n = y\}} \right]. \end{aligned}$$

Comme sous \mathbf{P}_x , $X_0 = X_{\tau_x^1} = x$, on peut aussi écrire $\nu(y) = \mathbf{E}_x \left[\sum_{n=1}^{\tau_x^1} \mathbf{1}_{\{X_n = y\}} \right]$, ce qui donne par les mêmes calculs avec des bornes différentes :

$$\nu(y) = \sum_{n=1}^{\infty} \mathbf{E}_x \left[\mathbf{1}_{\{\tau_x^1 \geq n\}} \mathbf{1}_{\{X_n = y\}} \right]. \quad (3.14)$$

Remarquons d'ores et déjà que l'événement $\{\tau_x^1 \geq n\}$ appartient à \mathcal{F}_{n-1} puisque c'est le complémentaire de l'événement $\{\tau_x^1 \leq n-1\}$. Pour $y \neq x$, en appliquant les

propriétés de l'espérance conditionnelle et la propriété de Markov forte :

$$\begin{aligned}
& \sum_{n=0}^{\infty} \mathbf{E}_x [\mathbf{1}_{\{\tau_x^1 \geq n\}} \mathbf{1}_{\{X_n=y\}}] \\
&= \sum_{n=1}^{\infty} \mathbf{E}_x [\mathbf{1}_{\{\tau_x^1 \geq n\}} \mathbf{E}_x [\mathbf{1}_{\{X_n=y\}} \mid \mathcal{F}_{n-1}]] \\
&= \sum_{n=1}^{\infty} \mathbf{E}_x [\mathbf{1}_{\{\tau_x^1 \geq n\}} \mathbf{E}_x [\mathbf{1}_{\{X_n=y\}} \mid X_{n-1}]] \\
&= \sum_{n=1}^{\infty} \sum_{z \in E} \mathbf{E}_x [\mathbf{1}_{\{\tau_x^1 \geq n\}} \mathbf{1}_{\{X_{n-1}=z\}} \mathbf{E}_x [\mathbf{1}_{\{X_n=y\}} \mid X_{n-1} = z]] \\
&= \sum_{z \in E} p(z, y) \sum_{n=1}^{\infty} \mathbf{E}_x [\mathbf{1}_{\{\tau_x^1 \geq n\}} \mathbf{1}_{\{X_{n-1}=z\}}] \\
&= \sum_{z \in E} \nu(z) p(z, y).
\end{aligned}$$

Pour $y = x$, il est clair que $\nu(x) = 1$ et d'autre part :

$$\begin{aligned}
\sum_{z \in E} \nu(z) p(z, x) &= \sum_{n=0}^{\infty} \sum_{z \in E} p(z, x) \mathbf{P}_x(X_n = z, \tau_x^1 > n) \\
&= \sum_{n=0}^{\infty} \sum_{z \in E} \mathbf{P}_x(X_n = z, X_{n+1} = x, \tau_x^1 > n) \\
&= \sum_{n=0}^{\infty} \sum_{z \in E} \mathbf{P}_x(X_n = z, \tau_x^1 = n + 1) \\
&= \sum_{n=0}^{\infty} \mathbf{P}_x(\tau_x^1 = n + 1) \\
&= \mathbf{P}_x(\tau_x^1 < \infty) = 1.
\end{aligned}$$

On a donc bien $\nu = \nu P$, et il ne reste qu'à vérifier que $\nu(y) < \infty$ pour tout y . C'est vrai pour $x = y$. Pour $y \neq x$, de deux choses l'une : soit x ne conduit pas à y et alors $\nu(y) = 0$, soit x conduit à y et comme x est récurrent, d'après le théorème 3.7, y conduit à x , c'est-à-dire qu'il existe m tel que $p^{(m)}(y, x) > 0$. Comme ν est invariante, $\nu.P^m = \nu$, ce qui implique que :

$$1 = \nu(x) = \sum_{z \in E} \nu(z) p^{(m)}(z, x) \geq \nu(y) p^{(m)}(y, x),$$

et donc $\nu(y) < \infty$. \square

COROLLAIRE 3.14.— Soit X une chaîne de Markov irréductible et récurrente de mesure invariante ν . Soit F un sous-ensemble de E et X^F la chaîne restreinte à F . Alors, X^F est irréductible et récurrente et admet, comme X , pour mesure invariante celle donnée par le théorème 3.13.

Démonstration. Les deux premiers points sont évidents. Pour $y \in F$, le nombre de visites à y de X^F et de X sont les mêmes donc X et X^F la même mesure invariante donnée par le théorème 3.13. \square

THÉORÈME 3.15.— Si la chaîne de Markov X est irréductible et récurrente alors il existe une unique (à un coefficient près) mesure invariante ν telle que pour tout y , $0 < \nu(y) < \infty$. L'unicité à un coefficient multiplicatif près signifie que si ν et ν' sont deux telles mesures alors il existe $c > 0$ tel que $\nu(x) = c\nu'(x)$ pour tout $x \in E$.

Démonstration. Soit μ une mesure invariante et soit $a \in E$. Soit ν la mesure invariante construite dans le théorème 3.13 avec a comme point de départ. Par construction, $\nu(a) = 1$ donc pour toute mesure invariante μ , $\mu(a) = \nu(a)\mu(a)$. Par définition, pour $z \in E \setminus \{a\}$:

$$\mu(z) = \sum_{y \in E} \mu(y)p(y, z) = \mu(a)p(a, z) + \sum_{y \neq a} \mu(y)p(y, z).$$

En itérant cette relation, on obtient :

$$\mu(z) = \mu(a)p(a, z) + \mu(a) \sum_{y \neq a} p(a, y)p(y, z) + \sum_{i \neq a} \sum_{y \neq a} \mu(x)p(x, y)p(y, z),$$

ce qui peut se récrire de la façon suivante :

$$\begin{aligned} \mu(z) &= \mu(a)\mathbf{P}_a(X_1 = z) \\ &\quad + \mu(a)\mathbf{P}_a(X_1 \neq a, X_2 = z) + \mathbf{P}_\mu(X_0 \neq a, X_1 \neq a, X_2 = z). \end{aligned}$$

Par récurrence sur n , on montre alors que pour tout n :

$$\mu(z) = \mu(a) \sum_{m=1}^n \mathbf{P}_a(\tau_a^1 > m, X_m = z) + \mathbf{P}_\mu \left(\bigcap_{y=0}^n (X_y \neq a) \cap X_n = z \right).$$

La dernière probabilité est un terme positif et l'on reconnaît, quand n tend vers l'infini, dans la première somme la définition de ν donc :

$$\mu(z) \geq \mu(a)\nu(z) \text{ pour tout } z \in E.$$

D'autre part, puisque pour tout n , $\mu.P^n = \mu$, on a aussi :

$$\mu(a) = \sum_x \mu(x)p^{(n)}(x, a) \geq \mu(a) \sum_x \nu(x)p^{(n)}(x, a) = \mu(a)\nu(a) = \mu(a).$$

Par conséquent, l'inégalité intermédiaire est une égalité et comme $\mu(x) \geq \mu(a)\nu(x)$, on doit avoir $\mu(x) = \mu(a)\nu(x)$ chaque fois que n est tel que $p^{(n)}(x, a) > 0$. Étant donné que X est irréductible, un tel entier n existe toujours donc $\mu(x) = \mu(a)\nu(x)$ pour tout $x \in E$. \square

THÉORÈME 3.16.— S'il existe une probabilité invariante ν , alors tous les états vérifiant $\nu(y) > 0$ sont récurrents.

Démonstration. Comme $\nu = \nu P^n$, le théorème de Fubini implique que :

$$\sum_{x \in E} \nu(x) \sum_{n \geq 1} p^{(n)}(x, y) = \sum_{n \geq 1} \nu(y) = \infty \text{ si } \nu(y) > 0.$$

D'autre part, d'après le lemme 3.5 :

$$\sum_{n \geq 1} p^{(n)}(x, y) = \frac{\mathbf{P}_x(\tau_y^1 < \infty)}{1 - \mathbf{P}_y(\tau_y^1 < \infty)}.$$

Comme $\mathbf{P}_x(\tau_y^1 < \infty) \leq 1$, on a :

$$\infty \leq \sum_{x \in E} \nu(x) \cdot \frac{1}{1 - \mathbf{P}_y(\tau_y^1 < \infty)},$$

donc $\mathbf{P}_y(\tau_y^1 < \infty) = 1$ puisque ν est finie, ce qui signifie que y est récurrent. \square

THÉORÈME 3.17.— Si X est irréductible et admet ν comme probabilité invariante alors la relation suivante est satisfaite :

$$\nu(x) = \frac{1}{\mathbf{E}_x[\tau_x^1]}.$$

Démonstration. S'il existe x tel que $\nu(x) = 0$ alors comme pour tout n , on a :

$$\nu(x) = \sum_{y \in E} p^{(n)}(y, x)\nu(y),$$

cela signifie que pour tout n et tout y , le produit de $\nu(y)$ et de $p^{(n)}(y, x)$ est nul. Or la chaîne est irréductible donc pour tout y , il existe n_y tel que $p^{(n_y)}(y, x) > 0$, donc $\nu(y) = 0$. Mais alors ν n'est pas une probabilité donc pour tout $x \in E$, $\nu(x) > 0$. D'après le théorème précédent, tous les états sont donc récurrents. On sait donc

que ν est à un coefficient c près donné par le théorème 3.13. Ce coefficient vérifie $c \sum_{y \in E} \nu(y) = 1$, or l'on sait que :

$$\sum_{y \in E} \nu(y) = \sum_{y \in E} \mathbf{E}_x \left[\sum_{n=0}^{\tau_x^1 - 1} \mathbf{1}_{\{X_n=y\}} \right] = \mathbf{E}_x [\tau_x^1],$$

d'après le théorème de Fubini et pour x choisi de façon quelconque dans E . \square

Le théorème suivant résume les principaux résultats précédents.

THÉORÈME 3.18.– Si X est irréductible, les trois assertions suivantes sont équivalentes :

- 1) l'un des états est récurrent positif;
- 2) il existe une probabilité invariante;
- 3) tous les états sont récurrents positifs.

De plus, la probabilité invariante est donnée par :

$$\nu(y) = \frac{1}{\mathbf{E}_x [\tau_x^1]} \mathbf{E}_x \left[\sum_{n=0}^{\tau_x^1 - 1} \mathbf{1}_{\{X_n=y\}} \right].$$

Démonstration. 1) \Rightarrow 2). En combinant les théorèmes 3.13 et 3.17, on voit que :

$$\nu(y) = \frac{1}{\mathbf{E}_x [\tau_x^1]} \mathbf{E}_x \left[\sum_{n=0}^{\tau_x^1 - 1} \mathbf{1}_{\{X_n=y\}} \right]$$

définit une probabilité invariante. Comme le terme de gauche ne dépend pas de x , on peut choisir $x = y$ et on retrouve bien $\nu(y) = \mathbf{E}_y [\tau_y^1]^{-1}$.

2) \Rightarrow 3). Puisque X est irréductible, on sait que la probabilité invariante est un multiple de celle construite dans le théorème 3.13 et donc que $\pi(y) > 0$ pour tout $y \in E$. D'après le théorème 3.17, cela signifie que tous les états sont récurrents positifs.

3) \Rightarrow 1) est trivial. \square

COROLLAIRE 3.19.– Toute chaîne de Markov irréductible sur E de cardinal fini est récurrente positive.

Démonstration. Il existe une mesure invariante μ . Comme l'espace d'états est fini, on peut toujours la normaliser en posant :

$$\nu(x) = \frac{1}{\sum_{y \in E} \mu(y)} \mu(y),$$

et l'on obtient une probabilité invariante. D'après le point 2) du théorème précédent, on en déduit qu'elle est récurrente positive. \square

Quand l'espace d'états est de cardinal infini, on peut utiliser le théorème suivant qui se démontre avec les mêmes outils que le théorème 3.12.

THÉORÈME 3.20 (Critère de Foster).— Supposons qu'il existe $h : E \rightarrow \mathbf{R}$ et $\epsilon > 0$ tels que :

- $\liminf_y h(y) > -\infty$;
- $h(X_1)$ est intégrable ;
- pour tout $y \in E_0^c$, on ait :

$$h(y) - \epsilon \geq \mathbf{E}[h(X_1) | X_0 = y].$$

Dans ces conditions, X est récurrente positive.

Soit X une chaîne de Markov irréductible récurrente sur E polonais. Pour tout $x \in E$, on pose $(Y_k^n, k \geq 0) = (X_{k \wedge \tau_x^n})$ et pour tout $n \geq 1$, on définit la n^{e} excursion Y^n de X depuis x par $(Y_k^n, k \geq 0)$, où :

$$Y_k^n = X_{k \wedge \tau_x^n} \circ \theta^{\tau_x^n}.$$

La 0^e excursion coïncide avec X jusqu'à la première visite à x , après Y^0 reste à x . La n^{e} excursion est une chaîne de Markov qui part de x et suit le comportement de la chaîne initiale jusqu'au temps d'atteinte de x suivant. Ensuite, elle reste constamment égale à x . L'évolution de la chaîne est reflétée par Y^{n+1} .

D'après la propriété de Markov forte, les processus $(Y^n, n \geq 0)$ sont indépendants les uns des autres et à partir de $n = 1$, ils ont tous la même loi : pour toute fonction $\psi : E^{\mathbf{N}} \rightarrow \mathbf{R}$:

$$\mathbf{E}[\psi(Y^n)] = \mathbf{E}[\psi(Y^1 \circ \theta^{\tau_x^n})] = \mathbf{E}[\psi(Y^1)].$$

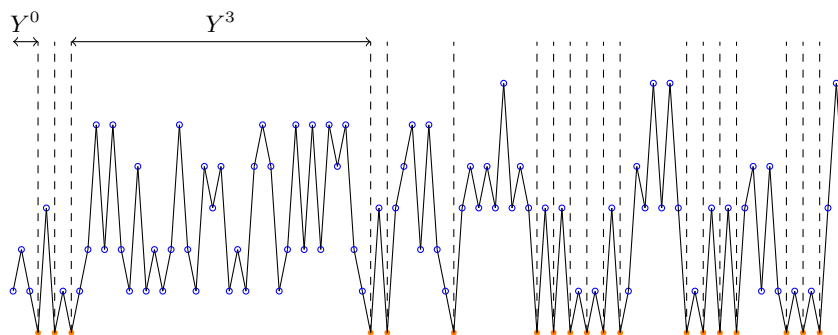


Figure 3.3 – Excursions de la chaîne de Markov de l'exemple 3.1

THÉORÈME 3.21.— Soit X récurrente, irréductible de loi invariante ν . Quelle que soit la condition initiale $x \in E$, pour toute fonction f dans $L^1(\nu)$, pour toute fonction $g \geq 0$ telle que $\sum_y g(y)\nu(y) > 0$, on a :

$$\frac{\sum_{j=0}^n f(X_j)}{\sum_{j=0}^n g(X_j)} \xrightarrow{n \rightarrow \infty} \frac{\sum_{y \in E} f(y)\nu(y)}{\sum_{y \in E} g(y)\nu(y)}, \mathbf{P}_x \text{ presque sûrement.}$$

En particulier, pour $f \in L^1(\nu)$, on a :

$$\frac{1}{n} \sum_{j=0}^n f(X_j) \xrightarrow{n \rightarrow \infty} \sum_{y \in E} f(y)\nu(y), \mathbf{P}_x \text{ presque sûrement.}$$

✎ On peut découper toute fonctionnelle additive en morceaux dépendant de chaque excursion. D'après l'indépendance et l'équidistribution de celles-ci, on peut appliquer la loi forte des grands nombres. Il reste à prouver que les termes de « bord », c'est-à-dire qui dépendent de Y^0 et ceux qui dépendent de l'excursion incomplète disparaissent dans la division par n .

Démonstration. La probabilité invariante est proportionnelle à la mesure invariante donnée dans le théorème 3.13. En particulier, il existe $c > 0$ tel que pour toute fonction $g \geq 0$, on ait :

$$c \sum_{y \in E} g(y)\nu(y) = \mathbf{E}_x \left[\sum_{n=0}^{\tau_x^1 - 1} g(X_n) \right].$$

Par homogénéité, on peut supposer que $c = 1$.

Soit $Z = (Z_k, k \geq 1)$ définie par :

$$Z_k = \sum_{n=\tau_x^k}^{\tau_x^{k+1}-1} f(X_n) = \sum_{n=0}^{\tau_x^1-1} f(Y_n^k) = Z_1 \circ \theta^{\tau_x^k}.$$

D'après la propriété de Markov forte, les variables aléatoires $(Z_k, k \geq 1)$ sont indépendantes et identiquement distribuées. De plus, on a :

$$\mathbf{E}_x [|Z_1|] \leq \mathbf{E}_x \left[\sum_{n=0}^{\tau_x^1-1} |f(X_n)| \right] = \sum_{y \in E} |f(y)| \nu(y) < \infty,$$

puisque $f \in L^1(\nu)$. On peut donc appliquer la loi forte des grands nombres, qui stipule que :

$$\frac{1}{n} \sum_{k=0}^{\tau_x^n-1} f(X_k) = \frac{1}{n} \sum_{k=1}^n Z_k \xrightarrow{n \rightarrow \infty} \mathbf{E}_x [Z_1] = \sum_{y \in E} f(y) \nu(y), \quad \mathbf{P}_x \text{ p.s.} \quad (3.15)$$

Si l'on applique ce résultat à $f \equiv 1$, on obtient :

$$\frac{\tau_x^n}{n} \xrightarrow{n \rightarrow \infty} 1, \quad \mathbf{P}_y \text{ presque sûrement.} \quad (3.16)$$

Soit $e(n)$ le nombre de visites à x entre les instants 0 et n . Notons que $e(n)$ est aussi le nombre d'excursions partant de x complètes avant l'instant n . Par définition, $\tau_x^{e(n)} \leq n < \tau_x^{e(n)+1}$, donc :

$$\frac{\tau_x^{e(n)}}{e(n)} \leq \frac{n}{e(n)} < \frac{\tau_x^{e(n)+1}}{e(n)+1} \frac{e(n)+1}{e(n)}.$$

En vertu de (3.16), les termes extrêmes de la ligne précédente tendent p.s. vers 1 donc $n^{-1}e(n)$ aussi. On écrit alors :

$$\begin{aligned} \frac{1}{n} \sum_{k=0}^n f(X_k) = \\ \frac{1}{n} \sum_{k=0}^{\tau_x^1-1} f(X_k) + \frac{e(n)}{n} \frac{1}{e(n)} (Z_1 + \dots + Z_{e(n)}) + \frac{1}{n} \sum_{k=\tau_x^{e(n)}+1}^n f(X_k). \end{aligned} \quad (3.17)$$

Supposons que $f \geq 0$. Le premier terme tend vers 0 \mathbf{P}_y -presque sûrement pour tout $y \in E$. D'après la définition de la convergence presque sûre, cela équivaut à ce que l'on ait pour tout $\epsilon > 0$:

$$\mathbf{P}_y \left(\limsup_n \left(\frac{1}{n} \sum_{k=0}^{\tau_x^1-1} f(X_k) > \epsilon \right) \right) = 0. \quad (3.18)$$

Par conséquent, compte tenu de la propriété de Markov forte, pour tout $\epsilon > 0$, on a :

$$\begin{aligned} \mathbf{P}_y \left(\limsup_n \left(\frac{1}{n} \sum_{k=\tau_x^{e(n)}}^{\tau_x^{e(n+1)}-1} f(X_k) > \epsilon \right) \right) \\ = \mathbf{P}_x \left(\limsup_n \left(\frac{1}{n} \sum_{k=0}^{\tau_x^1-1} f(Y_k^n) > \epsilon \right) \right) \\ = \mathbf{P}_x \left(\limsup_n \left(\frac{1}{n} \sum_{k=0}^{\tau_x^1-1} f(X_k) > \epsilon \right) \right) = 0, \end{aligned}$$

en vertu de (3.18). Comme :

$$\sum_{k=\tau_x^{e(n)}+1}^n f(X_k) \leq \sum_{k=\tau_x^{e(n)}}^{\tau_x^{e(n+1)}-1} f(X_k),$$

on a bien :

$$\frac{1}{n} \sum_{k=\tau_x^{e(n)}}^n f(X_k) \xrightarrow{n \rightarrow \infty} 0, \mathbf{P}_y \text{ presque sûrement.}$$

Pour f de signe quelconque, en appliquant le raisonnement précédent à $|f|$, on montre que les premier et troisième termes de (3.17) tendent presque sûrement vers 0.

D'après la première partie de la preuve (voir (3.15)), le terme médian de (3.17) tend presque sûrement vers $\sum_{y \in E} f(y)\nu(y)$.

Le cas particulier s'obtient en prenant $g = 1$. □

DÉFINITION 3.10.– Un état x est dit périodique s'il existe un entier $\delta \geq 2$ tel que :

$$\sum_{k=1}^{\infty} \mathbf{P}_x(\tau_x^1 = \delta k) = 1. \quad (3.19)$$

Le plus petit δ vérifiant (3.19) est appelé la période de x et nous la noterons $d(x)$. Les états qui ne sont pas périodiques sont appelés apériodiques.

EXEMPLE (Suite de l'exemple 3.1).– Dans ce cas, un peu de réflexion montre que l'on n'atteindra une case de numéro impair que tous les deux pas, de même pour les cases de numéro pair : si le rat part de 1 il ne peut être en case 3, 5 ou 7 au coup d'après. La période est donc 2. On voit que l'on peut classer les états en deux paquets, les cases $\{1, 3, 5, 7\}$ d'une part, les cases $\{2, 4, 6\}$ d'autre part. Le rat sera tous les pas pairs dans le sous-ensemble de départ et tous les pas impairs dans l'autre sous-ensemble.

Plus généralement, on a le théorème suivant.

THÉORÈME 3.22.— Soit X une chaîne de Markov récurrente irréductible de période d . Soit x fixé dans E , il existe une partition de E en d ensembles C_0, C_1, \dots, C_{d-1} tels que :

- x appartient à C_0 ;
- soit $y \in C_r$ et $z \in C_s$, si $p^{(n)}(y, z) > 0$ alors $n = (s - r) \pmod{d}$;
- C_0, \dots, C_{d-1} sont des classes de récurrence irréductibles apériodiques pour la chaîne de Markov de matrice de transition P^d .

La décomposition est unique à une renumérotation près. La chaîne de Markov de matrice de transition P^d est irréductible, récurrente, apériodique. Si sa condition initiale est dans C_r pour $r \in \llbracket 0, d - 1 \rrbracket$, alors tous ces éléments sont dans C_r .

La démonstration de ce théorème nécessite deux lemmes techniques.

LEMME 3.23.— Soit a_1, \dots, a_n des nombres entiers premiers dans leur ensemble, tout entier $m \geq \prod_x (1 + a_x)$ peut s'écrire sous la forme :

$$m = \sum_k x_k a_k \text{ avec } x_k \geq 0 \text{ pour tout } k. \quad (3.20)$$

Démonstration. Montrons par récurrence sur n que si a_1, \dots, a_n sont n entiers (non nécessairement premiers entre eux) et que $m \in \mathbf{N}$ s'écrit $m = \sum_k x_k a_k$ alors on peut toujours trouver une autre écriture satisfaisant les conditions de (3.20). Plus précisément, il existe une permutation σ de $\{1, \dots, N\}$ dans lui-même telle que :

$$x_{\sigma(i)} \leq \min_{l \neq \sigma(1), \dots, \sigma(i-1)} (a_l) \text{ pour tout } l \leq n - 1.$$

Supposons d'abord que $n = 2$. Comme $m \geq 0$, l'un des deux coefficients x_1 ou x_2 est positif. A une renumérotation près, on peut toujours supposer qu'il s'agit de x_1 . Montrons que l'on peut toujours supposer que $x_1 < a_2$. Si ce n'est pas le cas, on écrit alors $x_1 = ka_2 + r$ avec $0 \leq r < a_2$:

$$\begin{aligned} m &= x_1 a_1 + x_2 a_2 + ka_1 a_2 - ka_1 a_2 \\ &= (x_1 - ka_2) a_1 + (x_2 + ka_1) a_2 = r a_1 + (x_2 + ka_1) a_2. \end{aligned}$$

En conclusion, tout entier m peut s'écrire sous la forme $m = x_1 a_1 + x_2 a_2$ avec $0 \leq x < b$. En particulier, si $m \geq a_1 a_2$, x_2 doit être positif.

Supposons le résultat démontré pour $(n - 1)$. A une renumérotation près, on peut toujours supposer que x_1 est positif et appliquer l'hypothèse de récurrence à $m - x_1 a_1$

et aux $(n - 1)$ nombres restants. La renumérotation que l'on a appliquée au cours de cette opération définit la permutation σ .

Maintenant, si a_1, \dots, a_n sont premiers dans leur ensemble, le lemme de Bezout garantit l'existence de la représentation $m = \sum_k x_k a_k$ pour tout entier. D'après la première partie de la démonstration, on peut toujours supposer que $\sum_{k \leq n-1} x_k a_k$ est positive et inférieure à :

$$\sup_x \left(a_1 \dots a_n + \prod_{y \neq x} a_y + \dots \right) \leq a_1 \dots a_n + \sum_x \prod_{y \neq x} a_y + \dots = \prod (1 + a_x) - 1.$$

Par conséquent, pour m supérieur ou égal à $\prod (1 + a_x)$, il existe toujours une écriture de la forme (3.20). \square

LEMME 3.24.— Si x est apériodique alors il existe n_0 tel que si $n \geq n_0$ alors $p^{(n)}(x, x) > 0$.

Démonstration. Définissons l'ensemble :

$$I_x = \left\{ n \in \mathbf{N}, p^{(n)}(x, x) > 0 \right\}.$$

D'après la propriété de Markov, I_x est un semi-groupe : si m et n appartiennent à I_x alors $m + n$ aussi. En effet, on a :

$$p^{(m+n)}(x, x) \geq p^{(m)}(x, x)p^{(n)}(x, x).$$

On ordonne I_x par l'ordre naturel. Soit u_n le nombre de diviseurs communs des n premiers éléments de I_x . $(u_n, n \in \mathbf{N})$ est une suite décroissante positive donc convergente, et puisque x est apériodique, cette limite est 1. Comme u_n est à valeurs entières, il y a nécessairement un rang à partir duquel elle est constante, soit n_0 ce rang et soit a_1, \dots, a_{n_0} les n_0 premiers éléments de I_x . D'après le lemme précédent, pour n assez grand, $n \in I_x$. \square

Démonstration du théorème 3.22. Soit $K_y = \left\{ n, p^{(n)}(x, y) > 0 \right\}$. Pour k et l deux entiers, d'après la propriété de Markov, on a l'identité suivante :

$$\mathbf{P}_x(X_{k+l} = x) \geq \mathbf{P}_x(X_k = y)\mathbf{P}_y(X_l = x).$$

Par conséquent, n ne peut appartenir à K_y que si d divise $n + l$, c'est-à-dire si n s'écrit $ad + r$ où $r \in \{0, \dots, d - 1\}$ est le reste de la division de l par d . On définit C_r comme l'ensemble des points de E qui ont le même r . Ces ensembles forment clairement une partition et en prenant $l = 0$, on prouve que $x \in C_0$.

Soit m et n tels que $p^{(m)}(y, z) > 0$ et $p^{(n)}(x, y) > 0$. Comme $p^{(n+m)}(x, z) > 0$, il en découle d'après i) que $n + m \equiv s \pmod{d}$ et comme $n \equiv r \pmod{d}$, le résultat s'ensuit.

L'irréductibilité découle immédiatement du point précédent, l'apériodicité de la définition de la période. \square

Nous pouvons maintenant énoncer le résultat.

THÉORÈME 3.25.– Soit X une chaîne de Markov irréductible, récurrente positive, apériodique, de matrice de transition P et de probabilité invariante ν . Alors, on a :

$$\lim_{n \rightarrow \infty} p^{(n)}(x, y) = \nu(y), \text{ pour tout } x \text{ et tout } y.$$

✎ La démonstration se fait par couplage : on va montrer que deux chaînes de Markov indépendantes de même matrice de transition mais de conditions initiales différentes finissent toujours par se rencontrer. Notons qu'à partir de cet instant de rencontre, elles coïncident en loi.

Démonstration. Sur $E \times E$, on définit la chaîne de Markov $Z_n = (W_n, y_n)$ de matrice de transition :

$$\hat{p}((x_1, x_2), (y_1, y_2)) = p(x_1, y_1)p(x_2, y_2).$$

En d'autres termes, les deux coordonnées W et Y évoluent indépendamment l'une de l'autre selon la loi de la chaîne de Markov originelle.

On va montrer premièrement que Z est une chaîne de Markov irréductible. Comme tous les états sont apériodiques, d'après le lemme 3.24, à partir d'un certain rang M :

$$p^{(l)}(y_1, y_1) > 0 \text{ et } p^{(l)}(x_2, x_2) > 0.$$

Comme X est irréductible et récurrent, il existe $K \geq M$ et $L \geq M$ tels que :

$$p^{(K)}(x_1, x_2) > 0 \text{ et } p^{(L)}(y_1, y_2) > 0.$$

Par conséquent, le chemin :

$$(x_1, y_1) \rightarrow (x_2, y_1) \rightarrow (x_2, y_2)$$

est de probabilité strictement positive pour l'indice $K + L + M$. En effet, d'après la propriété de Markov :

$$\begin{aligned} p^{(K+L)}((x_1, y_1), (x_2, y_2)) \\ \geq p^{(K)}(x_1, x_2)p^{(K)}(y_1, y_1) \cdot p^{(L)}(x_2, x_2)p^{(L)}(y_1, y_2) > 0. \end{aligned}$$

Il est clair que $\hat{\nu}(x, y) = \nu(x)\nu(y)$ définit une probabilité invariante pour la chaîne de Markov Z . Par conséquent d'après le théorème 3.18, tous les états sont récurrents positifs. Soit T le temps d'atteinte de la diagonale de $E \times E$ par Z :

$$\begin{aligned}\Delta &= \{(x, y) \in E \times E, x = y\}; \\ T &= \inf \{n > 0, Z_n \in \Delta\}.\end{aligned}$$

Comme Z est irréductible, récurrente, le temps d'atteinte d'un état (x, x) de la diagonale est presque sûrement fini. Comme T est le minimum de tous ces temps, il est aussi presque sûrement fini. Montrons que sur $\{T \leq n\}$, W_n et Y_n ont même loi :

$$\begin{aligned}\mathbf{P}(W_n = y, T \leq n) &= \sum_x \mathbf{E} [\mathbf{1}_{\{W_n=y\}} \mathbf{1}_{\{W_T=x\}} \mathbf{1}_{\{T \leq n\}}] \\ &= \sum_x \mathbf{E} [\mathbf{1}_{\{W_T=x\}} \mathbf{1}_{\{T \leq n\}} \mathbf{E} [\mathbf{1}_{\{W_n=y\}} \mid \mathcal{F}_T]] \\ &= \sum_x \mathbf{E} [\mathbf{1}_{\{W_T=x\}} \mathbf{1}_{\{T \leq n\}} \mathbf{E}_x [\mathbf{1}_{\{W_{n-T}=y\}}]] \\ &= \sum_x \mathbf{E} [\mathbf{1}_{\{Y_T=x\}} \mathbf{1}_{\{T \leq n\}} \mathbf{E}_x [\mathbf{1}_{\{Y_{n-T}=y\}}]] \\ &= \mathbf{P}(Y_n = y, T \leq n).\end{aligned}$$

D'après ce qui précède :

$$\begin{aligned}\mathbf{P}(W_n = y) &= \mathbf{P}(W_n = y, T \leq n) + \mathbf{P}(W_n = y, T > n) \\ &= \mathbf{P}(Y_n = y, T \leq n) + \mathbf{P}(W_n = y, T > n) \\ &\leq \mathbf{P}(Y_n = y) + \mathbf{P}(W_n = y, T > n).\end{aligned}$$

Symétriquement, on a :

$$\mathbf{P}(Y_n = y) \leq \mathbf{P}(W_n = y) + \mathbf{P}(Y_n = y, T > n),$$

dont on déduit que :

$$|\mathbf{P}(W_n = y) - \mathbf{P}(Y_n = y)| \leq \mathbf{P}(Y_n = y, T > n) + \mathbf{P}(W_n = y, T > n).$$

D'où en sommant sur toutes les valeurs possibles de y , on obtient :

$$\sum_y |\mathbf{P}(Y_n = y) - \mathbf{P}(W_n = y)| \leq 2\mathbf{P}(T > n).$$

Comme T est presque sûrement fini, le terme de droite tend vers 0 quand n croît indéfiniment. Si l'on prend $W_0 = x$ et Y ayant la loi ν , on en déduit :

$$\sum_y |p^{(n)}(x, y) - \nu(y)| \xrightarrow{n \rightarrow \infty} 0,$$

ce qui implique le résultat annoncé. \square

REMARQUE.— On remarque que l'hypothèse d'apériodicité ne sert que dans la démonstration de l'irréductibilité de la chaîne de Markov Z . Pour se convaincre que cela est essentiel, regardons encore l'exemple du rat dans son labyrinthe. Formons la chaîne de Markov $Z_n = (X_n, Y_n)$ qui représente les positions de deux rats lâchés dans le même labyrinthe, qui évoluent indépendamment l'un de l'autre selon les mêmes règles que précédemment. Soit C_1 la classe cyclique de 1 et C_2 celle de 2 pour la chaîne de Markov X . Si Z part d'un état de $C_1 \times C_2$ alors Z évolue entre des états de cet ensemble et des états de $C_2 \times C_1$ mais n'atteint jamais d'états de $C_1 \times C_1$, par conséquent Z n'est pas irréductible.

Dans le cas périodique, on a toutefois le résultat suivant :

THÉORÈME 3.26.— Soit X une chaîne de Markov irréductible, récurrente positive, périodique de période d et de probabilité invariante ν . Soit $x \in E$ et C_0, \dots, C_{d-1} les classes cycliques associées à x . Si $y \in C_r$, on a :

$$\lim_{n \rightarrow \infty} p^{(nd+r)}(x, y) = d\nu(y).$$

✎ L'idée est d'appliquer le théorème précédent à la chaîne de matrice de transition P^d . Il faut pour cela déterminer la probabilité invariante de cette chaîne de Markov. On remarque d'après le théorème 3.13 qu'à un coefficient près la probabilité invariante d'un état y est égal à la proportion du nombre de visites à cet état entre deux visites à un état fixe x . Etant donné que dans la chaîne de Markov de matrice P^d on divise par d le nombre de pas, cette proportion est multiplié par d .

Démonstration. Par définition de la période et de C_k , C_k est un sous-ensemble fermé pour la chaîne X^k définie par $X_n^k = X_{nd+k}$ pour $k = 0, \dots, d-1$. Ces chaînes sont irréductibles et récurrentes positives. En vertu du corollaire 3.14, la probabilité invariante ν^k de X^k est proportionnelle à ν , c'est-à-dire qu'il existe α_k tel que $\nu^k(y) = \alpha_k \nu(y)$ pour tout $y \in C_k$. Par ailleurs, puisque ν est la probabilité invariante de X , pour tout k et tout l appartenant à $0, \dots, d-1$, on a :

$$\alpha_k = \mathbf{P}_\nu(X_{nd+k} \in C_k) = \mathbf{P}_\nu(X_{nd+k} \in C_k \cup C_l) = \mathbf{P}_\nu(X_{nd+l} \in C_l) = \alpha_l.$$

Il s'ensuit que $\alpha_k = d^{-1}$. Le dernier point découle du théorème 3.25. \square

Enfin, le dernier résultat utile pour les simulations est le théorème central limite suivant.

THÉORÈME 3.27.— Soit X une chaîne de Markov récurrente positive de probabilité invariante ν . Pour $f : E \times E \rightarrow \mathbf{R}$, on note :

$$Pf : \begin{cases} E & \longrightarrow \mathbf{R} \\ x & \longmapsto Pf(x) = \sum_y f(x, y)p(x, y) = \mathbf{E}_x [f(X_0, x_1)]. \end{cases}$$

Pour toute fonction f telle que $\mathbf{E}_\nu[P(f^2)] < \infty$ on a la convergence en loi suivante :

$$\frac{1}{\sqrt{n}} \sum_{k=1}^n (f(X_{k-1}, X_k) - Pf(X_{k-1})) \xrightarrow[n \rightarrow \infty]{\text{Loi}} \mathcal{N}(0, \sigma^2(f)),$$

où $\mathcal{N}(0, \sigma^2(f))$ représente une loi gaussienne centrée de variance donnée par :

$$\sigma^2(f) = \mathbf{E}_\nu[P(f^2)] - \mathbf{E}_\nu[(Pf)^2].$$

Démonstration. Le lemme 3.12 implique que pour f bornée, la suite :

$$M_n^f = \sum_{j=0}^n f(X_j, X_{j+1}) - \sum_{j=0}^n Pf(X_j, \cdot)(X_j)$$

est une martingale. D'autre part, son processus croissant est défini par :

$$\begin{aligned} \Delta \langle M^f \rangle_n &= \mathbf{E}[(\Delta M_n^f)^2 | \mathcal{F}_n] \\ &= \mathbf{E}\left[\left(f(X_n, X_{n+1}) - Pf(X_n, \cdot)(X_n)\right)^2 | \mathcal{F}_n\right] \\ &= Pf^2(X_n) + Pf(X_n)^2 - 2Pf(X_n)^2 \\ &= \Gamma f(X_n), \end{aligned}$$

où $\Gamma f = Pf^2 - (Pf)^2$ est l'opérateur carré du champ associé à P . Par conséquent, on a :

$$\langle M^f \rangle_n = \sum_{j=0}^n \Gamma f(X_j).$$

Par hypothèse, Γf est intégrable par rapport à ν , le théorème 3.21 implique que :

$$\frac{\langle M^f \rangle_n}{n} \xrightarrow[n \rightarrow \infty]{} \sigma^2(f), \mathbf{P}_x \text{ presque sûrement.}$$

Le résultat découle du théorème de la limite centrée pour les incréments de martingales. \square

Si l'on prend comme cas particulier, $f(X_{k-1}, X_k) = \mathbf{1}_{\{X_k=x\}}$, on obtient :

$$\mathbf{P}(\sqrt{n}(N_x^n - \pi(x)) \in [a, b]) \xrightarrow[n \rightarrow \infty]{} \int_a^b \exp(-x^2/2\sigma^2) \frac{dx}{\sigma\sqrt{2\pi}},$$

avec $\sigma^2 = \nu(y) - \sum_x p(x, y)^2 \nu(x)$.

EXEMPLE (Suite de l'exemple 3.1).— C'est le cas le plus simple dans lequel on n'a qu'à résoudre le système $\nu = \nu P$ et $\sum \pi(x) = 1$. Tous calculs faits, on trouve :

$$\nu = \left(\frac{1}{8}, \frac{3}{16}, \frac{1}{8}, \frac{3}{16}, \frac{3}{16}, \frac{1}{8}, \frac{1}{16} \right).$$

EXEMPLE (Suite de l'exemple 3.3).— Il faut restreindre la chaîne de Markov à une quelconque classe d'équivalence de la relation « communique ». Dans ce cas, il est clair que la probabilité invariante est la mesure uniforme sur ces états.

3.5. Calcul pratique de la probabilité invariante

Le principe est simple : la probabilité invariante est l'unique vecteur à composantes positives ou nulles, de poids total 1 qui satisfait l'équation $\nu(P - \text{Id}) = 0$. Si l'on veut résoudre un tel système par ordinateur, il faut prendre garde au fait que ce système est de co-rang 1 : il faut donc supprimer une colonne de P (par exemple, la dernière) et la remplacer par une colonne composée uniquement de 1. Soit \hat{P} la matrice ainsi obtenue. Il nous faut alors résoudre le système :

$$\pi(\hat{P} - \hat{I}) = b, \text{ avec } b = (0, \dots, 0, 1) \text{ et } \hat{I} = \begin{pmatrix} 1 & & & 0 \\ & 1 & (0) & 0 \\ & (0) & \ddots & \\ & & & 0 \end{pmatrix}.$$

Pratiquement, les chaînes que l'on utilise ont un espace d'états fini mais de cardinal très grand (plusieurs milliers d'états). Se pose alors le problème de la détermination de la probabilité invariante. Il s'agit *a priori* « simplement » de résoudre un système linéaire mais la taille de ce système oblige à utiliser des méthodes d'analyse numérique.

3.5.1. Méthode itérative

On a donc à résoudre l'équation $\pi = \pi P$ où P est la matrice de transition. D'après le théorème 3.25, si la chaîne est apériodique alors $\pi_{n+1} = \pi_n P$ tend vers la probabilité invariante. Pratiquement, on prend un π_0 quelconque et on itère. Ce procédé peut s'avérer coûteux si le calcul des coefficients de P est long. Néanmoins, la convergence est extrêmement rapide puisque géométrique de raison le module de la deuxième plus grande valeur propre de P .

Dans le cas où la chaîne est périodique (voir l'exemple du rat) de période d , il faut être plus précautionneux. Le théorème 3.25 nous indique que la suite π_n a d valeurs d'adhérence. Précisément, par définition même des classes cycliques, si π_0 est une masse de Dirac en x , les termes π_{kn} ont des composantes positives seulement pour les états de la classe cyclique de x , les termes π_{kn+j} ont des composantes positives seulement pour les états de la classe cyclique C_j , pour tout $j \in \{1, \dots, d-1\}$.

EXEMPLE (Suite de l'exemple 3.1).– Si la condition initiale est $\pi_0 = (1, 0, \dots)$ alors on a :

$$\begin{aligned} v(2) &= \left(0, \frac{1}{2}, 0, \frac{1}{2}, 0, 0, 0\right) \\ v(3) &= \left(\frac{1}{3}, 0, \frac{1}{6}, 0, \frac{1}{3}, 0, \frac{1}{6}\right) \\ v(4) &= \left(0, \frac{13}{36}, 0, \frac{4}{9}, 0, \frac{7}{36}, 0\right) \\ v(5) &= \left(\frac{29}{108}, 0, \frac{47}{216}, 0, \frac{79}{216}, 0, \frac{4}{27}\right) \\ v(6) &= \left(0, \frac{473}{1296}, 0, \frac{131}{324}, 0, \frac{299}{1296}, 0\right) \\ v(7) &= \left(\frac{997}{3888}, 0, \frac{1843}{7776}, 0, \frac{2891}{7776}, 0, \frac{131}{972}\right) \end{aligned}$$

Un moyen pour éviter ce désagrément est de considérer les sommes de Césaro, $\hat{\pi}_n = d^{-1} \sum_{i=n-d}^n \pi_i$. Cela exige de connaître la période, si c'est impossible on peut alors utiliser la méthode ergodique.

3.5.2. Méthode ergodique

Le théorème 3.21 affirme que pour une chaîne de Markov irréductible et récurrente positive de probabilité invariante ν , on a pour toute condition initiale et tout $x \in E$:

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n \mathbf{1}_{\{X_k=x\}} = \nu(x).$$

On simule donc une trajectoire de la chaîne de Markov le plus longtemps possible et on compte le pourcentage des fois où l'on passe dans l'état x . Le théorème 3.27 indique que la vitesse de convergence est en $1/\sqrt{n}$ ce qui se compare très désavantageusement avec les deux méthodes précédentes. En échange, on n'a pas à stocker tous les $\nu(x)$ mais seules les fréquences des valeurs qui nous intéressent. Il est en fait très fréquent que seules nous intéressent quelques composantes de ν .

EXEMPLE (File M/GI/1/K).– Il y a un buffer de taille K et l'on note X_n le nombre de clients dans le système juste après le départ du client n . Alors $(X_n, n \in \mathbf{N})$ vérifie la récurrence $X_{n+1} = \min((X_n - 1)^+ + A_{n+1}, K + 1)$. On a donc bien une chaîne de Markov irréductible et nécessairement récurrente puisque l'espace d'états est fini. On ne peut pas facilement calculer les probabilités invariantes par la méthode de la fonction

génératrice à cause des effets de bord. Cependant, pour le dimensionnement du buffer seule nous intéresse la probabilité de perte, c'est-à-dire $\nu(K + 1)$ et éventuellement le nombre moyen de clients dans la file. Ces deux quantités s'obtiennent comme suit :

$$\nu(K + 1) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n \mathbf{1}_{\{X_k = K+1\}} \text{ et } \hat{N} = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n X_k.$$

3.6. Problèmes

EXERCICE 1.– Sur un échiquier de 8 par 8, on place un cavalier dans le coin $A1$, on suppose que le cavalier se déplace au hasard (il choisit une direction au hasard parmi celles possibles à chaque coup) et sans mémoire. On rappelle qu'un cavalier se déplace de deux cases dans une direction (horizontale ou verticale) et d'une case dans l'autre direction. En utilisant la réversibilité et des considérations de symétrie, calculer le temps moyen de retour à la case $A1$.

Même question, si l'on identifie les bords opposés de l'échiquier; le cavalier se déplace alors sur un tore !

EXERCICE 2.– Construire (dans les cas où c'est possible) une chaîne de Markov à deux états telle que :

- les deux états soient récurrents ;
- les deux états soient transients ;
- l'un des états soit transient, l'autre récurrent ;
- les deux soient transients ;
- les deux soient récurrents nuls.

EXERCICE 3.– On considère la chaîne de Markov à valeurs dans $\{1, 2, 3\}$ dont la matrice de transition est donnée par :

$$\begin{pmatrix} 0 & 1/2 & 1/2 \\ f(p) & 0 & 1 - f(p) \\ 1 - f(p) & 0 & f(p) \end{pmatrix}$$

où $p \in [0, 1]$ et $f(p)$ est définie par :

$$f(p) = \begin{cases} 0 & \text{si } p \leq 1/4 \\ 2p - 1/2 & \text{si } 1/4 < p \leq 3/4 \\ 1 & \text{si } p \geq 3/4. \end{cases}$$

- 1) Donner la classification des états en fonction des valeurs de p .

2) Pour quelles valeurs de p existe-il une probabilité invariante ? La calculer lorsqu'elle existe.

3) Partant de 2, quelle est, en fonction de p , la valeur du temps moyen de retour en 2 ?

4) Soit h la fonction définie par :

$$h(1) = -1, h(2) = 1, h(3) = 1.$$

Que vaut la limite :

$$\frac{1}{n} \sum_{j=1}^n h(X_j)$$

quand n tend vers $+\infty$ pour $p < 3/4$?

5) Si l'on dispose d'un nombre arbitrairement grand d'exemples de trajectoires, comment sait-on si p est supérieur à $3/4$? Comment sait-on si $p < 1/4$? Comment peut-on estimer p dans le cas où il est compris entre $1/4$ et $3/4$?

EXERCICE 4.– Soit X , une chaîne de Markov irréductible et récurrente sur E , et soit F un sous-ensemble de E . Montrer que la chaîne restreinte à F ($X_n^F, n \in \mathbf{N}$) (voir définition 3.8) est une chaîne de Markov sur E .

EXERCICE 5.– On considère la chaîne de Markov homogène X à deux états A et B de matrice de transition :

$$P = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix}.$$

On cherche le temps de première apparition de la séquence ABA . Pour ce faire, on construit le processus $Y_n = (X_n, X_{n+1}, X_{n+2})$.

1) Montrer que Y est une chaîne de Markov homogène dont on donnera la matrice de transition (sous forme de matrice ou de graphe).

2) Cette chaîne est-elle irréductible ? apériodique ? récurrente positive ?

3) Calculer la probabilité invariante de Y . On pourra numéroter les états dans l'ordre lexicographique : $AAA = 1, AAB = 2, \dots$

4) En déduire le temps moyen (à l'état invariante) qui s'écoule entre deux occurrences de ABA .

5) On suppose que $X_0 = A, X_1 = B$. Donner les équations qui permettent de calculer $\mathbf{E} [\tau_{ABA}^1]$. Il n'est pas demandé de les résoudre.

EXERCICE 6.– Soit un équipement qui émet, sur une ligne de transmission, des paquets de taille constante. On note T le temps de transmission d'un paquet. Par la suite on considère un modèle à temps discret du système, c'est-à-dire, un modèle pour lequel le temps est divisé en intervalles de longueur constante, que nous supposons égale à T . On appelle chaque intervalle un slot. La ligne de transmission peut introduire des erreurs et

on définit une suite $(Y_n, n \in \mathbf{N})$ tel que $Y_n = 1$ si, à l'instant $n + 1$, la ligne est dans un état pour lequel elle introduit des erreurs et $Y_n = 0$ si, à l'instant $n + 1$, elle est dans un état où elle n'introduit pas des erreurs. On suppose que $(Y_n, n \in \mathbf{N})$ est une chaîne de Markov invariante et que $P(Y_1 = 1 | Y_0 = 1) = 0,9$ et $P(Y_1 = 0 | Y_0 = 0) = 0,1$.

L'émission se fait avec le protocole « arrêt et attente » (*stop and wait*). Selon ce protocole, chaque paquet doit être acquitté. S'il n'y a pas d'erreur, le paquet est acquitté positivement et le paquet suivant peut être transmis. En cas contraire, le paquet doit être retransmis. Pour simplifier le problème nous considérons que l'acquiescement arrive instantanément.

1) Calculer la distribution de probabilité invariante de Y_n .

2) On suppose que les paquets arrivent selon un processus géométrique. C'est-à-dire qu'au n^{e} slot il y a une arrivée avec probabilité q et aucune arrivée avec probabilité $1 - q$. Un paquet peut être transmis au même slot que celui où il arrive. Soit X_n le nombre de paquets dans le système, au slot n , après la transmission (s'il y a un paquet à transmettre). Le couple (X_n, Y_n) est une chaîne de Markov. On ordonne les états dans l'ordre lexicographique, c'est-à-dire :

0 1 2 3 4 5 6 7 8 9 ...
00 01 10 11 20 21 30 31 40 41 ...

3) Trouver Q la matrice de transition de (X_n, Y_n) .

4) Montrer que :

$$\begin{aligned} \nu_0 &= 1, \\ \nu_{2n} &= 9 \left(3\sqrt{q/1-q} \right)^{2n}, \\ \nu_{2n+1} &= 9\nu_{2n} \end{aligned}$$

est une mesure invariante pour la chaîne de Markov (X_n, y_n) .

5) Trouver l'ensemble de valeurs de q pour lesquelles tous les états sont récurrents positifs. Comparer le résultat obtenu avec le résultat de 1. Conclure.

EXERCICE 7.— On considère un paquet de N cartes. Pour le mélanger, on procède de la manière suivante : on choisit une carte au hasard parmi les N et on met cette carte sur le dessus du paquet sans bouger les ordres relatifs des autres.

1) Comment représenter l'état du paquet, noté X_n , après la n^{e} opération ?

2) En introduisant les permutations particulières :

$$\tau_k = \begin{pmatrix} 1 & 2 & \dots & k-1 & k & k+1 & \dots & N \\ 2 & 3 & \dots & k & 1 & k+1 & \dots & N \end{pmatrix},$$

pour $k \in \{1, \dots, N\}$; écrire les probabilités de transition de X .

L'écriture de τ_k signifie que $\tau_k(1) = 2, \tau_k(2) = 3, \dots$, ce que l'on peut aussi noter :

$$\tau_k = (2, 3, \dots, k, 1, k+1, \dots, n).$$

3) Montrer que cette chaîne est irréductible (on pourra avantageusement raisonner pour de petites valeurs de N comme $N = 4$ par exemple).

4) Montrer qu'au bout d'un nombre suffisamment grand d'opérations on obtient un « bon » mélange, caractérisé par l'équiprobabilité de tous les états possibles du paquet de cartes. (Cette question comporte deux parties : formaliser mathématiquement le problème puis le résoudre.)

EXERCICE 8.– On pose $E = \{1, \dots, 10\}$. On définit sur E , l'addition comme l'addition modulo 10, c'est-à-dire que $10 + 1 = 1$. On considère X , la chaîne de Markov de matrice de transition $P = (p_{i,j})$ donnée par :

$$p_{i,i+1} = p, p_{i,i-1} = 1 - p.$$

On suppose que p n'est égal ni à 0, ni à 1.

- 1) Cette chaîne est-elle irréductible ? récurrente ? apériodique ?
- 2) Quelle est sa probabilité invariante ?

On considère maintenant X_1 et X_2 deux copies indépendantes de cette chaîne. On pose $Y = (X_1, X_2)$.

- 3) Cette chaîne est-elle irréductible ?
- 4) Quelle est sa probabilité invariante ?
- 5) On pose maintenant $Z_n = Y_{2n}$. Cette chaîne est-elle irréductible ? Quels sont ces sous-ensembles fermés ? Est-elle récurrente ? apériodique ?

EXERCICE 9.– Soit $A = \{A_n : n \geq 1\}$ une suite de variables aléatoires indépendantes et identiquement distribuées à valeurs dans \mathbf{R}^k , soit h une application de $E \times \mathbf{R}^k$ dans E et soit X_0 une variable aléatoire indépendante de la suite A . On définit la suite $X = \{X_n : n \in \mathbf{N}\}$ par X_0 pour $n = 0$ et par $X_n = h(X_{n-1}, A_n)$, pour $n \geq 1$. Montrer que X est une chaîne de Markov.

3.7. Notes et commentaires

Le nombre d'ouvrages traitant des chaînes de Markov est incalculable, on ne saurait les citer tous. Parmi les plus récents et qui s'approchent le plus ou complètent notre approche, on peut noter [BAL 01, GRA 08]. Les chaînes de Markov constituent toujours un champ d'investigation très actif en raison de leur universalité. Les problèmes actuels se focalisent autour du calcul de la vitesse de convergence vers la probabilité stationnaire et son lien avec le « trou spectral », la réduction de l'espace d'états pour calculer plus facilement une approximation de la probabilité invariante, les applications en simulation et statistiques à travers les méthodes MCMC.

Résumé

- Une chaîne de Markov est définie par sa loi initiale ν et son opérateur de transition P .
- Un état récurrent est un état visité une infinité de fois. Un état transient est un état visité un nombre fini de fois.
- Deux états reliés entre eux ont même nature, c'est-à-dire transient ou récurrent.
- Une chaîne est irréductible si tous les états sont liés entre eux.
- Une mesure stationnaire s'identifie à un vecteur ligne solution de l'équation $\pi P = \pi$.
- Si l'on peut trouver π tel que $\sum_{x \in E} \pi(x) = 1$ alors π est une probabilité invariante, la chaîne est récurrente.
- Dans ce cas, quelle que soit la condition initiale, $\mathbf{P}(X_n = x) \xrightarrow{n \rightarrow \infty} \pi(x)$.
- Pour calculer π , on peut au choix résoudre le système $\pi P = \pi, \sum_{x \in E} \pi(x) = 1$ ou considérer la limite de la suite $\pi_{n+1} = \pi_n P, \pi_0$ quelconque.

Chapitre 4

Files d'attente stationnaires

Dans les chapitres 8 et 9, on suppose que les lois génériques des suites des interarrivées $(\xi_n, n \in \mathbf{N})$ et des temps de service $(\sigma_n, n \in \mathbf{N})$ sont exponentielles. Cette hypothèse permet d'étudier de nombreux modèles assez simplement, à travers l'étude en temps continu de processus markoviens de sauts. De nombreux résultats quantitatifs peuvent alors être obtenus sur les performances de ces systèmes.

Malheureusement, cette hypothèse n'est pas réaliste dans de nombreux cas et l'on est amené à considérer des suites de variables aléatoires indépendantes et identiquement distribuées de lois générales. Selon la nomenclature de Kendall, ce sont donc les modèles de type GI/GI. L'architecture des systèmes considérés amène souvent à relaxer encore ces hypothèses. En effet, une file d'attente modélise souvent le trafic dans un nœud intégré au sein d'un réseau et il est souhaitable que les caractéristiques probabilistes d'une file soient les mêmes que celles de la file suivante, autrement dit que le trafic d'entrée d'une file soit de même nature que le trafic de sortie. Or, alors même que le trafic de sortie est poissonnien quand le trafic d'entrée l'est aussi (théorème 8.8), il est facile de se convaincre que les temps interarrivées dans la deuxième file (qui sont les temps entre les sorties de la première file) ne sont pas indépendants en général, même si les interarrivées dans la première file l'étaient pour leur part, puisque leur ordre, par exemple, dépend de l'ordre de service dans la première file.

Il apparaît donc crucial de pouvoir considérer des modèles de files d'attente sous des hypothèses générales de stationnarité et non d'indépendance. Dans ce contexte, on comprendra aisément que des résultats quantitatifs fins sur les modèles puissent être plus difficiles à obtenir. Néanmoins, en se plaçant dans le cadre de travail introduit dans le chapitre 2, on peut dans bien des cas mener à bien l'étude de questions essentielles : existence et unicité d'un état d'équilibre, étude qualitative de cet équilibre

(comparaison de modèles, dépendance des caractéristiques à l'équilibre en la loi des variables aléatoires mises en jeu, etc.).

Dans un premier temps, nous abordons la file classique G/G/1, puis la file à plusieurs serveurs. Les résultats obtenus s'appuient essentiellement sur le théorème de Loynes (théorème 2.4) et donc sur la monotonie de la SRS mise en jeu. Nous considérons ensuite plusieurs modèles de files admettant des représentation dans des espaces plus complexes (comme la file *processor sharing* ou la file à une infinité de serveurs) ou dont la représentation naturelle par une SRS n'est pas monotone, comme les files à rejet ou les files avec clients impatientes.

4.1. Files à un serveur

4.1.1. Résultats généraux

On considère une file d'attente à un seul serveur travaillant sans vacances, traitant suivant une discipline de service non anticipative et conservative des requêtes se présentant suivant un flux d'arrivée stationnaire. Plus précisément, on se donne deux suites aléatoires $(\xi_n, n \in \mathbf{Z})$ et $(\sigma_n, n \in \mathbf{Z})$, respectivement à valeurs dans \mathbf{R}_*^+ et dans \mathbf{R}^+ , et qui représenteront respectivement les temps entre les arrivées des clients et leurs temps de service, comptés en unité de temps.

HYPOTHÈSE.— On suppose que $((\xi_n, \sigma_n), n \in \mathbf{Z})$ est stationnaire et ergodique. De plus, $\mathbf{E}[\xi_0] + \mathbf{E}[\sigma_0] < \infty$.

Comme on l'a vu dans le chapitre 2, on peut supposer que l'espace de probabilité canonique est $\Omega = F^{\mathbf{Z}}$ où $F = \mathbf{R}_*^+ \times \mathbf{R}^+$ muni de la tribu produit \mathcal{F} . La probabilité de référence \mathbf{P} est la mesure image de la suite des couples et l'automorphisme θ est l'opérateur de décalage vers la droite. Le quadruplet $(\Omega, \mathcal{F}, \mathbf{P}, \theta)$ est appelé *espace de Palm des arrivées et des services*.

Les variables aléatoires σ et ξ sont définies sur Ω par :

$$\sigma(((\xi_n, \sigma_n), n \in \mathbf{Z})) = \sigma_0 \text{ et } \xi(((\xi_n, \sigma_n), n \in \mathbf{Z})) = \xi_0,$$

et le décalage θ opère sur les deux composantes simultanément de sorte que :

$$\xi_n = \xi \circ \theta^n \text{ et } \sigma_n = \sigma \circ \theta^n.$$

Pour tout $n \in \mathbf{N}$, $\xi \circ \theta^n$ est donc interprété comme le temps entre l'arrivée du n^{e} client C_n et du $n + 1^{\text{e}}$ client C_{n+1} , et $\sigma \circ \theta^n$ comme le temps de service demandé par C_n . On fixe l'origine des temps $T_0 = 0$ à l'arrivée de C_0 , et on note pour tout $n \geq 1$, $T_n = \sum_{i=0}^{n-1} \xi \circ \theta^i$, qui est donc interprété comme dans le reste de l'ouvrage,

comme l'instant d'arrivée de C_n . Pour tout $n \in \mathbf{N}$, notons W_n la quantité de travail (*workload* en anglais, comptée en unité de temps) que le serveur a à accomplir juste avant l'arrivée du client C_n (c'est-à-dire à T_n^-), en partant d'une quantité de travail arbitraire W_0 à l'arrivée de C_0 . La quantité W_n correspond à la somme des temps de service des clients éventuellement en attente, plus le temps de service restant à traiter pour le client éventuellement en service à cet instant.

Dans le cas particulier où la discipline de service est FCFS (*First Come, First Served*), W_n représente donc la quantité de travail prioritaire sur celle apportée par C_n à son entrée, soit son temps d'attente proposé, avant de pouvoir atteindre le serveur.

LEMME 4.1.– Pour une file G/G/1, la suite des quantités de travail ($W_n, n \in \mathbf{N}$) obéit à l'équation de récurrence suivante (dite équation de Lindley) :

$$W_{n+1} = [W_n + \sigma \circ \theta^n - \xi \circ \theta^n]^+. \quad (4.1)$$

Démonstration. Pour tout $n \geq 0$, la quantité de travail W_{n+1} s'écrit comme la somme de la quantité de travail W_n restant à l'entrée de C_n , plus la quantité de travail $\sigma \circ \theta^n$ apportée par C_n , moins la quantité de travail J_n traitée par le serveur entre l'arrivée de C_n et celle de C_{n+1} . Donc :

$$W_{n+1} = W_n + \sigma \circ \theta^n - J_n.$$

Remarquons que juste après l'arrivée de C_n , le serveur a une charge de travail de $W_n + \sigma \circ \theta^n$. Il y a donc deux cas de figure :

– si $W_n + \sigma \circ \theta^n \geq \xi \circ \theta^n$, le serveur est occupé sans interruption entre l'arrivée de C_n et celle de C_{n+1} , et il traite donc une quantité de travail égale à $\xi \circ \theta^n$ entre ces deux instants. Dans ce cas, $J_n = \xi \circ \theta^n$;

– si $W_n + \sigma \circ \theta^n < \xi \circ \theta^n$, le serveur se libère avant l'arrivée de C_{n+1} , celui-ci trouve un système vide et est servi à son arrivée. En particulier $J_n = W_n + \sigma \circ \theta^n$.

Dans tous les cas, on obtient (4.1). □

On sait depuis l'exemple 2.3 qu'il existe une unique variable aléatoire W susceptible d'être la limite en loi des W_n et que cette variable aléatoire W satisfait l'équation :

$$W \circ \theta = [W + \sigma - \xi]^+ = \varphi(W), \mathbf{P} \text{ – presque sûrement.} \quad (4.2)$$

Intéressons-nous aux conditions sous lesquelles une solution W à valeurs dans \mathbf{R}^+ (et non dans $\mathbf{R}^+ \cup \{\infty\}$) existe. A ces conditions, une charge de travail stationnaire finie existe, dont on peut dire qu'elle décrit l'état stable du système.

THÉORÈME 4.2.– Si :

$$\mathbf{E}[\sigma] < \mathbf{E}[\xi], \quad (4.3)$$

la variable aléatoire W définie par (2.8) est l'unique solution presque sûrement finie de (4.2). Si :

$$\mathbf{E}[\sigma] > \mathbf{E}[\xi],$$

il n'y a pas de solution presque sûrement finie à (4.2).

Preuve du théorème 4.2. Nous pouvons former explicitement W . Rappelons (voir la preuve du théorème 2.4) que presque sûrement $M_0 = 0$, puis :

$$\begin{aligned} M_1(\omega) &= \varphi(M_0, \theta^{-1}\omega) \\ &= [\sigma(\theta^{-1}\omega) - \xi(\theta^{-1}\omega)]^+; \\ M_2(\omega) &= \varphi(M_1(\theta^{-1}\omega), \theta^{-1}\omega) \\ &= \left[[\sigma \circ \theta^{-1}(\theta^{-1}\omega) - \xi \circ \theta^{-1}(\theta^{-1}\omega)]^+ + \sigma(\theta^{-1}\omega) - \xi(\theta^{-1}\omega) \right]^+ \\ &= \left[\max_{1 \leq k \leq 2} \sum_{i=1}^k (\sigma(\theta^{-i}\omega) - \xi(\theta^{-i}\omega)) \right]^+ \end{aligned}$$

et par récurrence, on montre que pour tout $n \in \mathbf{N}^*$:

$$M_n = \left[\max_{1 \leq k \leq n} \sum_{i=1}^k (\sigma \circ \theta^{-i} - \xi \circ \theta^{-i}) \right]^+.$$

Ainsi, la solution minimale de (4.2) est donnée presque sûrement par :

$$W = M_\infty = \left[\sup_{k \geq 1} \sum_{i=1}^k (\sigma \circ \theta^{-i} - \xi \circ \theta^{-i}) \right]^+. \quad (4.4)$$

La variable aléatoire $\sigma - \xi$ étant intégrable, on a d'après le théorème ergodique 2.7 :

$$\frac{1}{n} \sum_{i=1}^n (\sigma \circ \theta^{-i} - \xi \circ \theta^{-i}) \xrightarrow{n \rightarrow \infty} \mathbf{E}[\sigma - \xi], \text{ presque sûrement.} \quad (4.5)$$

Notons $S_n = \sum_{i=1}^n (\sigma \circ \theta^{-i} - \xi \circ \theta^{-i})$ pour tout n .

Supposons que $\mathbf{E}[\sigma - \xi] > 0$. Dans ce cas, d'après (4.5), S_n tend vers $+\infty$ presque sûrement. D'après (4.4), $W = \limsup_n S_n^+$ donc $\mathbf{P}(W = +\infty) = 1$. Il n'existe donc pas de solution finie à (4.2) en vertu de la minimalité de W .

Supposons que $\mathbf{E}[\sigma - \xi] < 0$. D'après (4.5), la suite $(S, n \in \mathbf{N})$ tend vers $-\infty$ donc $\mathbf{P}(\limsup_n S_n^+ < +\infty) = 1$, soit $\mathbf{P}(W < +\infty) = 1$.

Il reste à vérifier que W est bien l'unique solution presque sûrement finie de (4.2). Remarquons d'abord que :

$$\mathbf{P}(Y = 0) > 0 \text{ pour toute solution finie } Y \text{ de (4.2).} \quad (4.6)$$

En effet, si $Y > 0$ presque sûrement alors $Y \circ \theta = Y + \sigma - \xi$ presque sûrement, ce qui implique que $\mathbf{E}[Y \circ \theta - Y] = \mathbf{E}[\sigma - \xi] < 0$, en contradiction avec le lemme 2.2. Or :

$$\{Y = 0\} \subset \{Y \leq W\},$$

ce qui implique d'après (4.6) que :

$$\mathbf{P}(Y \leq W) > 0. \quad (4.7)$$

Mais sur $\{Y \leq W\}$:

$$Y \circ \theta = \varphi(Y) \leq \varphi(W) = W \circ \theta,$$

par monotonie. L'événement $\{Y \leq W\}$ est donc θ -contractant, il est par conséquent presque-sûr d'après (4.7), soit $Y \leq W$ presque sûrement. Par minimalité de W , il s'ensuit que $Y = W$ presque sûrement. \square

Le cas où $\mathbf{E}[\sigma] = \mathbf{E}[\xi]$ est un cas limite où une charge de travail stationnaire finie peut exister ou ne pas exister suivant les lois de σ et ξ . Pour nous en convaincre, considérons deux exemples très simples.

EXEMPLE 4.1.– Supposons que $\sigma_n = \xi_n = 1$ pour tout n , ce qui revient à $\sigma = \xi = 1$, \mathbf{P} -presque sûrement. Alors, il est immédiat de voir que pour tout $x \geq 0$, la variable aléatoire $W \equiv x$ est une solution de (4.2).

EXEMPLE 4.2.– Supposons maintenant que $(\sigma_n, n \in \mathbf{N})$ et $(\xi_n, n \in \mathbf{N})$ sont deux suites indépendantes de variables aléatoires indépendantes et identiquement distribuées de même moyenne et de variances respectives σ_1^2 et σ_2^2 . Soit $0 < \epsilon < \frac{1}{2}$ et $a \in \mathbf{R}_+^*$ tel que $1 - F(a) = \epsilon$, où F désigne la fonction de répartition de $\mathcal{N}(0, 1)$. Alors, pour tout $x > 0$, il existe un indice n suffisamment grand tel que $\sqrt{n}\sqrt{\sigma_1^2 + \sigma_2^2}a \geq x$, et donc :

$$\mathbf{P}(W > x) \geq \mathbf{P}(S_n > x) \geq \mathbf{P}\left(\frac{S_n}{\sqrt{n}\sqrt{\sigma_1^2 + \sigma_2^2}} > a\right) \xrightarrow[n \rightarrow \infty]{} \epsilon$$

d'après le TLC. Ceci montre que $\mathbf{P}(W = \infty) \geq \epsilon$, et donc $W = +\infty$ presque sûrement, puisque l'événement $\{W = \infty\}$ est θ -contractant.

Nous supposons dorénavant que la condition de stabilité (4.3) est vérifiée.

THÉORÈME 4.3.– Pour toute variable aléatoire Y , \mathbf{P} -presque sûrement finie et positive, la suite des charges de travail $(W_n^Y, n \in \mathbf{N})$ issue de Y couple avec $(W \circ \theta^n, n \in \mathbf{N})$. En particulier, $W_n^Y \xrightarrow[n \rightarrow \infty]{\mathcal{L}} W$. D'autre part, il y a couplage arrière fort si $Y \leq W$, presque sûrement.

Démonstration. Soit ψ l'application définie par :

$$\psi : \begin{cases} \mathbf{R}^+ \times \Omega & \longrightarrow \mathbf{R} \\ (x, \omega) & \longmapsto x + (\sigma(\omega) - \xi(\omega)) \end{cases}$$

et $(Z_n, n \in \mathbf{N})$ la SRS portée par ψ et issue de Y . Le théorème 2.7 montre que sur un événement \mathfrak{A} de probabilité 1 :

$$\frac{1}{n} Z_n = \frac{Y}{n} + \frac{1}{n} \sum_{i=1}^n (\sigma \circ \theta^i - \xi \circ \theta^i) \xrightarrow[n \rightarrow \infty]{} \mathbf{E}[\sigma - \xi] < 0.$$

Sur \mathfrak{A} , Z_n tend vers $-\infty$ donc :

$$N^Y = \inf\{n \geq 0, Z_n < 0\}$$

est presque sûrement fini. On remarque que $W_n^Y = Z_n > 0$ pour tout $n \leq N^Y$ et $W_{N^Y}^Y = 0$. Par ailleurs, comme $Y \geq 0$, une récurrence immédiate montre que $W_n^Y \geq W_n^0$ presque sûrement pour tout n . Ceci implique que $W_{N^Y}^0 = 0 = W_{N^Y}^Y$ d'où il s'ensuit par définition d'une SRS que $W_n^0 = W_n^Y$ pour tout $n \geq N^Y$. Pour une autre condition initiale, par exemple $Y = W$, il existe $N^W < \infty$ tel que que $W_n^W = W_n^0$ pour tout $n \geq N^W$. Pour $n \geq N^Y \vee N^W$, $W_n^Y = W_n^0 = W_n^W$, ce qui montre la propriété de couplage entre $(W_n^Y, n \in \mathbf{N})$ et $(W \circ \theta^n, n \in \mathbf{N})$.

Soit maintenant Y une variable aléatoire presque sûrement majorée par W . Par monotonie, on a clairement $W_n^Y \leq W_n^W = W \circ \theta^n$ presque sûrement pour tout $n \in \mathbf{N}$. Sur $\theta^{-n}\{W = 0\}$, on a donc $W_n^Y = 0$, ce qui montre que $(\theta^{-n}\{W = 0\}, n \in \mathbf{N})$ est une suite d'événements de renouvellement de longueur 1 pour $(W_n^Y, n \in \mathbf{N})$. Le théorème 2.11 permet de conclure à la propriété de couplage arrière fort. \square

Nous nous intéressons maintenant à la faculté de la file d'attente à pouvoir se vider une infinité de fois presque sûrement. Il est clair que si la file se vide entre l'arrivée de C_{n-1} et celle de C_n , on a $W_n = 0$. Pour parler le langage des chaînes de Markov, la propriété de se vider une infinité de fois presque sûrement correspond donc à la récurrence du point 0 pour la suite $(W_n, n \in \mathbf{N})$. Nous montrons que cette propriété est vérifiée à la condition (4.3).

COROLLAIRE 4.4.– Sous la condition de stabilité $\mathbf{E}[\sigma] < \mathbf{E}[\xi]$, la file G/G/1 issue de toute charge finie Y à l'arrivée de C_0 se vide \mathbf{P} -presque sûrement une infinité de fois.

Démonstration. D'après le théorème 2.7 et (4.6), on a \mathbf{P} -presque sûrement :

$$\begin{aligned} 0 < \mathbf{P}(W = 0) &= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \mathbf{1}_0(W \circ \theta^i) \\ &= \lim_{n \rightarrow \infty} \frac{1}{n} \left\{ \sum_{i=1}^{N-1} \mathbf{1}_0(W \circ \theta^i) + \sum_{i=N}^n \mathbf{1}_0(W_i^Y) \right\}. \end{aligned}$$

Autrement dit, $\sum_{i=N}^{+\infty} \mathbf{1}_0(W_i^Y) = +\infty$, \mathbf{P} -presque sûrement, ce qui signifie très exactement que la file se vide une infinité de fois. \square

4.1.2. Comparaisons de files G/G/1

THÉORÈME 4.5.– Soit deux files G/G/1, portées respectivement par les variables aléatoires (σ, ξ) et $(\bar{\sigma}, \bar{\xi})$ respectivement définies sur $(\Omega, \mathcal{F}, \mathbf{P}, \theta)$ et $(\bar{\Omega}, \bar{\mathcal{F}}, \bar{\mathbf{P}}, \bar{\theta})$. Si l'on a :

$$\bar{\sigma} - \bar{\xi} \leq_{st} \sigma - \xi, \quad (4.8)$$

alors les solutions respectives W et \bar{W} des équations :

$$W \circ \theta = [W + \sigma - \xi]^+, \quad \mathbf{P} \text{ - presque sûrement,}$$

$$\bar{W} \circ \bar{\theta} = [\bar{W} + \bar{\sigma} - \bar{\xi}]^+, \quad \bar{\mathbf{P}} \text{ - presque sûrement}$$

vérifient :

$$\bar{W} \leq_{st} W. \quad (4.9)$$

Démonstration. On applique le théorème 2.15 à $\bar{\alpha} = \bar{\sigma} - \bar{\xi}$ et $\alpha = \sigma - \xi$ et $f(x, y) = (x + y)^+$. \square

REMARQUE.– En particulier, si l'on suppose que $(\sigma \circ \theta^n, n \in \mathbf{N})$ (respectivement $(\bar{\sigma} \circ \bar{\theta}^n, n \in \mathbf{N})$) est indépendante de $(\xi \circ \theta^n, n \in \mathbf{N})$ (respectivement $(\bar{\xi} \circ \bar{\theta}^n, n \in \mathbf{N})$), il est facile de voir que l'une des deux conditions :

$$\xi \stackrel{\mathcal{L}}{=} \bar{\xi} \text{ et } \bar{\sigma} \leq_{st} \sigma,$$

$$\sigma \stackrel{\mathcal{L}}{=} \bar{\sigma} \text{ et } \xi \leq_{st} \bar{\xi}$$

implique (4.8) et donc (4.9).

THÉORÈME 4.6.— Soit une file G/G/1 stable portée par les variables aléatoires σ et ξ , où l'on suppose que les suites $(\sigma \circ \theta^n, n \in \mathbf{N})$ et $(\xi \circ \theta^n, n \in \mathbf{N})$ sont indépendantes et satisfont la condition de stabilité $\mathbf{E}[\sigma - \xi] < 0$. Définissons sur le même espace de probabilité, les deux files alternatives suivantes :

– la première est une file G/D/1 de même charge, où le temps de service générique est donné par $\hat{\sigma} = \mathbf{E}[\sigma]$ et d'interarrivée générique ξ . Cette file est également stable et l'on note \hat{W} , sa charge de travail stationnaire ;

– la deuxième est une D/G/1 portée par $\check{\xi} = \mathbf{E}[\xi]$ et σ . On note \check{W} sa charge de travail stationnaire.

Pour toute fonction F croissante convexe, on a :

$$\mathbf{E} \left[F \left(\hat{W} \right) \right] \leq \mathbf{E} \left[F \left(W \right) \right]; \quad (4.10)$$

$$\mathbf{E} \left[F \left(\check{W} \right) \right] \leq \mathbf{E} \left[F \left(W \right) \right]. \quad (4.11)$$

Démonstration. Il est clair qu'à la condition (4.3), les deux systèmes sont stables, et l'on note \hat{W} et \check{W} leurs charges de travail stationnaires respectives. Le théorème 2.17 est vérifié, en posant pour \hat{W} :

$$\alpha = \sigma - \xi, \bar{\alpha} = \mathbf{E}[\sigma] - \xi \text{ et } \mathcal{F}_n = \sigma(\xi \circ \theta^j; j = 0, \dots, n)$$

et pour \check{W} :

$$\alpha = \sigma - \xi, \bar{\alpha} = \sigma - \mathbf{E}[\xi] \text{ et } \mathcal{F}_n = \sigma(\sigma \circ \theta^i; i = 0, \dots, n).$$

En effet, par exemple pour le premier cas, par indépendance on a presque sûrement pour tout $n \in \mathbf{N}$ et tout $i \leq n$:

$$\begin{aligned} \mathbf{E}[\alpha \circ \theta^i | \mathcal{F}_n] &= \mathbf{E}[\sigma \circ \theta^i - \xi \circ \theta^i | \xi \circ \theta^j; j = 0, \dots, n] \\ &= \mathbf{E}[\sigma] - \xi \circ \theta^i. \\ &= \bar{\alpha} \circ \theta^i. \end{aligned}$$

L'autre cas se traite de manière analogue. On conclut d'après le corollaire 2.18. \square

⚡ En ce sens, le déterminisme minimise à l'équilibre la charge de travail moyenne et donc le temps d'attente moyen si la discipline est FIFO. Si l'on admet que dans un routeur, le temps de traitement des paquets est proportionnel à leur longueur, cela signifie que le délai moyen est minimisé lorsque l'on prend des paquets de longueur constante.

A la limite, remarquons que le système déterministe, d'interarrivées et de temps de service égaux à leurs moyennes respectives, admet bien sûr l'unique solution stationnaire $W = 0$ presque sûrement.

4.1.3. Représentation des disciplines de service

Dans une file d'attente, la discipline de service caractérise la politique suivie par le(s) serveur(s) pour choisir un client lorsqu'il(s) se libère(nt) et qu'il y a plusieurs clients en attente. Notons que tous les résultats obtenus jusqu'ici dans la section 4.1 sont indépendants de la discipline considérée. Ainsi, afin de représenter cette information dans l'état du système, il convient d'enrichir le modèle, puisque la charge de travail seule ne suffit pas à caractériser la discipline de service.

Pour rendre compte de la discipline de service, nous représentons le système dans un espace plus grand. Concrètement, on décrit le système juste avant l'arrivée du client C_n , $n \geq 0$, par une suite ordonnée S_n , représentant les temps de service résiduels des clients présents dans le système à cet instant. Plus précisément, on note X_n le nombre de clients dans le système à T_n^- , et pour tout $i < n$ tel que C_i est dans le système à T_n^- , on note $\varphi_n(i) \in \{1, \dots, X_n\}$, la place de C_i dans la file d'attente dans l'ordre des priorités, la première étant celle occupée par le client en service à T_n^- . Pour chaque tel client C_i , on note $R_n(\varphi_n(i))$ le temps de service résiduel de C_i à T_n^- , c'est-à-dire :

$$R_n(\varphi_n(i)) = \begin{cases} \sigma \circ \theta^i & \text{si } C_i \text{ n'a pas reçu de service avant } T_n^-, \\ \sigma \circ \theta^i - \gamma_i & \text{si } C_i \text{ a déjà reçu la quantité de service } \gamma_i < \sigma \circ \theta^i \text{ à } T_n^-. \end{cases}$$

Pour tout $n \in \mathbf{Z}$, on définit $S_n \in \mathcal{S}$ (voir l'appendice A.3) la suite représentant les temps de service résiduels des clients présents dans le système au même instant, rangés dans l'ordre inverse des priorités, et en fixant à 0 les autres composantes de S_n , c'est-à-dire :

$$S_n(i) = R_n(X_n + 1 - i), \quad i \leq X_n \text{ et } S_n(i) = 0, \quad i > X_n,$$

ou encore :

$$S_n = \left\{ R_n(X_n), R_n(X_n - 1), \dots, R_n(2), R_n(1), 0, 0, \dots \right\}.$$

On appelle S_n , le *profil* des services résiduels à T_n^- .

Nous détaillons maintenant la dynamique de la suite de suites $(S_n, n \in \mathbf{N})$, en fonction de la discipline de service. On part d'un profil arbitraire $S_0 \in \mathcal{S}$ à l'arrivée de C_0 . Soit S_n , la valeur du profil juste avant l'arrivée de C_n . A T_n , on insère dans le profil le temps de service $\sigma \circ \theta^n$ du client entrant, arbitrairement à la première place, et l'on décale les autres termes de la suite d'un cran à droite. En appelant S_{n+} la suite résultante, on a donc :

$$S_{n+} = \left\{ \sigma \circ \theta^n, S_n(1), S_n(2), \dots \right\} = F^1(S_n, \sigma \circ \theta^n). \quad (4.12)$$

Ensuite, la discipline de service Φ est représentée par une application $F^\Phi : \mathcal{S} \rightarrow \mathcal{S}$ de la manière suivante :

- FCFS, F^{FCFS} est l'identité puisque le client entrant a la priorité la plus basse ;
- LCFS non préemptif :

$$F^{\text{LCFS}}(u) = \left\{ u(2), u(3), \dots, u(N(u) - 1), u(1), u(N(u)), 0, \dots \right\},$$

puisque le client entrant est inséré juste derrière le client en service dans l'ordre des priorités ;

- LCFS préemptif, dans ce cas :

$$F^{\text{LCFS}}(u) = \left\{ u(2), u(3), \dots, u(N_u - 1), u(N_u), u(0), 0, \dots \right\},$$

puisque le client entrant prend immédiatement la place du client en service (s'il y en a un) ;

- SRPT (*Shortest Remaining Processing Time*), on donne priorité préemptive au client qui a le temps de service résiduel le plus petit. Donc :

$$F^{\text{SRPT}}(u) = \begin{cases} \left\{ u(1), u(2), \dots, u(i-1), u(0), u(i), \dots, u(N(u)), 0, \dots \right\} \\ \hspace{15em} \text{si } u(i-1) \geq u(0) \geq u(i), \\ \left\{ u(0), u(1), u(2), \dots, u(N(u)), 0, \dots \right\} \text{ si } u(0) \geq u(1), \\ \left\{ u(1), u(2), \dots, u(N(u)), u(0), 0, \dots \right\} \text{ si } u(0) < u(N(u)). \end{cases}$$

Il vient ainsi que $F^{\text{SRPT}}(u)$ est décroissante dès que u l'est ;

- SPT (*Shortest Processing Time*) : on donne priorité non-préemptive au client qui a le temps de service résiduel le plus petit. Donc F^{SPT} égale F^{SRPT} à ceci près que $u(0)$ est inséré juste avant $u(N(u))$, même si $u(0) < u(N(u))$.

On voit bien que l'on peut ainsi représenter par une telle permutation des coordonnées non nulles de tout élément de \mathcal{S} , toute discipline de service ne dépendant que des dates d'arrivée et des demandes de service des clients depuis la dernière entrée dans un système vide, ou au plus depuis l'instant T_0- . Une telle discipline sera dite admissible.

On notera alors S_{n++} , le profil de la file juste après l'ordonnancement des clients, de sorte que :

$$S_{n++} = F^{\Phi}(S_{n+}). \quad (4.13)$$

Ensuite, le client en service à T_n (de service résiduel $S_{n+}(X_n)$ à cet instant) a juste avant l'arrivée de C_{n+1} un temps de service résiduel :

$$S_{n+1}(X_n) = [S_{n+1}(X_n) - \xi \circ \theta^n]^+.$$

Tout client le suivant (de service résiduel donné à T_n par $S_{n+}(j)$, $j = 0, \dots, X_n - 1$) reçoit un service avant T_{n+1} si et seulement si :

$$\xi \circ \theta^n > \sum_{i=j+1}^{+\infty} S_n(i),$$

en quantité donnée par :

$$\left(\xi \circ \theta^n - \sum_{i=j+1}^{+\infty} S_n(i) \right) \wedge S_{n+}(j).$$

Autrement dit, pour tout $j \in \mathbf{N}$:

$$\begin{aligned} S_{n+1}(j) &= S_{n+}(j) - \left(\xi \circ \theta^n - \sum_{i=j+1}^{+\infty} S_n(i) \right)^+ \wedge S_{n+}(j) \\ &= \left(S_{n+}(j) - \left(\xi \circ \theta^n - \sum_{i=j+1}^{+\infty} S_n(i) \right)^+ \right)^+. \end{aligned}$$

On notera alors $F^3(\cdot, \xi \circ \theta^n)$ l'application correspondante, de sorte que :

$$S_{n+1} = F^3(S_{n++}, \xi \circ \theta^n). \quad (4.14)$$

Les équations (4.12) à (4.14) indiquent que pour Φ fixée, la suite $(S_n, n \in \mathbf{N})$ est récurrente puisque pour tout $n \in \mathbf{N}$:

$$S_{n+1} = F^3(\cdot, \xi \circ \theta^n) \circ F^\Phi \circ F^1(\cdot, \sigma \circ \theta^n)(S_n). \quad (4.15)$$

Une suite de profils stationnaire correspond donc de manière univoque à une variable aléatoire S^Φ à valeurs dans \mathcal{S} , solution de l'équation :

$$S^\Phi \circ \theta = G^\Phi(S^\Phi), \quad (4.16)$$

où G^Φ est l'application définie par :

$$G^\Phi : \begin{cases} \mathcal{S} & \longrightarrow \mathcal{S} \\ u & \longmapsto F^3(u, \xi) \circ F^\Phi \circ F^1(u, \sigma). \end{cases}$$

THÉORÈME 4.7.— Soit Φ , une discipline de service admissible. Si la condition de stabilité (4.3) est vérifiée, il existe une unique solution S^Φ à (4.16) telle que $S^\Phi \in \mathcal{S}$ presque sûrement. De plus, il y a couplage arrière fort entre les suites $(S_n^{\Phi, \mu}, n \in \mathbf{N})$ et $(S^\Phi \circ \theta^n, n \in \mathbf{N})$, pour toute $\mu \in \mathcal{S}$ telle que :

$$Z(\mu) = \sum_{i \in \mathbf{N}^*} \mu(i) \leq W \text{ presque sûrement,} \quad (4.17)$$

où W est l'unique solution presque sûrement finie de (4.2).

Démonstration. Remarquons que pour tout $n \in \mathbf{N}$, la charge de travail à T_n^- (partant d'une condition initiale donnée) se déduit de S_n par :

$$W_n = \sum_{i \in \mathbf{N}} S_n(i).$$

Soit $\mu \in \mathcal{S}$ satisfaisant (4.17) et $S_n^{\Phi, \mu}$, la représentation du profil de la file d'attente à T_n^- sous la discipline Φ , en partant du profil μ à T_{0-} . Alors, on a clairement :

$$\sum_{i \in \mathbf{N}} S_n^{\Phi, \mu}(i) = W_n^{Z(\mu)}, \text{ presque sûrement,}$$

où $W_n^{Z(\mu)}$ est la n^e valeur de la SRS portée par φ définie par (4.2) et issue de $Z(\mu)$.

Par ailleurs, en vertu de la croissance presque sûrement de φ , il est facile de montrer par récurrence à partir de (4.17) que presque sûrement :

$$W_n^{Z(\mu)} \leq W \circ \theta^n, n \in \mathbf{N}.$$

Par conséquent, sur l'événement $\mathfrak{A}_n = \{W \circ \theta^n = 0\}$, on a $W_n^{Z(\mu)} = 0$ et donc $S_n^{\Phi, \mu} = \mathbf{0}$, la suite nulle de \mathcal{S} . Ainsi, $(\mathfrak{A}_n, n \in \mathbf{N}) = (\{W = 0\} \circ \theta^{-n}, n \in \mathbf{N})$ est une suite stationnaire d'événements de renouvellement de longueur 1 pour la suite $(S_n^{\Phi, \mu}, n \in \mathbf{N})$. Comme $\mathbf{P}(W = 0) > 0$ d'après (4.6), le théorème 2.11 appliqué à la classe de conditions initiales :

$$\mathcal{Z} = \left\{ \mu \in \mathcal{S}; \mu \text{ vérifie (4.17)} \right\} \quad (4.18)$$

implique l'existence d'une solution S^Φ à l'équation (4.16) pour la discipline Φ . Cette solution (qui s'écrit comme la limite presque sûrement d'une suite constante à partir d'un certain rang (voir la preuve du théorème 2.11) est presque sûrement finie dans le sens où ses composantes sont presque sûrement finies et qu'elle admet un nombre presque sûrement fini de composantes non nulles. En d'autres termes, S^Φ appartient à \mathcal{S} presque sûrement.

Soient S et S' , deux solutions de (4.16) à valeurs dans \mathcal{S} . En notant :

$$Z = \sum_{i \in \mathbf{N}} S(i) \text{ et } Z' = \sum_{i \in \mathbf{N}} S'(i)$$

les charges de travail respectives correspondant à ces deux profils, il est facile de vérifier avec (4.12) et (4.13) que Z et Z' sont deux solutions presque sûrement finies de (4.2). D'après le théorème 4.2, on a alors $Z = Z' = W$. Donc :

$$\{W = 0\} \subset \{Z = Z' = 0\} \subset \{S = S' = \mathbf{0}\}.$$

Comme l'événement de gauche est de probabilité non nulle et celui de droite est θ -invariant, on a $S = S'$ presque sûrement.

Le théorème 2.11 appliqué à la classe \mathcal{Z} implique en particulier la propriété de couplage arrière fort de $(S_n^\Phi, \mu, n \in \mathbf{N})$ avec $(S^\Phi \circ \theta^n, n \in \mathbf{N})$. \square

4.1.4. Autres caractéristiques à l'équilibre

Pour une discipline admissible donnée Φ , le vecteur S^Φ donne donc une information plus riche sur l'état d'équilibre du système que la seule charge de travail W . Montrons comment cette information peut être exploitée pour en déduire d'autres caractéristiques stationnaires du système, comme la congestion ou le temps d'attente.

4.1.4.1. Congestion

Notons $(X_n^\Phi, n \in \mathbf{N})$ la suite comptant pour tout $n \in \mathbf{N}$ le nombre de clients trouvés dans le système par C_n , partant d'un profil initial S^Φ . Dans ces conditions, la suite des profils trouvés par les clients successifs à leur arrivée est stationnaire et vaut $(S^\Phi \circ \theta^n, n \in \mathbf{N})$. En particulier, le client C_n trouve dans le système à son arrivée un profil $S^\Phi \circ \theta^n$, et donc :

$$X_n^\Phi = N(S^\Phi \circ \theta^n) = N(S^\Phi) \circ \theta^n \text{ presque sûrement pour tout } n \in \mathbf{N},$$

où $N(\cdot)$ est le nombre de coordonnées non nulles de la suite (voir A.3). Cela signifie que $(X_n^\Phi, n \in \mathbf{N})$ est stationnaire et qu'une congestion du système à l'équilibre existe, donnée par $X^\Phi = N(S^\Phi)$.

4.1.4.2. Temps d'attente et temps de séjour

On peut appliquer le même raisonnement pour montrer l'existence d'un temps d'attente à l'équilibre grâce au profil de service. Notons en général TA_n^Φ , le temps d'attente du client C_n pour entrer en service sous la discipline admissible Φ (rappelons que pour $\Phi = \text{FIFO}$, $\text{TA}_n^\Phi = W_n$).

Partons là encore d'un profil de service S^Φ . Le profil du système à l'arrivée du client C_n égale $S^\Phi \circ \theta^n$ et devient $(S^\Phi \circ \theta^n)_{++}$ après l'insertion de $\sigma \circ \theta^n$ (voir (4.13)).

Si i est le rang de $\sigma \circ \theta^n$ dans la suite $(S^\Phi \circ \theta^n)_{++}$, le temps d'attente de C_n égale donc la somme $\sum_{j>i} (S^\Phi \circ \theta^n)_{++}(j)$ des temps de service des clients prioritaires sur C_n , plus ceux des clients arrivés après l'entrée de C_n et avant son départ du système (ou son entrée en service si la Φ est non préemptive).

Cette quantité peut prendre une forme très complexe suivant la discipline Φ , mais l'essentiel est de se convaincre qu'elle ne dépend que de S_n^Φ , de $\sigma \circ \theta^n$ et des temps de service $\{\sigma \circ \theta^j; j > n\}$ et temps interarrivées $\{\sigma \circ \theta^j; j > n\}$ des clients arrivés après C_n . Autrement dit, il existe une fonction déterministe $J^\Phi : (\mathbf{R}^+)^{\mathbf{N}} \times (\mathbf{R}^+)^{\mathbf{N}} \rightarrow \mathbf{R}^+$ ne dépendant que de Φ telle que presque sûrement pour tout $n \in \mathbf{N}$:

$$\begin{aligned} \text{TA}_n^\Phi &= \sum_{j>i} (S^\Phi \circ \theta^n)_{++}(j) \\ &\quad + J^\Phi \left(\left\{ \sigma \circ \theta^{n+1}, \sigma \circ \theta^{n+2}, \dots \right\}, \left\{ \xi \circ \theta^{n+1}, \xi \circ \theta^{n+2}, \dots \right\} \right) \\ &= \left(\sum_{j>i} \left(F^\Phi \circ F^1(\cdot, \sigma) (S^\Phi) \right) (j) \right. \\ &\quad \left. + J^\Phi \left(\left\{ \sigma \circ \theta, \sigma \circ \theta^2, \dots \right\}, \left\{ \xi \circ \theta, \xi \circ \theta^2, \dots \right\} \right) \right) \circ \theta^n, \end{aligned}$$

où l'on a utilisé (4.15). Ceci montre donc, là encore, l'existence d'un temps d'attente stationnaire :

$$\begin{aligned} \text{TA}^\Phi &= \sum_{j>i} \left(F^\Phi \circ F^1(\cdot, \sigma) (S^\Phi) \right) (j) \\ &\quad + J^\Phi \left(\left\{ \sigma \circ \theta, \sigma \circ \theta^2, \dots \right\}, \left\{ \xi \circ \theta, \xi \circ \theta^2, \dots \right\} \right). \quad (4.19) \end{aligned}$$

EXEMPLE 4.3.– Ecrivons TA^Φ , par exemple, pour $\Phi = \text{LIFO}$ non préemptive. Tout d'abord, comme le service du client entrant C_0 est placé directement en priorité juste derrière le client en service, la somme de gauche de (4.19) égale le temps de service résiduel du client en service à l'arrivée de C_0 , donné par le dernier terme non nul de S^{LIFO} , soit $S^{\text{LIFO}} \left(N(S^{\text{LIFO}}) \right)$.

Ensuite, le terme de droite égale la somme des temps de service des clients entrés avant que le client C_0 ne puisse atteindre le serveur. Autrement dit :

$$J^{\text{LIFO}}(\dots) = \begin{cases} \sum_{i=1}^{i_0-1} \sigma \circ \theta^i & \text{si } i_0 > 1, \\ 0 & \text{sinon,} \end{cases}$$

où

$$i_0 = \inf \left\{ j \geq 1; \tilde{S}^{\text{LIFO}} \left(N \left(\tilde{S}^{\text{LIFO}} \right) \right) + \sum_{j=1}^{i-1} \sigma \circ \theta^j - \sum_{k=1}^i \xi \circ \theta^k \leq 0 \right\}$$

est le premier indice d'un client entré après C_0 qui termine son service avant que le client suivant ne soit entré.

REMARQUE.— A la condition de stabilité (4.3), pour toute discipline admissible Φ il existe donc également un temps de séjour stationnaire TS^Φ dans le système, donné par :

$$\text{TS}^\Phi = \text{TA}^\Phi + \sigma. \quad (4.20)$$

4.1.5. Optimalité de SRPT

Forts de la représentation exhaustive introduite dans la section précédente, nous montrons maintenant comment comparer à l'équilibre, les différentes disciplines de service. On travaille ici avec l'ordre Schur-convexe, noté « $<_c$ », présenté en A.3.

THÉORÈME 4.8.— Soit une file G/G/1 satisfaisant la condition de stabilité (4.3). Soit Φ une discipline admissible quelconque. Alors :

$$S^\Phi <_c S^{\text{SRPT}} \text{ presque sûrement,}$$

où S^Φ et S^{SRPT} sont les uniques solutions de (4.16) pour Φ et SRPT, respectivement.

Démonstration. Soit Φ , une discipline admissible et S^Φ l'unique solution presque sûrement finie de (4.16). Avec les notations précédentes, on a presque sûrement :

$$S^\Phi \xrightarrow{F^1(\cdot, \sigma)} S_+^\Phi \xrightarrow{F^\Phi} S_{++}^\Phi \xrightarrow{F^2(\cdot, \xi)} S^\Phi \circ \theta.$$

Fixons $j \in \mathbf{N}$. Si $\sigma < \underline{S}^\Phi(j-1)$, alors $\sigma \in \{\underline{S}_+^\Phi(i); i \geq j\}$, alors que si $\sigma \geq \underline{S}^\Phi(j-1)$, on a $\underline{S}_+^\Phi(i) = \underline{S}^\Phi(i-1)$ pour tout $i \geq j-1$. On a donc dans tous les cas :

$$\sum_{i=j}^{+\infty} \underline{S}_+^\Phi(i) = \left(\sigma + \sum_{i=j}^{+\infty} \underline{S}^\Phi(i) \right) \wedge \left(\sum_{i=j-1}^{+\infty} \underline{S}^\Phi(i) \right). \quad (4.21)$$

Par ailleurs, les clients de temps de service initialement égaux à $\underline{S}_{++}^\Phi(i)$, $i \geq j$, reçoivent à eux tous dans un intervalle de temps de longueur ξ un temps de service au plus égal à :

$$\left(\sum_{i=j}^{+\infty} \underline{S}_{++}^\Phi(i) \right) \wedge \xi.$$

Par conséquent :

$$\begin{aligned}
\sum_{i=j}^{+\infty} (\underline{S}^\Phi \circ \theta)(i) &\geq \left[\sum_{i=j}^{+\infty} \underline{S}_{++}^\Phi(i) - \xi \right]^+ \\
&= \left[\sum_{i=j}^{+\infty} \underline{S}_+^\Phi(i) - \xi \right]^+ \\
&= \left[\min \left\{ \sigma + \sum_{i=j}^{+\infty} \underline{S}^\Phi(i); \sum_{i=j-1}^{+\infty} \underline{S}^\Phi(i) \right\} - \xi \right]^+, \quad (4.22)
\end{aligned}$$

avec (4.21).

Remarquons maintenant que par définition de SRPT, on a $S^{\text{SRPT}} = \underline{S}_{++}^{\text{SRPT}}$, ce qui avec (4.21) et (4.22) implique que $S^{\text{SRPT}} \circ \theta$ est ordonné et que l'on a :

$$\begin{aligned}
\sum_{i=j}^{+\infty} (\underline{S}^{\text{SRPT}} \circ \theta)(i) &= \sum_{i=j}^{+\infty} (S^{\text{SRPT}} \circ \theta)(i) \\
&= \left[\sum_{i=j}^{+\infty} \min \left\{ \sigma + \sum_{i=j}^{+\infty} \underline{S}^{\text{SRPT}}(i); \sum_{i=j-1}^{+\infty} \underline{S}^{\text{SRPT}}(i) \right\} - \xi \right]^+. \quad (4.23)
\end{aligned}$$

Par conséquent, pour tout discipline admissible Φ , sur $\{S^\Phi < S^{\text{SRPT}}\}$ on a pour tout $j \in \mathbf{N}$:

$$\sum_{i=j}^{+\infty} \underline{S}^\Phi(i) \geq \sum_{i=j}^{+\infty} \underline{S}^{\text{SRPT}}(i).$$

Ceci implique avec (4.22) et (4.23) que :

$$\sum_{i=j}^{+\infty} (\underline{S}^\Phi \circ \theta)(i) \geq \sum_{i=j}^{+\infty} (\underline{S}^{\text{SRPT}} \circ \theta)(i).$$

Comme cette dernière égalité est vraie pour tout $j \in \mathbf{N}$, et que pour $j = 1$:

$$\sum_{i=1}^{+\infty} (\underline{S}^\Phi \circ \theta)(i) = \sum_{i=1}^{+\infty} (\underline{S}^{\text{SRPT}} \circ \theta)(i) = W \circ \theta,$$

on a $\underline{S}^\Phi \circ \theta < \underline{S}^{\text{SRPT}} \circ \theta$. L'événement $\{S^\Phi < S^{\text{SRPT}}\}$ est donc θ -contractant. Il est par ailleurs, de probabilité positive puisqu'il inclut l'événement $\{W = 0\}$. Le théorème est démontré. \square

Les preuves des corollaires suivants sont laissées au lecteur.

COROLLAIRE 4.9.– Soit Φ , une discipline quelconque. Alors :

– $X^{\text{SRPT}} \leq X^\Phi$ presque sûrement, où X^{SRPT} et X^Φ désignent le nombre de clients à l'équilibre dans le système sous SRPT et Φ , respectivement ;

– $\mathbf{E}[T_{\text{S}}^{\text{SRPT}}] \leq \mathbf{E}[T_{\text{S}}^\Phi]$, où $T_{\text{S}}^{\text{SRPT}}$ et T_{S}^Φ désignent le temps de séjour stationnaire, respectivement sous SRPT et Φ (voir (4.20)).

4.1.6. File GI/GI/1 : optimalité de FIFO

Considérons une file d'attente GI/GI/1, où λ , μ et ρ représentent les paramètres habituels. En plus des hypothèses courantes, nous supposons donc que les suites des interarrivées et des temps de service sont indépendantes et identiquement distribuées et indépendantes l'une de l'autre. Nous supposons, de plus, que la condition de stabilité (4.3) est vérifiée. Nous notons à nouveau pour tout n T_{A_n} , le temps d'attente de C_n avant d'atteindre le serveur, $T_{\text{S}_n} = T_{A_n} + \sigma \circ \theta^n$ le temps de séjour de C_n et $T'_n = T_n + T_{\text{S}_n}$, l'instant de sortie de C_n . Par la suite, nous soulignons quand cela est nécessaire la dépendance en la discipline de service en ajoutant des exposants FIFO et Ψ aux différents paramètres. On a vu en particulier qu'un temps d'attente stationnaire $T_{\text{A}}^{\text{FIFO}}$ (respectivement T_{A}^Ψ) et un temps de séjour stationnaire $T_{\text{S}}^{\text{FIFO}}$ (respectivement T_{S}^Ψ) existent sous FIFO (respectivement Ψ).

THÉORÈME 4.10.– Pour toute fonction convexe $g : \mathbf{R} \rightarrow \mathbf{R}$ et toute discipline admissible Ψ non préemptive et ne dépendant pas des temps de service :

$$\mathbf{E}[g(T_{\text{S}}^{\text{FIFO}})] \leq \mathbf{E}[g(T_{\text{S}}^\Psi)]. \quad (4.24)$$

☞ La discipline FIFO est optimale pour le temps de séjour parmi toutes les disciplines admissibles non préemptives et indépendantes des temps de service.

Démonstration. Nous couplons deux systèmes « nourris » par les mêmes arrivées et temps de service, l'un traité par FIFO et l'autre par Ψ . Nous supposons que C_0 trouve un système vide à son arrivée. Comme le système est stable, il existe \mathbf{P} -presque sûrement un entier fini τ (commun aux deux systèmes) tel que C_τ entre dans un système vide. Notons pour tout $k \geq 0$, $\psi(k)$ l'indice du k^{e} client servi par Ψ , en considérant bien sûr que C_0 est servi en « 0^e » position (puisque'il est seul dans le système à son arrivée), c'est-à-dire $\psi(0) = 0$.

Définissons presque sûrement les deux vecteurs de taille τ suivants :

$$N = ((\xi_0, \sigma_0), \dots, (\xi_{\tau-1}, \sigma_{\tau-1})) \text{ et } N^\psi = ((\xi_0, \sigma_{\psi(0)}), \dots, (\xi_{\tau-1}, \sigma_{\psi(\tau-1)})), \quad (4.25)$$

qui représentent respectivement, durant la première période d'activité, les arrivées de clients avec leurs temps de service et les arrivées de clients en réarrangeant les temps de service suivant ψ . L'argument d'échange pour les suites de variables aléatoires indépendantes et identiquement distribuées est le résultat intuitif suivant :

$$N \text{ et } N^\psi \text{ ont la même loi.} \quad (4.26)$$

Moralement, comme les temps de service sont indépendants, identiquement distribués et indépendants de tout le reste, on ne modifie pas la loi des différents paramètres du système en échangeant les temps de service des clients. Tout se passe comme si c'était le serveur qui décidait des temps de service des clients en réalisant un tirage indépendant de temps de service à l'arrivée de chaque client. Pour une preuve rigoureuse de ce résultat, voir les références en fin de chapitre.

Nous ajouterons par la suite, quand cela sera nécessaire, un argument (N) (respectivement (N^ψ)) quand les entrées et temps de service de la première période d'activité sont donnée par N (respectivement N^ψ). Pour tout $n \in \llbracket 0, \tau - 1 \rrbracket$, l'instant où le client C_n termine son service sous FIFO, si les temps de service suivent N^ψ , est donné par :

$$\begin{aligned} T_n^{\text{FIFO}}(N^\psi) &= \sum_{i=0}^n \sigma_{\psi(i)} \\ &= \sum_{i=0}^{\psi^{-1}(\psi(n))} \sigma_{\psi(i)} \\ &= T_{\psi(n)}^\psi(N), \end{aligned} \quad (4.27)$$

c'est-à-dire l'instant où le client $C_{\psi(n)}$ termine son service sous ψ si les temps de service sont donnés par N . Par conséquent :

$$\text{TS}_n^{\text{FIFO}}(N^\psi) = T_n^{\text{FIFO}}(N^\psi) - T_n = T_{\psi(n)}^\psi(N) - T_n. \quad (4.28)$$

Notons à partir de maintenant par des lettres en gras les vecteurs à τ composantes représentant les différentes grandeurs pour chaque client de la première période d'activité, par exemple :

$$\mathbf{T}^{\Psi}(N) = (T_0^{\Psi}(N), \dots, T_{\tau-1}^{\Psi}(N)).$$

Remarquons que \mathbf{T}_ψ^{Ψ} est, par définition, la version totalement ordonnée de \mathbf{T}^{Ψ} . Donc, d'après le lemme A.14 :

$$\mathbf{T}_\psi^{\Psi}(N) - \mathbf{T} <_c \mathbf{T}^{\Psi}(N) - \mathbf{T} = \mathbf{A}^{\Psi}(N),$$

c'est-à-dire avec (4.28) :

$$\text{TS}^{\text{FIFO}}(N^\psi) <_c \text{TS}^{\Psi}(N).$$

Donc d'après le (i) du lemme A.14, pour toute fonction convexe et symétrique $F : \mathbf{R}^\tau \rightarrow \mathbf{R}$:

$$F(\text{TS}^{\text{FIFO}}(N^\psi)) \leq F(\text{TS}^\Psi(N))$$

et en particulier pour toute fonction convexe $g : \mathbf{R} \rightarrow \mathbf{R}$:

$$\sum_{n=1}^{\tau-1} g(\text{TS}_n^{\text{FIFO}}(N^\psi)) \leq \sum_{n=1}^{\tau-1} g(\text{TS}_n^\Psi(N)).$$

Finalement, comme les périodes d'activités sont indépendantes et indistinguables en loi d'après l'hypothèse d'indépendance :

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N g(\text{TS}_n^{\text{FIFO}}(N^\psi)) \leq \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N g(\text{TS}_n^\Psi(N))$$

et on retrouve (4.24) avec (4.26) et le théorème de Birkhoff. \square

Les deux corollaires suivants sont à démontrer en exercice.

COROLLAIRE 4.11.– Pour toute fonction convexe $g : \mathbf{R} \rightarrow \mathbf{R}$, les temps d'attente stationnaires sous FIFO et Ψ vérifient :

$$\mathbf{E}[g(\text{TA}^{\text{FIFO}})] \leq \mathbf{E}[g(\text{TA}^\Psi)].$$

COROLLAIRE 4.12.– Pour toute fonction convexe g :

$$\mathbf{E}[g(\text{TS}^\Psi)] \leq \mathbf{E}[g(\text{TS}^{\text{LIFO}})] \text{ et } \mathbf{E}[\text{TA}^\Psi] \leq \mathbf{E}[g(\text{TA}^{\text{LIFO}})].$$

4.1.7. Files avec délais : optimalité d'EDF

On suppose maintenant que les clients ont une échéance pour entrer en service. On note E_n , l'échéance du client C_n et $D_n = E_n - T_n$, le délai initial avant l'échéance de C_n . Nous supposons que la suite $(D_n, n \in \mathbf{Z})$ est stationnaire et nous nous plaçons sur l'espace canonique $(\Omega, \mathcal{F}, \mathbf{P}, \theta)$ des arrivées, des services et des délais. On note alors D , la projection de $(D_n, n \in \mathbf{Z})$ sur sa première coordonnée, interprétée comme le délai avant l'échéance du client C_0 .

Nous supposons que $(\sigma_n, n \in \mathbf{Z})$ est une suite de variables aléatoires indépendantes et identiquement distribuées, indépendantes du processus des arrivées (et donc de $(\xi_n, n \in \mathbf{Z})$ et de $(D_n, n \in \mathbf{Z})$), et que les variables aléatoires ξ , σ et D sont intégrables. Les délais des clients sont souples, par opposition au cas d'impatiences dures abordé en section 4.6, dans le sens où un client qui n'est pas en service avant la fin de son délai ne quitte pas le système, mais continue d'attendre son tour. Les délais

doivent donc être vus comme des indicateurs de l'exigence des clients en termes de délai d'exécution. Nous étudions ci-après la faculté du système à minimiser le retard par rapport à cette exigence, en comparant les différentes politiques de service. Supposons là encore que la condition de stabilité (4.3) pour ce système est vérifiée. Nous notons à nouveau TA_n , le temps d'attente de C_n avant d'atteindre le serveur, et $B_n = T_n + TA_n$ l'instant où C_n entre en service. A tout instant $t \geq T_n$, le délai résiduel (c'est-à-dire le temps résiduel avant l'échéance) de C_n à t est donné par $R_n(t) = E_n - t$. Le délai résiduel de C_n au moment où il commence son service est donc donné par :

$$R_n = R_n(B_n),$$

alors que le retard de C_n par rapport à son échéance s'écrit :

$$L_n = (R_n)^- = -R_n \wedge 0.$$

On définit deux disciplines sensibles aux échéances :

- la discipline *Earliest Deadline First* (EDF) donne toujours priorité au client ayant l'échéance la plus précoce ;
- la discipline *Latest Deadline First* (LDF) donne priorité au client ayant l'échéance la plus tardive.

Par ailleurs, le système étant stable, il existe pour toute Φ un délai résiduel stationnaire à l'entrée en service et un retard stationnaire. Ceux-ci sont donnés respectivement par :

$$R^\Phi = D - TA^\Phi \text{ et } L^\Phi = (R^\Phi)^-.$$

Nous établissons maintenant l'analogie du théorème 4.10 dans le cas d'une file avec délais.

THÉORÈME 4.13.– Pour toute fonction convexe $g : \mathbf{R} \rightarrow \mathbf{R}$ et toute discipline admissible et non préemptive Ψ , ne dépendant pas des temps de service :

$$\mathbf{E} [g(R^{\text{EDF}})] \leq \mathbf{E} [g(R^\Psi)]. \quad (4.29)$$

Démonstration. Les notations déjà introduites dans la preuve du théorème 4.10 ne sont pas rappelées ici. Nous notons pour tout $j \geq 0$, $C_{\alpha(j)}$ le j^{e} client dans l'ordre croissant des échéances (c'est-à-dire $D_{\alpha(i)} \leq D_{\alpha(j)}$ pour $i < j$), et définissons l'application :

$$\zeta = \alpha \circ \psi \circ \phi^{-1} \circ \alpha^{-1},$$

où $\phi(k)$ est pour tout k l'indice du k^{e} client servi par EDF. Le temps d'arrêt τ , indépendant de la discipline considérée, est défini comme précédemment et l'on note maintenant les vecteurs aléatoires suivants :

$$N = ((\xi_0, \sigma_0, D_0), \dots, (\xi_{\tau-1}, \sigma_{\tau-1}, D_{\tau-1}))$$

$$N^\zeta = ((\xi_0, \sigma_{\psi(0)}, D_0), \dots, (\xi_{\tau-1}, \sigma_{\psi(\tau-1)}, D_{\tau-1})),$$

autrement dit N^ζ réarrange les services suivant ζ pendant la première période d'activité. Alors, N et N^ζ ont même distribution comme en (4.26). Pour tout $n \in \llbracket 0, \tau - 1 \rrbracket$:

$$\begin{aligned}
B_{\alpha(n)}^{\text{EDF}}(N^\zeta) &= \sum_{i=0}^{F^{-1}(n)-1} \sigma_{\zeta \circ \alpha \circ \phi(i)} \\
&= \sum_{i=0}^{\phi^{-1}(n)-1} \sigma_{\alpha \circ \psi \circ \phi^{-1} \circ \alpha^{-1} \circ \alpha \circ \phi(i)} \\
&= \sum_{i=0}^{\phi^{-1}(n)-1} \sigma_{\alpha \circ \psi(i)} \\
&= B_{\zeta \circ \alpha(n)}^\psi(N).
\end{aligned}$$

On a donc :

$$\begin{aligned}
R_{\alpha(n)}^{\text{EDF}}(N^\zeta) &= E_{\alpha(n)}(N^\zeta) - B_{\alpha(n)}^{\text{EDF}}(N^\zeta) \\
&= E_{\alpha(n)}(N^\zeta) - B_{\zeta \circ \alpha(n)}^\Psi(N) \\
&= E_{\alpha(n)}(N) - B_{\zeta \circ \alpha(n)}^\Psi(N). \tag{4.30}
\end{aligned}$$

Mais d'autre part, en renvoyant le lecteur aux notions introduites en appendice A.3.

LEMME 4.14.– ζ est la composition de permutations ordonnantes de $\mathbf{B}_\alpha^\Psi(N)$.

Preuve du lemme 4.14. Le premier entier n , s'il existe, tel que $\zeta(\alpha(n)) \neq \alpha(n)$ est tel qu'à la $\alpha(n)$ ^e fin de service sous Ψ (qui est aussi la $\alpha(n)$ ^e fin de service sous EDF puisque $\zeta(k) = k$ pour $k \in \llbracket 0, \alpha(n) - 1 \rrbracket$), il y a dans le système deux clients C_{i_1} et C_{i_2} tels que $D_{i_1} < D_{i_2}$, et EDF choisit C_{i_1} alors que Ψ choisit C_{i_2} . En d'autres termes, en notant pour $\ell = 1, 2$, $j_\ell = \alpha^{-1}(i_\ell)$, on a $B_{\alpha(j_2)}^\Psi(N) < B_{\alpha(j_1)}^\Psi(N)$ alors que $i_2 = \alpha(j_2) > \alpha(j_1) = i_1$. Mais comme EDF donne priorité à C_{i_1} sur C_{i_2} , on a $\phi^{-1}(j_1) < \phi^{-1}(j_2)$. Donc :

$$B_{\zeta \circ \alpha(j_1)}^\Psi(N) = B_{\alpha \circ \psi \circ \phi^{-1}(j_1)}^\Psi(N) < B_{\alpha \circ \psi \circ \phi^{-1}(j_2)}^\Psi(N) = B_{\zeta \circ \alpha(j_2)}^\Psi(N).$$

Donc la permutation ζ_1 échangeant i et j ordonne $B_\alpha^\Psi(N)$. On conclut en remarquant que ζ s'écrit $\zeta = \zeta_p \circ \dots \circ \zeta_1$, où les ζ_i sont de telles permutations. \square

D'après le lemme A.14 et le lemme 4.14, on a donc :

$$\mathbf{E}_\alpha(N) - \mathbf{B}_{\zeta \circ \alpha}^{\text{EDF}}(N) <_c \mathbf{E}_\alpha(N) - \mathbf{B}_\alpha^\Psi(N) = \mathbf{R}_\alpha^\Psi(N),$$

et donc d'après (4.30) :

$$\mathbf{R}_\alpha^{\text{EDF}}(N^\zeta) <_c \mathbf{R}_\alpha^\Psi(N).$$

On conclut alors comme dans le preuve du théorème 4.10. \square

Comme $g : x \mapsto x^-$ est une fonction convexe, et par définition de EDF et LDF, on a en particulier le corollaire suivant.

COROLLAIRE 4.15.– Le retard moyen à l'équilibre est minimisé par EDF et maximisé par LDF. Autrement dit, pour toute discipline admissible Ψ :

$$\mathbf{E}[L^{\text{EDF}}] \leq \mathbf{E}[L^\Psi] \leq \mathbf{E}[L^{\text{LDF}}].$$

4.2. La file *processor sharing*

Nous introduisons maintenant un système de service d'un type particulier, qui a la capacité de servir tous les clients en même temps (il n'y a donc par définition, pas de salle d'attente). Le prix à payer pour un tel mécanisme (qui modélise de nombreux systèmes physiques – voir les références en fin de chapitre) est que la vitesse de traitement instantanée pour chaque client est divisée par le nombre de clients dans le système. Autrement dit, s'il y a p clients dans le système à un instant donné, leurs temps de service résiduels respectifs décroissent de $1/p$ par unité de temps.

Nous faisons les mêmes hypothèses probabilistes, et conservons les mêmes notations qu'auparavant. Comme le serveur travaille, quoiqu'il arrive, à vitesse unité dès que le système n'est pas vide, il est facile de se convaincre que la suite $(W_n, n \in \mathbf{N})$ des charges de travail satisfasse sur l'espace canonique l'équation de Lindley (4.1). Il existe donc une charge de travail stationnaire à la condition (4.3).

Pour caractériser plus complètement l'état d'équilibre, nous cherchons à construire les versions stationnaires de caractéristiques remarquables, telles que la congestion du système et le temps d'attente ou de séjour. Or, le profil de service à l'équilibre, dont nous déduirons ces quantités, n'a pas la même forme sous ces hypothèses que pour une file G/G/1 classique. Nous montrons ici comment le construire, en utilisant les événements de renouvellement.

Nous renvoyons le lecteur aux notions introduites en appendice A.3 concernant l'espace de suites \mathcal{S} . Nous définissons pour tout n , S_n^{PS} le profil de service à T_n^- partant d'un profil arbitraire $S_0^{\text{PS}} \in \mathcal{S}$, en rangeant par convention les termes non nuls de S_n^{PS} par ordre décroissant. Clairement, $S_n^{\text{PS}} \in \mathcal{S}$ pour tout $n \in \mathbf{N}$. On a la résultat suivant.

LEMME 4.16.– La suite $(S_n^{\text{PS}}, n \in \mathbf{N})$ est récurrente dans \mathcal{S} : notons pour tout $u \in \mathcal{S}$ et $x \in \mathbf{R}^+$:

$$\left\{ \begin{array}{ll} \gamma_i(u, x) = \frac{1}{i} \left(x - \sum_{j=i+1}^{+\infty} u(j) \right) & \text{pour tout } i \in \mathbf{N}^*; \\ i_0(u, x) = \min\{i \in \mathbf{N}^*; u(i) \leq \gamma_i(u, x)\}; \\ \gamma(u, x) = \gamma_{(i_0(u, x)-1) \vee 1}(u, x); \\ (F^{\text{PS}}(u, x))(k) = [u(k) - \gamma(u, x)]^+ & \text{pour tout } k \in \mathbf{N}^*. \end{array} \right.$$

Pour tout $n \in \mathbf{N}$, on a :

$$S_{n+1}^{\text{PS}} = F^{\text{PS}}(\underline{F^1(S_n^{\text{PS}}, \sigma \circ \theta^n)}, \xi \circ \theta^n), \quad (4.31)$$

où \underline{u} désigne la version réordonnée de u par ordre décroissant et $F^1(\cdot, \sigma)$ est définie par (4.12).

Démonstration. Nous notons comme plus haut :

$$S_{n++}^{\text{PS}} = \underline{F^1(S_n^{\text{PS}}, \sigma \circ \theta^n)},$$

le profil juste après T_n et après le réordonnement des temps de service à T_n par valeurs décroissantes.

Notons pour tout $i \in \mathbf{N}^*$ tel que $S_{n++}^{\text{PS}}(i) \neq 0$, \tilde{C}_i le client de temps de service résiduel $S_{n++}^{\text{PS}}(i)$ à T_n et \tilde{T}'_i , le temps de sortie virtuel de \tilde{C}_i si aucun client ne rentrait après T_n . Bien sûr, \tilde{T}'_i n'est pas égal à $T_n + S_{n++}^{\text{PS}}(i)$ si C_n n'est pas entré dans un système vide.

Rappelons que $N(S_{n++}^{\text{PS}})$ désigne le nombre de termes non nuls de S_{n++}^{PS} . Pour tout $i \in \{1, \dots, N(S_{n++}^{\text{PS}})\}$, \tilde{C}_i et \tilde{C}_{i-1} reçoivent tous deux la quantité de service $S_{n++}^{\text{PS}}(i)$ sur l'intervalle de temps $[T_n, \tilde{T}'_i]$. Donc, le temps de service résiduel de \tilde{C}_{i-1} à \tilde{T}'_i égale $S_{n++}^{\text{PS}}(i-1) - S_{n++}^{\text{PS}}(i)$, et ce client ainsi que tous ceux qui le suivent sont servis à la vitesse $\frac{1}{i-1}$ sur l'intervalle de temps $[\tilde{T}'_i, \tilde{T}'_{i-1}]$. On a par conséquent la formule de récurrence :

$$\tilde{T}'_{i-1} = \tilde{T}'_i + (S_{n++}^{\text{PS}}(i-1) - S_{n++}^{\text{PS}}(i))(i-1), \quad i \in \llbracket 2, N(S_{n++}^{\text{PS}}) \rrbracket,$$

dont nous déduisons que pour tout $i \in \llbracket 1, N(S_{n++}^{\text{PS}}) \rrbracket$:

$$\tilde{T}'_i = T_n + i S_{n++}^{\text{PS}}(i) + \sum_{j=i+1}^{N(S_{n++}^{\text{PS}})} S_{n++}^{\text{PS}}(j).$$

Pour tout i , \tilde{C}_i est servi avant T_{n+1} si $\tilde{T}'_i - T_n \leq \xi \circ \theta^n$ ou en d'autres termes si :

$$S_{n++}^{\text{PS}}(i) < \frac{1}{i} \left(\xi \circ \theta^n - \sum_{j=i+1}^{N(S_{n++}^{\text{PS}})} S_{n++}^{\text{PS}}(j) \right).$$

En particulier, $i_0 = i_0(S_{n++}^{\text{PS}}, \xi \circ \theta^n)$ représente l'indice du dernier client à quitter le système avant T_{n+1} (ou 0 s'il n'y a aucun départ entre T_n et T_{n+1}).

Alors, le système n'est pas vide juste avant T_{n+1} si $i_0 < N$, et dans ce cas $\{\tilde{C}_i, i \in \llbracket 1, i_0 - 1 \rrbracket\}$ est la famille des clients présents dans le système à cet instant T_{n+1} . Pour un tel client \tilde{C}_i , le temps de service résiduel à T_{n+1} égale son temps de service résiduel à \tilde{T}'_{i_0} moins la quantité de travail reçue entre \tilde{T}'_{i_0} et T_{n+1} , c'est-à-dire :

$$S_{n++}^{\text{PS}}(i) - S_{n++}^{\text{PS}}(i_0) - \frac{T_{n+1} - \tilde{T}'_{i_0}}{i_0 - 1} = S_{n++}^{\text{PS}}(i) - \gamma(S_{n++}^{\text{PS}}, \xi \circ \theta^n).$$

D'où le résultat. □

Un profil stationnaire correspond donc à l'unique solution S^{PS} de l'équation :

$$S^{\text{PS}} \circ \theta = G^{\text{PS}}(S^{\text{PS}}), \mathbf{P} \text{ - presque sûrement,} \quad (4.32)$$

où G^{PS} est l'application aléatoire : $\mathcal{S} \rightarrow \mathcal{S}$ définie sur Ω par :

$$G^{\text{PS}}(u) = F^{\text{PS}}\left(\underline{F^1}(u, \sigma), \xi\right).$$

THÉORÈME 4.17.— A la condition (4.3), l'équation (4.32) admet une unique solution à valeurs dans \mathcal{S} .

Démonstration. La charge de travail à un instant donné égale la somme des termes du profil à cet instant. Donc, comme dans la preuve du théorème 4.7, la même suite $(\mathfrak{A}_n, n \in \mathbf{N})$ est une suite stationnaire d'événements de renouvellement de longueur 1 pour toute suite $(S_n^{\text{PS}, \mu}, n \in \mathbf{N})$ issue de $\mu \in \mathcal{S}$ telle que $\sum_{i \in \mathbf{N}^*} \mu(i) \leq W$, où W est l'unique solution de (4.2). Là encore, le théorème 2.11 implique l'existence d'une solution à (4.32). L'unicité découle encore de l'égalité de deux solutions sur l'événement non négligeable $\{W = 0\}$. □

On peut alors raisonner comme en section 4.1.4.

COROLLAIRE 4.18.— Il existe une unique congestion stationnaire X^{PS} à la condition (4.3). De plus, pour toute condition initiale de la famille \mathcal{Z} (définie par (4.18)), la suite des congestions converge avec couplage arrière fort vers X^{PS} .

REMARQUE.— On peut également, comme en section 4.1.4, construire un temps de séjour (ou autrement dit, un temps de service) stationnaire dans le système à partir de S^{PS} . Ce développement est laissé en exercice.

4.3. Files en parallèle

Considérons maintenant un système soumis au même type de flux d'arrivée G/G , mais où S serveurs (où $S \geq 1$) traitent les demandes des clients de manière conservative et sans vacation. Il y a une file d'attente de capacité illimitée pour chaque serveur. On affecte les clients à leur arrivée à l'un des serveurs libres ou s'il n'y en a pas, au serveur ayant la charge de travail la plus petite. Une fois affecté à un serveur, tout client le reste jusqu'à sa sortie du système, il n'y a donc pas d'échange.

4.3.1. Résultat préliminaire

On commence par introduire un résultat technique, utile pour la suite. On se place sur un quadruplet stationnaire ergodique donné $(\Omega, \mathcal{F}, \mathbf{P}, \theta)$, sur lequel on définit α et β , deux variables aléatoires intégrables à valeurs dans \mathbf{R}_+ . On suppose de plus que $\mathbf{P}(\beta > 0) > 0$. Soit $F^{\alpha, \beta}$ la fonction aléatoire : $\mathbf{R}^+ \rightarrow \mathbf{R}^+$ définie pour tout $x \in \mathbf{R}$ par :

$$F^{\alpha, \beta}(x) = [x \vee \alpha - \beta]^+. \quad (4.33)$$

Le résultat suivant découle de la théorie de Loynes, comme pour la file $G/G/1$. Sa preuve est laissée en exercice.

THÉORÈME 4.19.— Il existe une unique solution \mathbf{P} -presque sûrement finie à l'équation :

$$Z \circ \theta = F^{\alpha, \beta}(Z), \quad (4.34)$$

donnée par :

$$Y^{\alpha, \beta} = \left[\sup_{j \in \mathbf{N}^*} \left(\alpha \circ \theta^{-j} - \sum_{i=1}^j \beta \circ \theta^{-i} \right) \right]^+. \quad (4.35)$$

De plus, pour toute variable aléatoire Z presque sûrement finie et positive, la SRS $(Y_n^Z, n \in \mathbf{N})$ couple avec $(Y^{\alpha, \beta} \circ \theta^n, n \in \mathbf{N})$, et il existe une presque sûrement une infinité d'indices n tels que $Y_n^Z = 0$ si et seulement si :

$$\mathbf{P}(Y^{\alpha, \beta} = 0) > 0. \quad (4.36)$$

Nous représentons le système à S files en parallèle, à l'arrivée de chaque client, par une variable aléatoire à valeurs dans l'espace $(\mathbf{R}^+)^S$ (voir A.3), représentant la charge de travail de chaque serveur à cet instant, rangée dans l'ordre croissant. On

part, à l'instant 0, d'un état initial $V_0 = (V_0(1), \dots, V_0(s)) \in \overline{(\mathbf{R}^+)^S}$, où pour tout i , $V_0(i)$ représente la charge de travail du serveur ayant la i^e charge de travail la plus petite. Ensuite, on représente le système à l'arrivée du client C_n , $n \geq 0$ par le vecteur $V_n \in \overline{(\mathbf{R}^+)^S}$, où $V_n(i)$ est la i^e plus petite charge de travail d'un serveur juste avant l'arrivée de C_n . On appelle alors V_n le profil de service à cet instant.

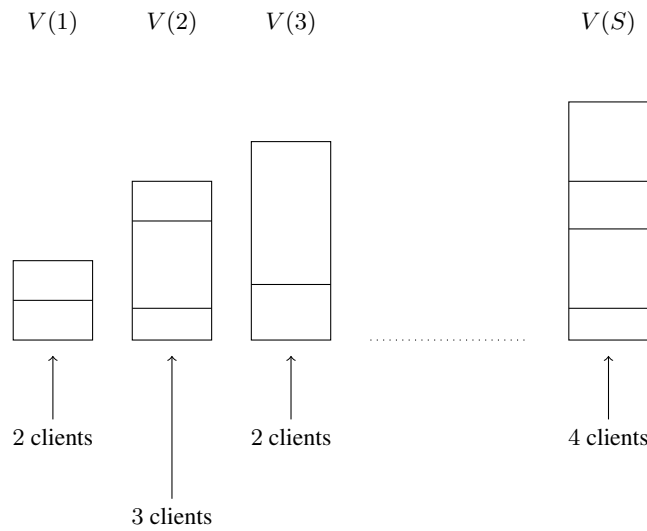


Figure 4.1 – Le vecteur des charges de travail. Les portions de la colonne $V(i)$ représentent les temps de service des clients qui seront par le serveur de charge $V(i)$.

Il est alors facile de vérifier, pour tout $n \in \mathbf{N}$, l'équation dite de Kiefer et Wolfowitz :

$$V_{n+1} = \overline{[V_n + \sigma \circ \theta^n \cdot \mathbf{e}_1 - \xi \circ \theta^n \cdot \mathbf{1}]^+}. \quad (4.37)$$

Un profil de service correspond donc de manière univoque à une solution Y à valeurs dans $\overline{(\mathbf{R}^+)^S}$ pour l'équation :

$$Y \circ \theta = \overline{[Y + \sigma \cdot \mathbf{e}_1 - \xi \cdot \mathbf{1}]^+} = G(Y), \quad (4.38)$$

où σ et ξ sont définies dans l'introduction de la section 4.1. L'application G est clairement presque sûrement continue. Il est facile de remarquer qu'elle est presque

sûrement \prec -croissante : si u et v sont tels que $u < v$ dans $\overline{(\mathbf{R}^+)^S}$, alors pour tout $i \in \llbracket 1, S \rrbracket$, presque sûrement :

$$\begin{aligned} G(u)(i) &= [u(i) \vee ((u(1) + \sigma) \wedge u(i+1)) - \xi]^+ \\ &\leq [v(i) \vee ((v(1) + \sigma) \wedge v(i+1)) - \xi]^+ \\ &= G(v)(i), \end{aligned} \quad (4.39)$$

en comprenant que $u(S+1) = v(S+1) = \infty$. On peut donc appliquer le théorème 2.4, et conclure à l'existence d'une solution \prec -minimale Y_∞ , donnée par la limite presque sûre de la suite de Loynes correspondante à $(V_n, n \in \mathbf{N})$, que nous notons dans ce cas $(Y_n, n \in \mathbf{N})$.

4.3.2. Stabilité

Comme pour la file G/G/1, rien ne garantit que la solution minimale Y_∞ que nous venons d'introduire soit presque sûrement finie (au sens où toutes ses coordonnées sont presque sûrement finies). Nous donnons ci-après la condition de stabilité du système, c'est-à-dire une condition suffisante pour que Y_∞ soit bien presque sûrement à valeurs dans $\overline{(\mathbf{R}^+)^S}$.

THÉORÈME 4.20.– Si :

$$\mathbf{E}[\sigma] < S\mathbf{E}[\xi], \quad (4.40)$$

$Y_\infty(i) < \infty$ presque sûrement pour tout $i = 1, \dots, S$, et si :

$$\mathbf{E}[\sigma] > S\mathbf{E}[\xi],$$

$Y_\infty(i) = \infty$ presque sûrement pour tout $i = 1, \dots, S$.

Démonstration. La suite de Loynes vérifie presque sûrement pour tout $n \in \mathbf{N}$:

$$Y_{n+1} = \overline{[Y_n \circ \theta^{-1} + \sigma \circ \theta^{-1} \cdot \mathbf{e}_1 - \xi \circ \theta^{-1} \cdot \mathbf{1}]^+},$$

ce qui implique en particulier d'après (4.39) que :

$$Y_{n+1}(S) = \left[\left((Y_n \circ \theta^{-1}(1) + \sigma \circ \theta^{-1}) \vee Y_n \circ \theta^{-1}(S) \right) - \xi \circ \theta^{-1} \right]^+. \quad (4.41)$$

Comme pour tout i , $(Y_n(i), n \in \mathbf{N})$ tend en croissant presque sûrement vers $Y_\infty(i)$, le passage à la limite presque sûre dans (4.41) implique que :

$$\begin{aligned} (Y_\infty(S)) \circ \theta &= \left[((Y_\infty(1) + \sigma) \vee Y_\infty(S)) - \xi \right]^+ \\ &= F^{Y_\infty(1) + \sigma, \xi}(Y_\infty(S)), \end{aligned} \quad (4.42)$$

en rappelant la notation (4.35). D'après le théorème 4.19, on a donc presque sûrement :

$$Y_\infty(S) = \left[\sup_{j \in \mathbf{N}^*} \left((Y_\infty(1) + \sigma) \circ \theta^{-j} - \sum_{i=1}^j \xi \circ \theta^{-i} \right) \right]^+.$$

Par conséquent, comme $\mathbf{E}[\xi] \geq 0$, à un négligeable près on a :

$$\{Y_\infty(S) = \infty\} \subset \{Y_\infty(1) = \infty\}$$

et donc, comme Y_∞ est ordonné dans l'ordre croissant :

$$\{Y_\infty(S) = \infty\} = \{Y_\infty(i) = \infty, i = 1, \dots, S\}.$$

L'égalité (4.42) implique que l'événement de gauche est θ -contractant. Donc, on est dans l'alternative suivante :

$$Y_\infty(i) < \infty \text{ presque sûrement pour tout } i = 1, \dots, S \quad (4.43)$$

ou :

$$Y_\infty(i) = \infty \text{ presque sûrement pour tout } i = 1, \dots, S. \quad (4.44)$$

Comme chaque serveur travaille à vitesse unité s'il a quelqu'un à servir, on a presque sûrement pour tout $n \in \mathbf{N}$:

$$\left(\sum_{i=1}^S Y_{n+1}(i) \right) \circ \theta = \sum_{i=2}^S [Y_n(i) - \xi]^+ + [Y_n(1) + \sigma - \xi]^+. \quad (4.45)$$

Donc, en notant :

$$S_n = \sum_{i=1}^S Y_n(i)$$

la somme des charges de travail des serveurs à l'entrée de C_n , on a :

$$\begin{aligned} S_{n+1} \circ \theta - S_n &= \sum_{i=2}^S \left([Y_n(i) - \xi]^+ - Y_n(i) \right) + [Y_n(1) + \sigma - \xi]^+ - Y_n(1) \\ &= - \sum_{i=1}^S (\xi \wedge Y_n(i)) - ((\xi - \sigma) \wedge Y_n(1)). \end{aligned} \quad (4.46)$$

Comme la suite $(Y_n, n \in \mathbf{N})$ est presque sûrement $<$ -croissante, $(S_n, n \in \mathbf{N})$ est presque sûrement croissante. En particulier, par θ -invariance $\mathbf{E}[S_{n+1} \circ \theta] - \mathbf{E}[S_n] \geq 0$, ce qui donne avec (4.46) que :

$$\sum_{i=2}^S \mathbf{E}[\xi \wedge Y_n(i)] + \mathbf{E}[(\xi - \sigma) \wedge Y_n(1)] \leq 0.$$

En passant à la limite par convergence monotone :

$$\sum_{i=2}^S \mathbf{E} [\xi \wedge Y_\infty(i)] + \mathbf{E} [(\xi - \sigma) \wedge Y_\infty(1)] \leq 0. \quad (4.47)$$

Donc, (4.44) implique que :

$$\mathbf{E} [\sigma] \geq S \mathbf{E} [\xi],$$

ce qui montre la suffisance de la condition.

D'autre part, comme $x^+ + y^+ \geq (x + y)^+$ pour tous réels x et y , on a d'après (4.45) que pour tout $n \in \mathbf{N}$, presque sûrement :

$$S_{n+1} \circ \theta \geq [S_n + \sigma - S\xi]^+.$$

Rappelons que la suite de Loynes $(\tilde{M}_n, n \in \mathbf{N})$ correspondant à la charge de travail d'une file G/G/1 d'interarrivées générique $\tilde{\xi} = S\xi$, vérifie $\tilde{M}_0 = 0$ presque sûrement et pour tout $n \in \mathbf{N}$:

$$\tilde{M}_{n+1} \circ \theta = [\tilde{M}_n + \sigma - S\xi]^+.$$

L'application $x \mapsto [x + \sigma - S\xi]^+$ étant presque sûrement croissante, comme $S_0 = 0 = \tilde{M}_0$ presque sûrement, une récurrence immédiate montre que $S_n \geq \tilde{M}_n$ presque sûrement pour tout n . Notons \tilde{M}_∞ la solution minimale donnée par (4.4) pour $\tilde{\xi}$. D'après le théorème 4.2, on obtient donc en passant à la limite que dès que $\mathbf{E} [\sigma] > \mathbf{E} [\tilde{\xi}] = S \mathbf{E} [\xi]$:

$$S_\infty \geq \tilde{M}_\infty = \infty,$$

ce qui conclut la preuve. \square

REMARQUE.— Il est facile de construire des exemples où Y_∞ n'est pas la seule solution de l'équation (4.38) à composantes toutes finies. On peut construire une solution maximale à cette équation en s'intéressant aux translatées de la suite de Loynes par une constante.

Nous montrons maintenant qu'ici aussi, la file stable revient presque sûrement une infinité de fois dans un état de congestion finie (voir 4.4).

THÉORÈME 4.21.— A la condition (4.40), la solution minimale Y_∞ de (4.38) vérifie :

$$\mathbf{P} (Y_\infty(1) = 0) > 0. \quad (4.48)$$

En particulier, il y a \mathbf{P} -presque sûrement une infinité d'instant où le système compte un nombre fini de clients.

Démonstration. D'après (4.47), si l'on a $Y_\infty(1) \geq \xi - \sigma$ et $Y_\infty(2) \geq \xi$ presque sûrement, on a alors $\mathbf{E}[\sigma] \geq s\mathbf{E}[\xi]$, ce qui est une absurdité. Donc, sur un événement non négligeable, on a $Y_\infty(1) < \xi - \sigma$ ou $Y_\infty(2) < \xi$, ce qui implique que $Y_\infty(1) \circ \theta = 0$.

Nous montrons en section 4.4 qu'une file ne peut pas être inactive s'il y a un client en attente dans une autre file : ce client aurait choisi la file vide, qui avait à son arrivée la charge la plus faible. L'événement $Y_\infty(1) = 0$ correspond aux états d'équilibre où au plus $S - 1$ serveurs sont occupés, et donc d'après la remarque précédente, où il y a au plus $S - 1$ clients dans le système. On déduit alors de (4.48) que l'on visite un système dans un état de congestion finie presque sûrement une infinité de fois comme dans le corollaire 4.4. \square

REMARQUE.— On peut vérifier par l'exemple que la condition (4.40) n'implique pas que :

$$\mathbf{P}(X_\infty(S) = 0) > 0$$

et donc que le système se vide presque sûrement une infinité de fois : il suffit par exemple de fixer le couple de réels (x, y) tel que $x > 0$ et $x < y \leq 2x$, et de considérer le système déterministe où $\xi = x$, $\sigma = y$ presque sûrement et $S = 2$. Alors, la solution minimale est le plus petit couple ordonné v tel que :

$$v = \overline{\left((v(1) + y - x)^+, (v(2) - x)^+ \right)},$$

qui est clairement donné par :

$$v = (0, y - x).$$

4.3.3. Comparaison de systèmes

Comme en section 4.1.2, la $<$ -monotonie de la SRS des profils nous permet de comparer les états d'équilibre des systèmes en fonction de l'ordre stochastique des variables aléatoires qui le dirigent.

THÉORÈME 4.22.— Soit deux systèmes à S files portés respectivement par les variables aléatoires (σ, ξ) et $(\bar{\sigma}, \bar{\xi})$, éventuellement définies sur deux quaduplets différents. Si l'on a :

$$(\bar{\sigma}, -\bar{\xi}) \leq_{st} (\sigma, -\xi), \quad (4.49)$$

alors les solutions respectives W et \bar{W} de l'équation (4.37) pour (σ, ξ) et $(\bar{\sigma}, \bar{\xi})$ vérifient :

$$\bar{W} \leq_{st} W. \quad (4.50)$$

Démonstration. On applique le théorème 2.15 à $\bar{\alpha} = (\bar{\sigma}, \bar{\xi})$, $\alpha = (\sigma, \xi)$ et :

$$f : \begin{cases} \mathbf{R}^S \times \mathbf{R}^2 & \rightarrow \mathbf{R}^S \\ (x, (y(1), y(2))) & \mapsto \left[x + y(1) \cdot \mathbf{e}_1 + y(2) \cdot \mathbf{1} \right]^+ . \end{cases}$$

On déduit facilement de (4.39) que f est $<$ -croissante en ses deux arguments. \square

REMARQUE.— Comme précédemment, si l'on suppose que $(\sigma \circ \theta^n, n \in \mathbf{N})$ (respectivement $(\bar{\sigma} \circ \bar{\theta}, n \in \mathbf{N})$) est indépendante de $(\xi \circ \theta^n, n \in \mathbf{N})$ (respectivement $(\bar{\xi} \circ \bar{\theta}^n, n \in \mathbf{N})$), il est facile de voir que le théorème précédent s'applique sous l'une des deux conditions suivantes :

$$\begin{aligned} \xi &\stackrel{L}{=} \bar{\xi} \text{ et } \bar{\sigma} \leq_{st} \sigma, \\ \sigma &\stackrel{L}{=} \bar{\sigma} \text{ et } \xi \leq_{st} \bar{\xi}. \end{aligned}$$

4.3.4. L'allocation optimale

Nous montrons ici en quel sens l'allocation d'un client entrant au serveur de plus petite charge de travail est optimale. Considérons un système où, partant d'un profil de service arbitraire U à l'arrivée de C_0 , on affecte tout client entrant à la I^e file (où I est un entier fixé dans $[[1, S]]$), dans l'ordre des charges de travail croissantes, plutôt qu'à la première. Pour ce modèle, la suite $(\tilde{V}_n^U, n \in \mathbf{N})$ des profils de service satisfait la relation de récurrence :

$$\tilde{V}_{n+1}^U = \left[\tilde{V}_n^U + \sigma \circ \theta^n \cdot \mathbf{e}_I - \xi \circ \theta^n \cdot \mathbf{1} \right]^+, \text{ presque sûrement.} \quad (4.51)$$

Notons comme précédemment, $(V_n^U, n \in \mathbf{N})$ la SRS des profils de service ordonnés issue de U si l'on applique toujours l'allocation au serveur de plus petite charge, définie sur le même espace canonique que $(\tilde{V}_n^U, n \in \mathbf{N})$. Dans le théorème suivant, « $<_*$ » désigne l'ordre partiel sur $(\mathbf{R}^+)^S$ introduit dans la définition A.21.

THÉORÈME 4.23.— Pour toute variable aléatoire U à valeurs dans $(\mathbf{R}^+)^S$, pour tout $n \in \mathbf{N}$, presque sûrement :

$$V_n^U <_* \tilde{V}_n^U; \quad (4.52)$$

$$\forall \ell \geq I, V_n^U(\ell) \leq \tilde{V}_n^U(\ell). \quad (4.53)$$

Démonstration. Nous raisonnons par récurrence. Les relations (4.52) et (4.53) sont clairement vérifiées pour $n = 0$, supposons qu'elles le sont au rang n . En posant encore

$u(S+1) = \infty$ pour tout $u \in \overline{(\mathbf{R}^+)^S}$, on a alors presque sûrement pour tout $\ell \geq I$:

$$\begin{aligned} V_{n+1}^U(\ell) &= [V_n^U(\ell) \vee ((V_n^U(1) + \sigma \circ \theta^n) \wedge V_n^U(\ell+1)) - \xi \circ \theta^n]^+ \\ &\leq [\tilde{V}_n^U(\ell) \vee ((\tilde{V}_n^U(I) + \sigma \circ \theta^n) \wedge \tilde{V}_n^U(\ell+1)) - \xi \circ \theta^n]^+ \\ &= \tilde{V}_{n+1}^U(\ell) \end{aligned}$$

et (4.53) est vérifiée au rang $n+1$. En particulier, ceci implique que :

$$\sum_{i=k}^S V_{n+1}^U(i) \leq \sum_{i=k}^n \tilde{V}_{n+1}^U(i) \text{ pour tout } k \geq I$$

et par conséquent il suffit de montrer que :

$$\sum_{i=k}^S V_{n+1}^U(i) \leq \sum_{i=k}^n \tilde{V}_{n+1}^U(i) \text{ pour tout } k \leq I-1 \quad (4.54)$$

pour vérifier (4.52) au rang $n+1$. Fixons donc $k \leq I-1$ et formons les sommes :

$$\begin{aligned} \sum_{i=k}^S V_{n+1}^U(i) &= \sum_{i=k+1}^S [V_n^U(i) - \xi \circ \theta^n]^+ \\ &\quad + [(V_n^U(1) + \sigma \circ \theta^n) \vee V_n^U(k) - \xi \circ \theta^n]^+ ; \end{aligned} \quad (4.55)$$

$$\begin{aligned} \sum_{i=k}^S \tilde{V}_{n+1}^U(i) &= \sum_{i=k; k \neq I}^S [\tilde{V}_n^U(i) - \xi \circ \theta^n]^+ \\ &\quad + [\tilde{V}_n^U(I) + \sigma \circ \theta^n - \xi \circ \theta^n]^+ . \end{aligned} \quad (4.56)$$

Si $V_n^U(k) \geq V_n^U(1) + \sigma \circ \theta^n$, alors (4.55) vaut :

$$\sum_{i=k}^S [V_n^U(i) - \xi \circ \theta^n]^+ \leq \sum_{i=k}^S [\tilde{V}_n^U(i) - \xi \circ \theta^n]^+ \leq \sum_{i=k}^S \tilde{V}_{n+1}^U(i) \text{ presque sûrement,}$$

où l'on a utilisé le (i) du lemme A.15 dans la première inégalité. Il ne reste donc qu'à traiter le cas où $V_n^U(k) \leq V_n^U(1) + \sigma \circ \theta^n$. Alors, (4.55) égale :

$$[V_n^U(1) + \sigma \circ \theta^n - \xi \circ \theta^n]^+ + \sum_{i=k+1}^S [V_n^U(i) - \xi \circ \theta^n]^+ . \quad (4.57)$$

Le vecteur $(V_n^U(1), V_n^U(k+1), \dots, V_n^U(S))$ est ordonné et presque sûrement :

$$(\xi - \sigma, \xi, \dots, \xi) \circ \theta^n = \overline{\left(\xi, \dots, \underbrace{\xi}_{I-1}, \underbrace{\xi - \sigma}_I, \underbrace{\xi}_{I+1}, \dots, \xi \right) \circ \theta^n} .$$

Donc, d'après (A.8) et (A.6) :

$$\begin{aligned} & (V_n^U(1), V_n^U(k+1), \dots, V_n^U(S)) - (\xi - \sigma, \xi, \dots, \xi) \circ \theta^n \\ & <_c (V_n^U(1), V_n^U(k+1), \dots, V_n^U(S)) \\ & \quad - \left(\xi, \dots, \underbrace{\xi}_{I-1}, \underbrace{\xi - \sigma}_I, \underbrace{\xi}_{I+1}, \dots, \xi \right) \circ \theta^n, \end{aligned}$$

où $<_c$ désigne l'ordre Schur-convexe. Comme, pour tout p , la fonction $u \rightarrow \sum_{i=1}^p u^+$ est symétrique et convexe de $(\mathbf{R}^+)^p$ dans \mathbf{R} , d'après (A.7) la somme (4.57) vérifie presque sûrement :

$$\begin{aligned} & [V_n^U(1) + \sigma \circ \theta^n - \xi \circ \theta^n]^+ + \sum_{i=k+1}^S [V_n^U(i) - \xi \circ \theta^n]^+ \\ & \leq [V_n^U(1) - \xi \circ \theta^n]^+ + \sum_{i=k+1; i \neq I}^S [V_n^U(i) - \xi \circ \theta^n]^+ \\ & \quad + [V_n^U(I) + \sigma \circ \theta^n - \xi \circ \theta^n]^+ \\ & \leq \sum_{i=k; i \neq I}^S [V_n^U(i) - \xi \circ \theta^n]^+ + [V_n^U(I) + \sigma \circ \theta^n - \xi \circ \theta^n]^+. \quad (4.58) \end{aligned}$$

Par ailleurs, comme $V_n^U(I) \leq \tilde{V}_n^U(I)$ en vertu de (4.52), les assertions (ii) puis (i) du lemme A.15 montrent que (4.52) implique que presque sûrement :

$$\left[\overline{V_n^U + \sigma \circ \theta^n \cdot \mathbf{e}_I - \xi \circ \theta^n \cdot \mathbf{1}} \right]^+ <_* \left[\overline{\tilde{V}_n^U + \sigma \circ \theta^n \cdot \mathbf{e}_I - \xi \circ \theta^n \cdot \mathbf{1}} \right]^+.$$

En particulier :

$$\begin{aligned} & \sum_{i=k; i \neq I}^S [V_n^U(i) - \xi \circ \theta^n]^+ + [V_n^U(I) + \sigma \circ \theta^n - \xi \circ \theta^n]^+ \\ & \leq \sum_{i=k; i \neq I}^S [\tilde{V}_n^U(i) - \xi \circ \theta^n]^+ + [\tilde{V}_n^U(I) + \sigma \circ \theta^n - \xi \circ \theta^n]^+ \quad (4.59) \end{aligned}$$

et l'on déduit (4.54) de (4.55), (4.57), (4.58) et (4.59). La relation (4.52) est donc bien vérifiée au rang $n + 1$. \square

Le résultat précédent montre en particulier que, partant d'un même profil initial de service et soumettant le système au même trafic, l'allocation au serveur de plus

petite charge optimise la charge de travail totale par rapport à n'importe quelle autre allocation figée à un autre serveur, puisque pour tout $I \in \llbracket 1, S \rrbracket$, presque sûrement :

$$\sum_{i=1}^S V_n^U(i) \leq \sum_{i=1}^S \tilde{V}_n^U(i).$$

D'autre part, comme \tilde{V}_n^U est ordonné, (4.53) implique que $V_n^U(1) \leq \tilde{V}_n^U(I)$ presque sûrement, autrement dit le temps d'attente proposé au n^e client est également minimisé.

Nous pouvons étendre ces résultats à l'état d'équilibre. D'après (4.51), un profil de service stationnaire pour l'allocation au I^e serveur est une solution à valeurs dans $(\mathbf{R}^+)^S$ de l'équation :

$$\tilde{V} \circ \theta = \left[\tilde{V} + \sigma \cdot \mathbf{e}_I - \xi \cdot \mathbf{1} \right]^+ = \tilde{G}(\tilde{V}) \text{ presque sûrement.} \quad (4.60)$$

Pour tous u et v tels que $u < v$, pour tout $i \geq I$, presque sûrement :

$$\begin{aligned} \tilde{G}(u)(i) &= [u(i) \vee ((u(I) + \sigma) \wedge u(i+1)) - \xi]^+ \\ &\leq [v(i) \vee ((v(I) + \sigma) \wedge v(i+1)) - \xi]^+ \\ &= \tilde{G}(v)(i), \end{aligned}$$

alors que pour tout pour tout $i < I$, presque sûrement :

$$\tilde{G}(u)(i) = [u(i) - \xi]^+ \leq [v(i) - \xi]^+ = \tilde{G}(v)(i).$$

L'application \tilde{G} est donc presque sûrement $< -$ croissante et clairement continue. En appliquant le théorème de Loynes, on obtient la solution minimale de (4.60), donnée par \tilde{Y}_∞ , la limite presque sûre, coordonnée par coordonnée, de la suite de Loynes correspondante $(\tilde{Y}_n, n \in \mathbf{N}) = (\tilde{V}_n^0 \circ \theta^{-n}, n \in \mathbf{N})$. D'après le théorème 4.23, en passant à la limite les solutions minimales vérifient presque sûrement :

$$Y_\infty <_* \tilde{Y}_\infty \text{ et } Y_\infty(1) \leq \tilde{Y}_\infty(I).$$

L'allocation au serveur de plus petite charge minimise donc à l'équilibre la charge de travail totale et le temps d'attente proposé.

Il est par ailleurs immédiat de remarquer que pour tout $n \in \mathbf{N}$, presque sûrement $V_n^0(\ell) = 0$ pour tout $\ell < I$, puisqu'au aucun service n'est jamais attribué aux $I - 1$ premières files, toujours inactives. Par conséquent, la restriction de V_∞ à ses $S - (I - 1)$ dernières coordonnées s'écrit clairement comme la solution minimale de (4.38), c'est-à-dire le profil stationnaire minimal pour l'allocation à la plus petite file parmi $S \equiv S - (I - 1)$ files. Nous avons donc montré les deux corollaires suivants.

COROLLAIRE 4.24.– Pour tout $I \in \llbracket 1, S \rrbracket$, la solution $<$ -minimale \tilde{Y}_∞ de (4.60) satisfait $\tilde{Y}_\infty(S) < \infty$ presque sûrement à la condition :

$$\mathbf{E}[\sigma] < (S - I + 1)\mathbf{E}[\xi].$$

De plus, si Y_∞ désigne la solution minimale de (4.38) on a :

$$\begin{aligned} Y_\infty &<_* \tilde{Y}_\infty; \\ \forall \ell \geq I, Y_\infty(\ell) &\leq \tilde{Y}_\infty(\ell), \end{aligned}$$

et en particulier $Y_\infty(1) \leq \tilde{Y}_\infty(I)$.

COROLLAIRE 4.25.– Notons Y_∞^S et $Y_\infty^{S'}$ les solutions minimales de l'équation (4.38) respectivement pour S et S' serveurs, où $1 \leq S' \leq S$. A la condition $\mathbf{E}[\sigma] < S'\mathbf{E}[\xi]$ où elles sont toutes deux presque sûrement finies, elles vérifient :

$$Y_\infty^S(S - i) \leq Y_\infty^{S'}(S' - i) \text{ presque sûrement pour tout } i \in \llbracket 0, S' - 1 \rrbracket.$$

En particulier, $Y_\infty^S(1) \leq Y_\infty^{S'}(1)$.

⌚ Le dernier résultat précise la propriété qu'une augmentation du nombre de files réduit la charge de travail à l'équilibre : si les deux systèmes sont stables, la charge de travail de chaque file du système le plus petit est plus grande que celle de la file correspondante (en partant de la file la plus chargée) dans le système le plus grand. La dernière inégalité signifie que le temps d'attente proposé au client entrant est minimisé par le plus grand système.

4.4. La file à S serveurs

Nous considérons maintenant un système voisin du précédent. Il y a S serveurs qui traitent les requêtes de manière conservative et sans vacation, mais l'architecture du système d'attente est différente : si tous les S serveurs sont occupés, les clients sont mis en attente dans une file unique et de taille infinie, et sont affectés au premier serveur qui se libère dans l'ordre premier entré, premier servi. Les notations et les hypothèses probabilistes sont les mêmes que précédemment – on considère donc une file stationnaire $G/G/S/\infty/\text{FIFO}$. Nous montrons ici que ce système équivaut en fait au précédent. En particulier la condition de stabilité est la même, donnée par (4.40).

Nous représentons à nouveau la file d'attente par la suite de ses profils de service, en gardant la trace des temps de service de tous les clients présents dans le système à l'instant courant. Plus précisément, on fixe $\hat{V}_0 \in \mathcal{S}$ et on note pour tout $n \in \mathbf{N}$, \hat{V}_n l'élément de \mathcal{S} qui représente le temps de service résiduel de tous les clients dans le système à l'arrivée du n^e :

- si les S serveurs sont occupés :
 - les S premières coordonnées de \hat{V}_n sont les temps de service résiduels des S clients en service, rangés par ordre décroissant,
 - les coordonnées suivantes représentent les temps de service demandés par les clients en attente, rangés dans l'ordre des priorités. Autrement dit, pour tout indice $i \in \llbracket S + 1, N(\hat{V}_n) \rrbracket$, $\hat{V}_n(i)$ représente le temps de service du i^{e} client en attente dans l'ordre des arrivées. En particulier, le client de temps de service $\hat{V}_n(N(S + 1))$ sera le prochain à entrer en service, etc. ;
- si $j \leq S$ serveurs sont occupés, $N(\hat{V}_n) = j$ et les $\hat{V}_n(i), i \in \llbracket 1, j \rrbracket$ représentent les temps de service résiduels des clients en service, rangés dans l'ordre décroissant.

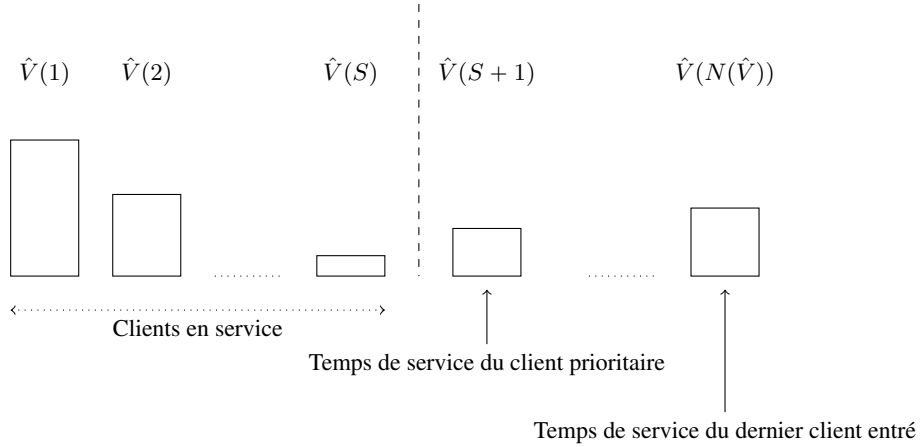


Figure 4.2 – Le profil de services de la file G/G/S

Il est facile de voir que la suite $(\hat{V}_n, n \in \mathbf{N})$ ainsi définie est récurrente sur l'espace canonique des arrivées et des services et d'expliciter la fonction de récurrence. Pour cela, nous construisons pour tout $u \in \mathcal{S}$, la famille d'ensembles d'indices $A_1(u), A_2(u), \dots, A_S(u)$, par récurrence de la façon suivante :

- on fixe au départ $A_1^0(u) = A_2^0(u) = \dots = A_S^0(u) = \emptyset$;
- ensuite, pour tout $j \in \llbracket 1, N(u) - S \rrbracket$, on note :

$$\varphi_u(j) = \operatorname{Argmin}_{i=1, \dots, S} \left\{ u(i) + \sum_{k \in A_i^{j-1}(u)} u(S+k) \right\} \quad (4.61)$$

et on pose :

$$\begin{cases} A_{\varphi_u(j)}^j(u) &= A_{\varphi_u(j)}^{j-1}(u) \cup \{j\}; \\ A_i^j(u) &= A_i^{j-1}(u), \text{ pour tout } i \neq \varphi_u(j). \end{cases}$$

– on pose finalement :

$$A_i(u) = A_i^{N(u)-S}(u) \text{ pour tout } i \in \llbracket 1; S \rrbracket.$$

Bien sûr, on comprendra que $\sum_{k \in \emptyset} \dots = 0$, et on fixera $A_i(u) = \emptyset$ si $N(u) \leq S$.

THÉORÈME 4.26.– Partant de $\hat{V}_0 \in \mathcal{S}$, on a pour tout $n \in \mathbf{N}$:

$$\hat{V}_{n+1} = \hat{G}^3 \circ \hat{G}^2(\cdot, \xi \circ \theta^n) \circ \hat{G}^1(\cdot, \sigma \circ \theta^n)(\hat{V}_n),$$

où \hat{G}^1 , \hat{G}^2 et \hat{G}^3 sont respectivement définies par (4.62), (4.65) et (4.66).

Démonstration. On suppose que C_n trouve un profil de service \hat{V}_n à son arrivée. Tout d'abord, le temps de service $\sigma \circ \theta^n$ apporté par C_n est placé à son arrivée à la place la moins prioritaire, autrement dit le profil de service devient tout d'abord :

$$\hat{V}_{n+} = \hat{V}_n + (\sigma \circ \theta^n) \cdot \mathbf{e}_{N(\hat{V}_n)+1} = \hat{G}^1(\hat{V}_n, \sigma \circ \theta^n). \quad (4.62)$$

Il convient ensuite d'affecter les clients éventuellement en attente aux différents serveurs :

(i) si $N(\hat{V}_n) < S$, il y a des serveurs libres à l'arrivée de C_n et donc le temps de service $\sigma \circ \theta^n$ est affecté au premier serveur libre. \hat{V}_{n+} reste inchangé car le temps de service du client entrant est par construction donné par $\hat{V}_{n+}(N(\hat{V}_{n+}) + 1)$;

(ii) si $N(\hat{V}_{n+}) \geq S$, il n'y a pas de serveur libre à l'arrivée de C_n . Il suffit alors de bien comprendre la construction des ensembles $A_i(\hat{V}_n)$ précédents pour écrire simplement la fonction de récurrence. Appelons « serveur i », $i = 1, \dots, S$, le serveur dont le client en service a un service résiduel égal à $\hat{V}_n(i)$ à l'arrivée de C_n (en particulier, le serveur 1 a la plus grande charge de travail résiduelle et le serveur s a la plus petite à cet instant). Notons également \tilde{C}_j , $j = 1, \dots, N(\hat{V}_{n+}) - S$, le client (s'il existe) en attente à l'arrivée de C_n , dont le temps de service est donné par $\hat{V}_{n+}(i)$ (en particulier, \tilde{C}_1 est le client prioritaire à cet instant et $\tilde{C}_{N(\hat{V}_{n+})-S} = \tilde{C}_{N(\hat{V}_n)+1-S}$ est le client C_n à peine arrivé). Notons bien que ces deux indexations sont relatives à la situation à l'instant d'entrée de C_n , autrement dit elles dépendent de n .

Remarquons d'abord que $1 \in A_S(\hat{V}_{n+})$ par définition. Or, le premier client à éventuellement entrer en service après l'arrivée de C_n est \tilde{C}_1 . Celui-ci rejoindra le

premier serveur qui se libèrera, c'est-à-dire le serveur S . Le deuxième client à entrer en service \tilde{C}_2 rejoindra alors le serveur S si :

$$\hat{V}_{n+}(S) + \hat{V}_{n+}(S+1) \leq \hat{V}_{n+}(S-1),$$

ou le serveur $S-1$ si :

$$\hat{V}_{n+}(S) + \hat{V}_{n+}(S+1) > \hat{V}_{n+}(S-1).$$

Notons alors que dans le premier cas, $2 \in A_S(\hat{V}_{n+})$ et dans le deuxième, $2 \in A_{S-1}(\hat{V}_{n+})$, par définition. Ainsi de suite, on remarque que pour tout $j \in \llbracket 1, \dots, N(\hat{V}_{n+}) - S \rrbracket$, $\varphi_{\hat{V}_{n+}}(j)$ défini par (4.61) représente l'indice (à l'instant d'entrée de C_n) du serveur qui servira effectivement le client \tilde{C}_j , puisque c'est le premier dont la charge de travail résiduelle s'annulera après l'entrée en service du client \tilde{C}_{j-1} (ou après l'arrivée de C_n si $j=1$). En d'autres termes, pour tout $i \in \llbracket 1, S \rrbracket$, pour tout $j = 1, \dots, N(\hat{V}_{n+}) - S$, $A_i^j(\hat{V}_{n+})$ représente l'ensemble des indices (dans la numérotation des \tilde{C}_j) des clients présents dans le système après l'entrée de C_n et entrés strictement avant \tilde{C}_{j+1} , et qui entreront en service avec le serveur i . Par conséquent, $A_i(\hat{V}_{n+})$ désigne alors l'ensemble des indices de tous les clients présents juste après l'arrivée de C_n et qui seront affectés au serveur i .

Il est facile de voir, puisque la discipline est FCFS, que les ensembles $A_i^j(\hat{V}_{n+})$ ne diffèrent des $A_i^j(\hat{V}_n)$ que pour l'indice du client à peine entré $j = N(\hat{V}_{n+}) - S$, qui est ajouté à l'ensemble $A_{\varphi_{\hat{V}_{n+}}(N(\hat{V}_{n+})-S)}^{N(\hat{V}_n)-S}$.

Ainsi, entre l'entrée de C_n et celle de C_{n+1} , le serveur d'indice i à l'entrée de C_n sert une quantité de service égale à :

$$\xi \circ \theta^n \wedge \left(\hat{V}_{n+}(i) + \sum_{k \in A_i(\hat{V}_{n+})} \hat{V}_{n+}(S+k) \right).$$

Si $\xi \circ \theta^n$ est inférieur à la somme précédente, le serveur i a encore du travail à l'arrivée de C_{n+1} . Le dernier client à être entré en service au serveur i avant l'arrivée de C_{n+1} est alors :

- le client qui était déjà en service à l'arrivée de C_n si $\hat{V}_{n+}(i) > \xi \circ \theta^n$;
- sinon, c'est le client $\tilde{C}_{\psi(i)}$, où :

$$\psi(i) = \begin{cases} \min \left\{ j \in A_i(\hat{V}_{n+}) \mid \hat{V}_{n+}(i) + \sum_{k \in A_i^j(\hat{V}_{n+})} \hat{V}_{n+}(S+k) > \xi \circ \theta^n \right\}, \\ \text{ou} \\ \max A_i(\hat{V}_{n+}) \text{ si l'ensemble précédent est vide,} \end{cases}$$

puisque l'on comprend bien que $\psi(i)$ désigne l'indice du dernier client qui a eu le temps de rentrer en service avec le serveur d'indice i à l'entrée de C_n , entre l'instant d'arrivée de C_n et celui de C_{n+1} .

Autrement dit, pour tout $j \in \llbracket 1, N(\hat{V}_{n+}) - S \rrbracket$, le client \tilde{C}_j entre en service (avec le serveur $\varphi_{\hat{V}_{n+}}(j)$) avant l'arrivée de C_{n+1} si et seulement si $j \leq \psi(\varphi_{\hat{V}_{n+}}(j))$.

Dans les deux cas (i) et (ii), on peut donc écrire la suite \hat{V}_{n++} qui représente le profil de service juste avant l'arrivée de C_{n+1} et avant réordonnancement de la manière suivante, en comprenant éventuellement que $\sum_{k \in \emptyset} = 0$:

$$\hat{V}_{n++}(i) = \left[\hat{V}_{n+}(i) + \sum_{k \in A_i^{\psi(i)}(\hat{V}_{n+})} \hat{V}_{n+}(k) - \xi \circ \theta^n \right]^+ ; \quad i \in \llbracket 1, S \rrbracket, \quad (4.63)$$

et pour tout $j \in \llbracket 1, N(\hat{V}_{n+}) - S \rrbracket$:

$$\hat{V}_{n++}(S+j) = \begin{cases} 0 & \text{si } j \leq \psi(\varphi_{\hat{V}_{n+}}(j)); \\ \hat{V}_{n+}(S+j) & \text{sinon.} \end{cases} \quad (4.64)$$

Comme précédemment, on note $\tilde{G}^2(\cdot, \xi \circ \theta^n)$ l'application : $(\mathbf{R}^+)^N \times \mathbf{R} \rightarrow (\mathbf{R}^+)^N$ définie par (4.63) et (4.64) et telle que :

$$\hat{V}_{n++} = \tilde{G}^2(\hat{V}_{n+}, \xi \circ \theta^n). \quad (4.65)$$

Finalement, on réordonne par ordre décroissant les temps de service résiduels des clients en service (qui sont les S premières coordonnées de \hat{V}_{n++}). Quand aux éventuelles coordonnées non nulles suivantes, elles représentent les clients en attente dans l'ordre des priorités, jusqu'à l'indice $N(\hat{V}_{n++})$. En notant \hat{G}^3 , l'application : $(\mathbf{R}^+)^N \rightarrow \mathcal{S}$ qui ordonne les S premières composantes d'une suite dans l'ordre décroissant, et qui supprime les 0 des composantes suivantes en conservant leur ordre (pour ne garder que les services résiduels des clients encore en attente à l'arrivée de C_{n+1}), on a donc :

$$\hat{V}_{n+1} = \hat{G}^3(\hat{V}_{n++}). \quad (4.66)$$

D'où le résultat. \square

Définissons l'application Ψ suivante :

$$\Psi : \begin{cases} \mathcal{S} & \rightarrow (\mathbf{R}^+)^S, \\ u & \mapsto \Psi(u) \text{ tel que } \Psi(u)(i) = u(i) + \sum_{j \in A_i(u)} u(S+j) \text{ pour tout } i \in \llbracket 1, S \rrbracket. \end{cases}$$

En l'espèce, pour tout $n \in \mathbf{N}$ et tout $i \in \llbracket 1, S \rrbracket$, la quantité $\Psi(\hat{V}_n)(i)$ représente donc la charge de travail « virtuelle » du serveur d'indice i juste avant l'entrée de C_n , composée des temps de service de tous les clients destinés à être servis par ce serveur. On a le résultat suivant.

LEMME 4.27.– Pour tout $n \in \mathbf{N}$:

$$\overline{\Psi(\hat{V}_{n+1})} = \left[\overline{\Psi(\hat{V}_n) + \sigma \circ \theta^n \cdot \mathbf{e}_1 - \xi \circ \theta^n \cdot \mathbf{1}} \right]^+. \quad (4.67)$$

Démonstration. A l'arrivée de C_n :

(i) si certains serveurs sont libres (c'est-à-dire $N(\hat{V}_n) < S$), $\overline{\Psi(\hat{V}_n)}$ n'est autre que la version ordonnée par ordre croissant de la restriction de \hat{V}_n à ses S premières composantes. Ensuite :

$$\begin{aligned} \hat{V}_{n+} &= \hat{V}_n + \sigma \circ \theta^n \cdot \mathbf{e}_{N(\hat{V}_n)+1}, \\ A_i(\hat{V}_{n+1}) &= \emptyset, \quad i \in \llbracket 1, S \rrbracket, \end{aligned}$$

ce qui implique que :

$$\hat{V}_{n+1} = \left[\hat{V}_n + \sigma \circ \theta^n \cdot \mathbf{e}_{N(\hat{V}_n)+1} - \xi \circ \theta^n \cdot \mathbf{1} \right]^+$$

et donc (4.67) est vérifiée dans ce cas ;

(ii) si tous les serveurs sont occupés, C_n sera affecté au serveur d'indice $\varphi_{\hat{V}_{n+}}(N(\hat{V}_{n+}) - S)$. En FCFS, les affectations des clients aux différents serveurs ne dépendent pas des arrivées ultérieures, et donc tout client en attente à l'arrivée de C_n reste destiné au même serveur après l'arrivée de C_n , après les arrivées suivantes et ainsi de suite jusqu'à son service (même si l'indice du serveur changera éventuellement lors des arrivées successives). Donc, pour tout serveur i auquel C_n ne sera pas affecté, l'ensemble des indices des clients destinés à être servis par i juste avant l'arrivée de C_n est le même que celui juste après l'arrivée de C_n . Autrement dit :

$$\Psi(\hat{V}_{n+})(i) = \Psi(\hat{V}_n)(i); \quad i \neq \varphi_{\hat{V}_{n+}}(N(\hat{V}_{n+}) - S).$$

Ensuite, pour tout i l'ensemble des indices des clients destinés à un serveur donné juste avant l'entrée de C_{n+1} est égal à celui juste après l'arrivée de C_n , auquel on enlève les indices des clients entrés en service entre les deux arrivées, pendant un intervalle de temps $\xi \circ \theta^n$. D'autre part, l'indice du serveur varie éventuellement de i à ℓ suivant l'ordre des temps de service résiduels des clients en service à l'arrivée de C_{n+1} . Par conséquent, on a :

$$\begin{aligned} \forall i \neq \varphi_{\hat{V}_{n+}}(N(\hat{V}_{n+}) - S), \exists \ell \in \llbracket 1, S \rrbracket; \\ \Psi(\hat{V}_{n+1})(\ell) = \left[\Psi(\hat{V}_n)(i) - \xi \circ \theta^n \right]^+. \end{aligned} \quad (4.68)$$

D'autre part, C_n est effectivement affecté au serveur qui a, à son arrivée, la charge de travail virtuelle (constituée par les clients en attente à l'arrivée de C_n) la plus faible,

c'est-à-dire :

$$\Psi(\hat{V}_{n+1}) \left(\varphi_{\hat{V}_{n+1}}(N(\hat{V}_{n+1}) - s) \right) = \overline{\Psi(\hat{V}_n)}(1) + \sigma \circ \theta^n$$

et donc :

$$\exists k \in \llbracket 1, S \rrbracket \text{ tel que } \Psi(\hat{V}_{n+1})(k) = \left[\overline{\Psi(\hat{V}_n)}(1) + \sigma \circ \theta^n - \xi \circ \theta^n \right]^+. \quad (4.69)$$

Clairement, (4.68) et (4.69) impliquent également (4.67) dans ce cas.

□

⌚ La SRS des vecteurs des charges virtuelles des serveurs suit donc l'équation de Kiefer et Wolfowitz. Ceci montre bien que le système équivaut à celui à S files en parallèle où les clients rejoignent la queue la moins chargée : chaque file correspond à un instant donné à un serveur et aux clients qu'il est destiné à servir.

Nous pouvons maintenant étudier la stabilité du système. Notons \hat{G} , l'application aléatoire : $\mathcal{S} \rightarrow \mathcal{S}$ définie pour tout $u \in \mathcal{S}$ par :

$$\hat{G}(u) = \hat{G}^3 \circ \hat{G}^2(., \xi) \circ \hat{G}^1(., \sigma)(u) \text{ presque sûrement.} \quad (4.70)$$

Un profil de service stationnaire correspond comme d'habitude à une solution \hat{V} à l'équation :

$$\hat{V} \circ \theta = \hat{G}(\hat{V}) \text{ presque sûrement.} \quad (4.71)$$

THÉORÈME 4.28.— L'équation (4.71) admet une solution à valeurs dans \mathcal{S} à la condition (4.40). Sinon, elle n'admet aucune solution à valeurs dans \mathcal{S} .

Démonstration. Soit $(V_n, n \in \mathbf{N})$, la suite des profils de service du système à S files en parallèle et Y_∞ , la solution minimale de (4.38). Soit $U \in \mathcal{S}$ tel que :

$$\overline{\Psi(U)} = Y_\infty \text{ presque sûrement.} \quad (4.72)$$

On déduit alors de (4.37) et (4.67) que les SRS :

$$\left(\overline{\Psi(\hat{V}_n^U)}, n \in \mathbf{N} \right) \text{ et } (V_n^{Y_\infty}, n \in \mathbf{N}) = (Y_\infty \circ \theta^n, n \in \mathbf{N})$$

sont presque sûrement confondues.

Soit l'événement :

$$\mathcal{E} = \{Y_\infty(1) = 0\}.$$

D'après la remarque précédente, pour tout $n \in \mathbf{N}$, sur $\theta^{-n}\mathcal{E}$ on a :

$$\overline{\Psi(\hat{V}_n^U)}(1) = 0$$

et donc $A_s(\hat{V}_n^U) = \emptyset$, ce qui implique que $\hat{V}_n^U(S+1) = \hat{V}_n^U(S+2) = \dots = 0$ puisque pour tout $u \in \mathcal{S}$, $S \in A_S(u)$ dès que $u(S+1) > 0$. Donc, sur l'événement $\theta^{-n}\mathcal{E}$:

$$\begin{aligned} \hat{V}_{n+1}^U &= \left[\hat{V}_n^U + \sigma \circ \theta^n \cdot \mathbf{e}_{N(\hat{V}_n^U)+1} - \xi \circ \theta^n \right]^+ \\ &= \left(\underbrace{\left[\Psi(\hat{V}_n^U) + \sigma \circ \theta^n \cdot \mathbf{e}_S - \xi \circ \theta^n \right]^+}_{S \text{ premières composantes}}, 0, 0, \dots \right) \\ &= \left(\left[Y_\infty \circ \theta^n + \sigma \circ \theta^n \cdot \mathbf{e}_1 - \xi \circ \theta^n \right]^+, 0, 0, \dots \right). \end{aligned}$$

La suite $(\theta^{-n}\mathcal{E}, n \in \mathbf{N})$ est donc une suite d'événements de renouvellement de longueur 1 pour $(S_n^U, n \in \mathbf{N})$, pour toute condition initiale U satisfaisant (4.72). En conclusion :

- si (4.40) est vérifiée, en vertu du fait que \mathcal{E} est non négligeable d'après le théorème 4.21, le corollaire 2.12 implique l'existence d'une solution \hat{V} à valeurs dans \mathcal{S} à l'équation (4.71). De plus, il est facile de montrer que $\overline{\Psi(\hat{V})} = Y_\infty$ presque sûrement ;
- si (4.40) n'est pas vérifiée, si (4.71) admettait une solution \hat{V} à valeurs dans \mathcal{S} , une solution V à valeurs dans $(\mathbf{R}^+)^S$ à l'équation (4.38) serait clairement donnée par $V = \overline{\Psi(\hat{V})}$, une absurdité d'après le théorème 4.20.

Le résultat est démontré. □

4.5. Le système à une infinité de serveurs

On considère maintenant un système idéal où tous les clients sont servis en même temps, à pleine vitesse. En d'autres termes, il y a un nombre infini de serveurs, de sorte que tout client soit servi dès son arrivée. Nous supposons là encore que le flux d'arrivée dans le système est de type G/G, et conservons les mêmes notations que précédemment. On notera donc G/G/ ∞ un tel système sans attente. Il est facile de voir que dans ce cas, la suite des charges de travail n'est pas récurrente, puisque la quantité de travail traitée par le serveur entre deux arrivées successives dépend du nombre de clients dans le système à chaque instant entre ces deux dates. Nous présentons ci-après une condition de stabilité pour ce système, et une représentation exhaustive à l'équilibre par le profil de service.

4.5.1. Le profil de services

Comme précédemment, nous nous plaçons sur l'espace \mathcal{S} (voir A.3). Nous notons S_n^∞ , le profil des services à T_n^- , en partant d'un profil S_0^∞ à l'origine. On ordonne arbitrairement les profils S_n^∞ , $n \in \mathbf{N}$ par ordre décroissant. Il est immédiat de remarquer que la suite des profils est récurrente sur \mathcal{S} : pour tout $n \in \mathbf{N}$:

$$S_{n+1}^\infty = \left[\underline{F^1(S_n^\infty, \sigma \circ \theta^n)} - \xi \circ \theta^n \cdot \mathbf{1} \right]^+,$$

où $F^1(\cdot, \sigma \circ \theta^n)$ est définie par (4.12).

En se plaçant sur Ω l'espace canonique des arrivées et des services, l'existence d'un profil de services stationnaire revient donc à celle d'une solution S^∞ à valeurs dans \mathcal{S} à l'équation :

$$S^\infty \circ \theta = G^\infty(S^\infty), \mathbf{P} - \text{presque sûrement}, \quad (4.73)$$

où $G^\infty(\cdot) = \left[\underline{F^1(\cdot, \sigma)} - \xi \cdot \mathbf{1} \right]^+$, \mathbf{P} -presque sûrement.

La condition de stabilité du système est donnée dans la proposition suivante.

THÉORÈME 4.29.– On suppose que σ et ξ sont intégrables et que $\mathbf{P}(\xi > 0) > 0$. A la condition :

$$\mathbf{P} \left(\sup_{j \in \mathbf{N}^*} \left(\sigma \circ \theta^{-j} - \sum_{i=1}^j \xi \circ \theta^{-i} \right) \leq 0 \right) > 0, \quad (4.74)$$

l'équation (4.73) admet une unique solution S^∞ à valeurs dans \mathcal{S} . De plus, en notant Z , l'unique solution de l'équation :

$$Z \circ \theta = [Z \vee \sigma - \xi]^+, \quad (4.75)$$

alors pour toute $\mu \in \mathcal{S}$ telle que :

$$\mu(1) \leq Z, \quad (4.76)$$

il y a couplage arrière fort entre les suites $(S_n^{\infty, \mu}, n \in \mathbf{N})$ et $(S^\infty \circ \theta^n, n \in \mathbf{N})$.

Démonstration. Partant d'un profil quelconque à l'origine, pour tout $n \in \mathbf{N}$ le plus grand temps de service résiduel à T_{n+1}^- est le maximum entre le plus grand terme à T_n^- et le temps de service initial de C_n , moins la quantité de service reçue pour tous les clients entre T_n et T_{n+1} . Autrement dit :

$$S_{n+1}^\infty(1) = [S_n^\infty(1) \vee \sigma \circ \theta^n - \xi \circ \theta^n]^+.$$

L'existence d'un plus grand temps de service résiduel revient donc à celle d'une solution proprement définie à l'équation (4.75). D'après le théorème 4.19, cette équation admet une unique solution presque sûrement finie, donnée par :

$$Z = \left[\sup_{j \in \mathbf{N}^+} \left(\sigma \circ \theta^{-j} - \sum_{i=1}^j \xi \circ \theta^{-i} \right) \right]^+. \quad (4.77)$$

On applique alors la même méthode que dans la preuve du théorème 4.7. Comme $u(1) = 0$ implique $u = \mathbf{0}$ pour toute $u \in \mathcal{S}$, en notant pour tout n , $\mathfrak{B}_n = \{Z \circ \theta^n = 0\}$, $(\mathfrak{B}_n, n \in \mathbf{N})$ est une suite stationnaire d'événements de renouvellement de longueur 1 pour toute suite $(S_n^{\infty, \mu}, n \in \mathbf{N})$ issue de μ , une variable aléatoire à valeurs dans \mathcal{S} telle que :

$$\mu(1) \leq Z, \text{ presque sûrement.}$$

Comme (4.74) revient à $\mathbf{P}(\mathfrak{B}_0) > 0$, le théorème 2.11 implique à nouveau l'existence d'une solution à (4.73). L'unicité de la solution à valeurs dans \mathcal{S} provient du fait que $S(1) = Z$ pour toute solution S , et donc que deux solutions coïncident et sont à valeurs dans \mathcal{S} sur l'événement non négligeable $\{Z = 0\}$. Finalement, la propriété de couplage arrière fort découle du théorème de Borovkov, comme dans le théorème 4.7. \square

Comme précédemment, on en déduit naturellement les corollaires suivants.

COROLLAIRE 4.30.— A la condition (4.74), il existe une unique congestion stationnaire X^∞ et il y a convergence avec couplage arrière fort de la suite des congestions issue de toute μ vérifiant (4.76) vers X^∞ .

COROLLAIRE 4.31.— A la condition (4.74), la file $G/G/\infty$ issue de tout profil fini se vide presque sûrement une infinité de fois.

Démonstration. Comme la plus grande composante $S^\infty(1)$ du profil stationnaire satisfait (4.75), elle est donnée explicitement par (4.77) et l'hypothèse (4.74) implique :

$$\mathbf{P}(S^\infty = \mathbf{0}) = \mathbf{P}(S^\infty(1) = 0) > 0.$$

On conclut alors par un raisonnement analogue à celui du corollaire 4.4. \square

4.5.2. File $GI/GI/\infty$

Supposons que les temps de services et les interarrivées forment deux suites indépendantes, identiquement distribuées et indépendantes l'une de l'autre (système $GI/GI/\infty$). Dans ce cas particulier, la condition de stabilité du système peut être réécrite plus explicitement.

COROLLAIRE 4.32.– Dans le cas GI/GI/ ∞ , les conclusions du théorème 4.29, des corollaires 4.30 et 4.31 sont valables, à la condition de stabilité :

$$\mathbf{P}(\sigma \leq \xi) > 0. \quad (4.78)$$

Démonstration. Il suffit de vérifier que (4.74) équivaut à (4.78) dans ce cas. Bien sûr, (4.74) implique toujours (4.78) puisque :

$$\mathbf{P}(\sigma \leq \xi) = \mathbf{P}(\sigma \circ \theta^{-1} - \xi \circ \theta^{-1} \leq 0)$$

et presque sûrement :

$$\sup_{j \in \mathbf{N}^*} \left(\sigma \circ \theta^{-j} - \sum_{i=1}^j \xi \circ \theta^{-i} \right) \geq \sigma \circ \theta^{-1} - \xi \circ \theta^{-1}.$$

Pour la réciproque, on a presque sûrement pour tout $j \in \mathbf{N}^*$:

$$\begin{aligned} \sigma \circ \theta^{-j} - \sum_{i=1}^j \xi \circ \theta^{-i} \\ = \sigma \circ \theta^{-1} - \xi \circ \theta^{-1} + \sum_{i=1}^{j-1} \left((\sigma \circ \theta^{-1} - \sigma) \circ \theta^{-i} - (\xi \circ \theta^{-1}) \circ \theta^{-i} \right), \end{aligned}$$

et donc en vertu de l'hypothèse d'indépendance :

$$\begin{aligned} & \mathbf{P} \left(\sup_{j \in \mathbf{N}^*} \left(\sigma \circ \theta^{-j} - \sum_{i=1}^j \xi \circ \theta^{-i} \right) \leq 0 \right) \\ & \leq \mathbf{P} \left[\left\{ \sigma \circ \theta^{-1} - \xi \circ \theta^{-1} \leq 0 \right\} \right. \\ & \quad \left. \cap \left\{ \sup_{j \in \mathbf{N}^*} \sum_{i=1}^{j-1} \left((\sigma \circ \theta^{-1} - \sigma) \circ \theta^{-i} - (\xi \circ \theta^{-1}) \circ \theta^{-i} \right) \leq 0 \right\} \right] \\ & = \mathbf{P}(\sigma - \xi \leq 0) \mathbf{P} \left(\sup_{j \in \mathbf{N}^*} \sum_{i=1}^j \left((\sigma \circ \theta^{-1} - \sigma) \circ \theta^{-i} - (\xi \circ \theta^{-1}) \circ \theta^{-i} \right) \leq 0 \right). \end{aligned}$$

La dernière probabilité du membre de droite est strictement positive en appliquant (4.6) aux variables aléatoires $\sigma \circ \theta^{-1} - \sigma$ et ξ et en remarquant que :

$$\mathbf{E}[\sigma \circ \theta^{-1} - \sigma - \xi] = \mathbf{E}[-\xi] < 0$$

d'après le lemme 2.2. Donc, (4.78) implique bien (4.74) dans ce cas. \square

4.6. Les files d'attente avec clients impatientes

Considérons un système à un serveur et de capacité infinie, dans lequel des clients entrent suivant un flux G/G (on conservera les mêmes notations qu'en section 4.1). On suppose de plus que ces clients sont impatientes : le client C_n exige d'être servi avant un certain délai D_n , où l'on supposera que la suite $((\sigma_n, \xi_n, D_n); n \in \mathbf{Z})$ est stationnaire. Ce délai (ou patience) fixe donc une « date butoir » $T_n + D_n$, au-delà de laquelle le client quitte le système, pour toujours, s'il n'a pas été satisfait. On notera un tel système $G/G/1/1 + G - X$, où le troisième G caractérise la suite des délais et où X désigne la discipline de service.

Nous nous intéressons principalement à deux types de systèmes :

- le client C_n exige d'être entré en service avant $T_n + D_n$. Si tel est le cas, on considèrera que le client restera dans le système jusqu'à la fin de son service, même si celle-ci intervient après $T_n + D_n$. Sinon, le client quitte le système à $T_n + D_n$, pour toujours, sans avoir atteint le serveur. Le délai est alors dit jusqu'au début du service. On dira alors que la file est de type (b) , comme *beginning* et on la notera $G/G/1/1 + G(b) - X$;

- le client C_n exige d'avoir été totalement servi avant $T_n + D_n$. Sinon, le client quitte le système pour toujours à $T_n + D_n$ et ce même si son service a été entamé avant cette date. On dit alors que le délai court jusqu'à la fin du service et que la file est de type (e) , comme *end*. On la notera $G/G/1/1 + G(e) - X$.

Dans toute cette section, nous travaillons sur l'espace canonique $(\Omega, \mathcal{F}, \mathbf{P}, \theta)$ de la suite $((\sigma_n, \xi_n, D_n); n \in \mathbf{Z})$ ou autrement dit, sur l'espace de Palm des arrivées, des services et des délais. Les variables aléatoires σ , ξ et D sont alors définies comme dans les sections précédentes, nous supposons qu'elles sont toutes intégrables et que $\mathbf{P}(\xi > 0) > 0$.

4.6.1. Le profil des services et des patiences

Nous donnons dans cette section une représentation exhaustive du système par une SRS gardant la trace de tous les temps de service résiduels et toutes les patiences résiduelles des clients en service et en attente. Pour simplifier l'exposé, nous considérons ici uniquement un système $G/G/1/1+G(b)$ avec discipline de service non préemptive. Il apparaîtra cependant, comme nous le remarquerons, que des descriptions semblables pourront être proposées pour des files (e) et des files avec plusieurs serveurs.

Comme précédemment, on note X_n le nombre de clients dans le système juste avant l'arrivée du client C_n (à T_n^-) et pour tout $i < n$ tel que C_i est dans le système à T_n^- , on note $\varphi_n(i) \in \{1, \dots, X_n\}$, la place de C_i dans la file d'attente dans l'ordre des priorités, la première étant celle occupée par le client en service à T_n^- . Pour chaque tel client C_i , on note :

- $R_n(\varphi_n(i))$ le temps de service résiduel de C_i à T_n^- (déjà défini en section 4.1.3);
- $\tilde{R}_n(\varphi_n(i))$ le temps de patience résiduel de C_i à T_n^- , c'est-à-dire le temps résiduel à T_n avant la fin de la patience de C_i . Autrement dit :

$$\tilde{R}_n(\varphi_n(i)) = T_i + D_i - T_n. \quad (4.79)$$

Pour tout $n \in \mathbf{N}$, on définit $\nu_n \in \mathcal{S}^2$ (voir la définition formelle de \mathcal{S}^2 et les notations complémentaires en appendice A.3), la suite représentant les temps de service résiduels et les patiences résiduelles des clients présents dans le système à cet instant, rangés dans l'ordre inverse des priorités. Par convention, on fixe à 0 le temps de patience résiduel du client en service (ce client ne sera de toute façon plus éliminé) et à $(0, 0)$ les autres composantes de ν_n . Autrement dit :

$$\nu_n(i) = \begin{cases} \left(R_n(X_n + 1 - i), \tilde{R}_n(X_n + 1 - i) \right), & i < X_n; \\ \left(R_n(X_n + 1 - i), 0 \right), & i = X_n; \\ (0, 0), & i > X_n. \end{cases}$$

On appelle ν_n , le profil des services et des patiences résiduels à T_n^- .

Nous détaillons la dynamique du processus des profils sur l'espace de Palm des arrivées, des services et des patiences. On suppose que le client C_n trouve à son arrivée un profil ν_n . Tout d'abord, le temps de service et la patience de C_n sont insérés dans le profil arbitrairement à la première place, c'est-à-dire :

$$\nu_{n+} = \left\{ (\sigma \circ \theta^n, D \circ \theta^n), \nu_n(1), \nu_n(2), \dots \right\} = H^1(\nu_n, \sigma \circ \theta^n). \quad (4.80)$$

Ensuite, comme dans le cas des profils de service d'une file G/G/1, on applique à \mathcal{S}_{n+} la fonction $H^\Phi : \mathcal{S}^2 \rightarrow \mathcal{S}^2$, qui réordonne les composantes de ν_{n+} suivant l'ordre des priorités pour la discipline Φ :

- si Φ ne dépend que des dates d'arrivée et de service des clients (c'est-à-dire $\Phi = \text{FIFO, LIFO, SRPT, etc.}$), H^Φ n'est autre que l'« extension » de l'application F^Φ définie en (4.13) à \mathcal{S}^2 , dans le sens où pour tout $u \in \mathcal{S}^2$:

$$H^\Phi(u)(i) = \left(F^\Phi(u^1)(i), u^2(j) \right), \text{ où } j \text{ est tel que } F^\Phi(u^1)(i) = u^1(j),$$

la deuxième coordonnée « suit » donc la première, réarrangée suivant F^Φ ;

- la discipline Φ peut aussi dépendre des patiences des clients :
 - la discipline *Earliest Deadline First* (EDF) donne priorité non préemptive, à la fin de chaque service, au client en attente dont la patience résiduelle est la plus courte. Par conséquent :

$$H^{\text{EDF}}(u) = \left\{ u(2), u(3), \dots, u(i), u(1), u(i+1), \dots \right\} \text{ si } u^2(i+1) \leq u^2(1) < u^2(i),$$

autrement dit la suite des deuxièmes coordonnées de $H^{\text{EDF}}(u)$ est ordonnée par ordre décroissant,

- la discipline *Latest Deadline First* (EDF) donne priorité non préemptive, à la fin de chaque service, au client en attente dont la patience résiduelle est la plus longue. Ainsi :

$$H^{\text{LDF}}(u) = \left\{ u(2), u(3), \dots, u(i), u(1), u(i+1), \dots \right\} \text{ si } u^2(i+1) \geq u^2(1) > u^2(i),$$

les deuxièmes coordonnées de $H^{\text{LDF}}(u)$ sont donc ordonnées par ordre croissant.

On note alors comme précédemment, pour toute discipline Φ :

$$\nu_{n++} = H^{2,\Phi}(\nu_{n+}). \quad (4.81)$$

Ensuite, les clients sont servis successivement dans l'ordre des priorités pendant une durée $\xi \circ \theta^n$.

Pour décrire cette opération, on définit pour tout $x \in \mathbf{R}^+$ et tout $u \in \mathcal{S}^2$, les ensembles d'indices $\mathcal{B}_x^j(u)$, $j = 1, \dots, N(u) - 1$ à la manière des ensembles $A_i^j(u)$ de la section 4.4, de la façon suivante :

$$\mathcal{B}_x^0(u) = \{0\},$$

et pour tout $j \in \mathbf{N}^*$:

$$\mathcal{B}_x^j(u) = \begin{cases} \mathcal{B}_x^{j-1}(u) \cup \{j\} & \text{si } \sum_{k \in \mathcal{B}_x^{j-1}(u)} u^1(N(u) - k) < x \wedge u^2(N(u) - j); \\ \mathcal{B}_x^{j-1}(u) & \text{sinon.} \end{cases}$$

Finalement, on note :

$$\mathcal{B}_x(u) = \mathcal{B}_x^{N(u)-1}(u) \text{ et } \psi_x(u) = \max \mathcal{B}_x(u).$$

Appelons \tilde{C}_j , $j = 0, \dots, N(\nu_{n++}) - 1$, le client présent dans le système juste après l'arrivée de C_n et dont le temps de service et la patience résiduels sont donnés par $\nu_{n++}(N(\nu_{n++}) - j)$, de telle sorte que \tilde{C}_0 est le client en service, \tilde{C}_1 le client qui le suit dans l'ordre des priorités, etc. et $\tilde{C}_{N(\nu_{n++})-1}$ est le client C_n à peine entré.

L'ensemble $\mathcal{B}_{\xi \circ \theta^n}^j(\nu_{n++})$ contient alors tous les indices (jusqu'à j inclus) des clients qui ont le temps de rentrer en service avant l'arrivée de C_{n+1} . En effet, \tilde{C}_j peut entrer en service avant l'arrivée de C_{n+1} si :

– le temps nécessaire pour servir les clients qui lui sont prioritaires est inférieur à $\xi \circ \theta^n$, c'est-à-dire :

$$\sum_{k \in \mathcal{B}_{\xi \circ \theta^n}^{j-1}(\nu_{n++})} \nu_{n++}^1 (N(\nu_{n++}) - k) \leq \xi \circ \theta^n;$$

– sa patience n'a pas été épuisée avant la fin des services des clients qui lui sont prioritaires, c'est-à-dire :

$$\sum_{k \in \mathcal{B}_{\xi \circ \theta^n}^{j-1}(\nu_{n++})} \nu_{n++}^1 (N(\nu_{n++}) - k) \leq \nu_{n++}^2 (N(\nu_{n++}) - j).$$

L'entier $\psi_{\xi \circ \theta^n}(\nu_{n++})$ représente donc le plus grand indice d'un client entré en service avant l'arrivée de C_{n+1} (ou le client en service à l'arrivée de C_n si son service n'est pas terminé à l'arrivée de C_{n+1}).

Juste avant l'arrivée de C_{n+1} , les composantes de ν_{n++} deviennent donc :

$$\nu_{n++}(N(\nu_{n++}) - j) = (0, 0), \quad j \in \llbracket 0, \psi_{\xi \circ \theta^n}(\nu_{n++}) - 1 \rrbracket, \quad (4.82)$$

puisque tous les clients d'indices correspondant ont soit fini leur service, soit épuisé leur patience avant l'arrivée de C_{n+1} . Les composantes correspondant au client en service à l'arrivée de C_{n+1} sont données par :

$$\begin{aligned} \nu_{n+++}^1 (N(\nu_{n++}) - \psi_{\xi \circ \theta^n}(\nu_{n++})) \\ = \left[\sum_{k=0}^{\psi_{\xi \circ \theta^n}(\nu_{n++})} \nu_{n+++}^1 (N(\nu_{n++}) - k) - \xi \circ \theta^n \right]^+; \end{aligned} \quad (4.83)$$

$$\begin{aligned} \nu_{n+++}^2 (N(\nu_{n++}) - \psi_{\xi \circ \theta^n}(\nu_{n++})) \\ = \left[\nu_{n+++}^2 (N(\nu_{n++}) - \psi_{\xi \circ \theta^n}(\nu_{n++})) - \xi \circ \theta^n \right]^+, \end{aligned} \quad (4.84)$$

la dernière quantité étant nulle si la patience de ce client a expiré après le début de son service. Enfin, pour tout $j \in \llbracket \psi_{\xi \circ \theta^n}(\nu_{n++}), N(\nu_{n++}) - 1 \rrbracket$:

$$\nu_{n+++}^2 (N(\nu_{n++}) - j) = \left[\nu_{n+++}^2 (N(\nu_{n++}) - j) - \xi \circ \theta^n \right]^+, \quad (4.85)$$

$$\nu_{n+++}^1 (N(\nu_{n++}) - j) = \nu_{n+++}^1 (N(\nu_{n++}) - j) \mathbf{1}_{\left\{ \nu_{n+++}^2 (N(\nu_{n++}) - j) > 0 \right\}}, \quad (4.86)$$

les éventuels clients moins prioritaires que le client actuellement en service voient donc leur temps de service résiduel inchangé et leur patience résiduelle diminuer du temps écoulé entre les deux arrivées. Ils sont éliminés (la coordonnée correspondante est fixée à $(0, 0)$) si celle-ci est négative. Les équations (4.82) à (4.86) décrivent donc une application $H^1 : \mathcal{S}^2 \times \mathbf{R}^+ \rightarrow \mathcal{S}^2$ telle que :

$$\nu_{n+++} = H^3(\nu_{n++}, \xi \circ \theta^n). \quad (4.87)$$

Finalement, ν_{n+1} est obtenu en effaçant les composantes égales à $(0, 0)$ intercalées entre des composantes non nulles (qui correspondent à des clients éliminés ou servis entre les arrivées de C_n et de C_{n+1}), en conservant l'ordre des composantes restantes. On écrit alors :

$$\nu_{n+1} = H^4(\nu_{n+++}). \quad (4.88)$$

On a donc démontré le résultat suivant.

THÉORÈME 4.33.— La suite $(\nu_n, n \in \mathbf{N})$ est récurrente pour toute discipline admissible Φ : pour toute valeur initiale $\nu_0 \in \mathcal{S}^2$, pour tout $n \in \mathbf{N}$:

$$\nu_{n+1} = H^\Phi \circ \theta^n(\nu_n),$$

où :

$$H^\Phi = H^4 \circ H^3(\cdot, \xi) \circ H^{2, \Phi} \circ H^1(\cdot, \sigma),$$

définies par (4.80), (4.81), (4.87) et (4.88).

Comme d'habitude, pour une discipline de service admissible Φ donnée, un profil stationnaire est une variable aléatoire ν^Φ à valeurs dans \mathcal{S}^2 satisfaisant l'équation :

$$\nu^\Phi \circ \theta = H^\Phi(\nu^\Phi) \text{ presque sûrement.} \quad (4.89)$$

On a le résultat suivant.

THÉORÈME 4.34.— Pour toute discipline admissible Φ , si :

$$\mathbf{P} \left(\sup_{j \in \mathbf{N}^*} \left((\sigma + D) \circ \theta^{-j} - \sum_{i=1}^j \xi \circ \theta^{-i} \right) \leq 0 \right) > 0, \quad (4.90)$$

la file $G/G/1/1+G(b)-\Phi$ admet un unique profil stationnaire ν^Φ .

Démonstration. Existence. D'après le théorème 4.19 appliqué à $\alpha \equiv \sigma + D$ et $\beta \equiv \xi$, il existe à la condition (4.90) une unique variable aléatoire positive, solution de l'équation :

$$Y \circ \theta = [Y \vee (\sigma + D) - \xi]^+,$$

donnée par :

$$Y^{\sigma+D,\xi} = \left[\sup_{j \in \mathbf{N}^*} \left((\sigma + D) \circ \theta^{-j} - \sum_{i=1}^j \xi \circ \theta^{-i} \right) \right]^+.$$

Définissons pour tout $u \in \mathcal{S}^2$:

$$Z(u) = \sup_{i \in \mathbf{N}^*} (u^1(i) + u^2(i)) \quad (4.91)$$

la plus grande somme des deux coordonnées d'une composante de u et soit χ une variable aléatoire à valeurs \mathcal{S}^2 telle que :

$$Z(\chi) \in \mathcal{Z} = \{ \text{variables aléatoires positives } Z \text{ t.q. } Z \leq Y^{\sigma+D,\xi} \text{ presque sûrement} \}. \quad (4.92)$$

Soit $(\nu_n^\chi, n \in \mathbf{N})$ la suite des profils issue de χ pour la discipline fixée Φ . Comme on l'a vu dans le calcul de la fonctionnelle de récurrence de $(\nu_n^\chi, n \in \mathbf{N})$, pour tout $n \in \mathbf{N}$ et pour tout $i \in \mathbf{N}^*$, il existe un $j \in \mathbf{N}^*$ tel que :

$$\nu_{n+1}^{\chi,1}(i) + \nu_{n+1}^{\chi,2}(i) \leq \left[\nu_{n+1}^{\chi,1}(j) + \nu_{n+1}^{\chi,2}(j) - \xi \circ \theta^n \right]^+, \quad (4.93)$$

puisqu'entre les arrivées de C_n et C_{n+1} , la patience résiduelle d'un client décroît de $\xi \circ \theta^n$ et le temps de service résiduel est au plus constant (l'indice du client dans la file peut changer de i à j entre ν_{n+} et ν_{n+1} , en fonction des priorités). Or, d'après (4.80) :

$$\left\{ \nu_{n+1}^{\chi,1}(j) + \nu_{n+1}^{\chi,2}(j); j \in \mathbf{N}^* \right\} = \left\{ \nu_n^{\chi,1}(j) + \nu_n^{\chi,2}(j); j \in \mathbf{N}^* \right\} \cup \left\{ \sigma \circ \theta^n + D \circ \theta^n \right\},$$

ce qui donne avec (4.93) en passant au *supremum* :

$$Z(\nu_{n+1}^\chi) \leq \left[Z(\nu_n^\chi) \vee (\sigma \circ \theta^n + D \circ \theta^n) - \xi \circ \theta^n \right]^+ \text{ presque sûrement.} \quad (4.94)$$

Par conséquent, pour tout $n \in \mathbf{N}$ tel que $Z(\nu_n^\chi) \leq Y^{\sigma+D,\xi} \circ \theta^n$, on obtient par monotonie avec (4.94) que :

$$\begin{aligned} Z(\nu_{n+1}^\chi) &\leq \left[(Y^{\sigma+D,\xi} \circ \theta^n) \vee (\sigma \circ \theta^n + D \circ \theta^n) - \xi \circ \theta^n \right]^+ \\ &= Y^{\sigma+D,\xi} \circ \theta^{n+1}, \end{aligned}$$

par définition de $Y^{\sigma+D,\xi}$. On a donc montré par récurrence avec (4.92) que :

$$Z(\nu_n^\chi) \leq Y^{\sigma+D,\xi} \circ \theta^n \text{ presque sûrement pour tout } n \in \mathbf{N}. \quad (4.95)$$

Soit maintenant l'événement :

$$A = \{Y = 0\}. \quad (4.96)$$

D'après (4.95), sur $\theta^{-n}A$ on a $Z(\nu_n^x) = 0$ et donc $\nu_n^x = \mathbf{0}$, la suite nulle de \mathcal{S}^2 . Par conséquent, $(\theta^{-n}A, n \in \mathbf{N})$ est une suite d'événements de renouvellement de longueur 1 pour toute condition initiale χ appartenant à la famille \mathcal{Z} définie par (4.92). En vertu de l'hypothèse (4.90), le corollaire 2.12 implique l'existence d'une solution pour (4.89) à valeurs dans \mathcal{S}^2 .

Unicité. Soit ς , une solution de (4.89) pour Φ fixée. D'après (4.94), on a presque sûrement :

$$Z(\varsigma \circ \theta) \leq [Z(\varsigma) \vee (\sigma + D) - \xi]^+. \quad (4.97)$$

Si l'on avait $Z(\varsigma) > \sigma + D$ presque sûrement (ce qui implique en particulier que $Z(\varsigma) \circ \theta > 0$ presque sûrement), on aurait presque sûrement d'après (4.97) :

$$Z(\varsigma) \circ \theta \leq [Z(\varsigma) \vee (\sigma + D) - \xi]^+ = Z(\varsigma) \vee (\sigma + D) - \xi = Z(\varsigma) - \xi,$$

et donc $\mathbf{E}[Z(\varsigma) \circ \theta - Z(\varsigma)] \leq -\mathbf{E}[\xi] < 0$, en contradiction avec le lemme 2.2. On a donc :

$$\mathbf{P}(Z(\varsigma) \leq \sigma + D) > 0. \quad (4.98)$$

D'après (4.97), sur l'événement $\{Z(\varsigma) \leq Y^{\sigma+D, \xi}\}$:

$$Z(\varsigma) \circ \theta \leq [Y^{\sigma+D, \xi} \vee (\sigma + D) - \xi]^+ = Y^{\sigma+D, \xi} \circ \theta.$$

Donc, $\{Z(\varsigma) \leq Y^{\sigma+D, \xi}\}$ est θ -contractant. D'autre part, sur $\{Z(\varsigma) \leq \sigma + D\}$:

$$\begin{aligned} Z(\varsigma) \circ \theta &\leq [Z(\varsigma) \vee (\sigma + D) - \xi]^+ = [\sigma + D - \xi]^+ \\ &\leq [Y^{\sigma+D, \xi} \vee (D + \sigma) - \xi]^+ = Y^{\sigma+D, \xi} \circ \theta, \end{aligned}$$

et (4.98) implique que $\mathbf{P}(Z(\varsigma) \leq Y^{\sigma+D, \xi}) > 0$ et donc :

$$Z(\varsigma) \leq Y^{\sigma+D, \xi} \text{ presque sûrement.}$$

En d'autres termes, pour toute solution ς , $Z(\varsigma)$ appartient à la famille \mathcal{Z} et donc $(\nu_n^{\varsigma}, n \in \mathbf{N}) = (\varsigma \circ \theta^n, n \in \mathbf{N})$ admet $(\theta^{-n}A, n \in \mathbf{N})$ comme suite d'événements de renouvellement de longueur 1. D'après le corollaire 2.13, $\mathbf{P}(A) > 0$ implique l'unicité de la solution ν^Φ . \square

Comme la file se vide à l'arrivée de C_n si et seulement si $\nu_n = 0$ et que ceci est vrai si et seulement si $Z(\nu_n) = 0$, on obtient comme d'habitude le résultat suivant par minoration.

COROLLAIRE 4.35.– Pour toute discipline admissible Φ , la file $G/G/1/1+G(b)-\Phi$ se vide presque sûrement une infinité de fois si (4.90) est vérifiée.

Comme nous l'avons déjà fait nous pouvons déduire du théorème 4.34, l'existence d'une charge de travail et d'une congestion stationnaire.

COROLLAIRE 4.36.– Pour toute discipline admissible Φ , la file $G/G/1/1+G(b)-\Phi$ admet une unique congestion stationnaire X^Φ et une unique charge de travail W^Φ à la condition (4.90).

REMARQUE.– Pour des systèmes $G/G/S/S+G(b)$, $G/G/1/1+G(e)$ et $G/G/S/S+G(e)$, des représentations par des SRS de profils peuvent être proposées. Elles permettent de retrouver des conditions de stabilité voisines, que le lecteur pourra retrouver en exercice.

4.6.2. File $GI/GI/1/1+GI$

On peut appliquer les arguments de la section 4.5.2 pour donner la condition de stabilité du système $GI/GI/1/1+GI(b)$.

COROLLAIRE 4.37.– Pour toute discipline admissible Φ , la file $GI/GI/1/1+GI(b)-\Phi$ est stable et les conclusions du théorème 4.34 et des corollaires 4.35 et 4.36 sont valides sous la condition :

$$\mathbf{P}(\sigma + D \leq \xi) > 0. \quad (4.99)$$

4.6.3. Optimalité d'EDF

Considérons le cas particulier d'une file $G/M/1/1+G(b)$: les temps de service des clients $(\sigma_n, n \in \mathbf{Z})$ sont supposés indépendants et identiquement distribués de loi exponentielle, indépendants des temps d'interarrivées $(\xi_n, n \in \mathbf{Z})$ et des patiences $(D_n, n \in \mathbf{Z})$. Nous montrons dans ce cas que la discipline EDF est optimale, en tant qu'elle est celle qui perd le moins de clients, en un sens que nous allons préciser.

Soient $x \in \mathbf{R}^+$ et $u, v \in \mathcal{S}^2$. On va modifier les deux suites pour en obtenir deux autres \hat{u} et \hat{v} , de la manière suivante :

- on ramène arbitrairement à la même valeur leur première composante non nulle :

$$\hat{u}^1(N(u)) = \hat{v}^1(N(v)) = 0 = u^1(N(u));$$

- les deuxièmes composantes de \hat{u} sont celles de u et celles de \hat{v} sont celles de v ;

– soit :

$$\ell_1 = \min \{j \geq 0; j \in \mathcal{B}_x^j(u) \text{ ou } j \in \mathcal{B}_x^j(v)\},$$

le premier indice appartenant à $\mathcal{B}_x(u)$ ou à $\mathcal{B}_x(v)$. On fixe alors pour tout $j < \ell_1$, $\hat{u}^1(N(u) - j) = u^1(N(u) - j)$ et $\hat{v}^1(N(v) - j) = v^1(N(v) - j)$. Ensuite, si $\ell_1 \in \mathcal{B}_x^j(u) = \mathcal{B}_x^j(\hat{u})$, en appelant :

$$k_1 = \min \{j \geq 0; j \in \mathcal{B}_x^j(v) = \mathcal{B}_x^j(\hat{v})\} \geq \ell_1$$

le premier indice de $\mathcal{B}_x(v)$, on fixe :

$$\hat{v}^1(N(v) - k_1) = u^1(N(u) - \ell_1),$$

c'est-à-dire que l'on rend artificiellement égales les premières coordonnées des premiers indices de $\mathcal{B}_x(u)$ et $\mathcal{B}_x(v)$;

– ensuite, on note :

$$\ell_2 = \min \{j > \ell_1; j \in \mathcal{B}_x^j(\hat{u})\} \wedge \min \{j > k_1; j \in \mathcal{B}_x^j(\hat{v})\},$$

le deuxième indice appartenant à $\mathcal{B}_x(u)$ ou $\mathcal{B}_x(v)$ et l'on recommence la même construction. Si là encore le minimum est donné par celui de gauche, on laisse les premières composantes intermédiaires inchangées, puis l'on note :

$$\hat{v}^1(N(v) - k_2) = u^1(N(u) - \ell_2) = \hat{u}^1(N(u) - \ell_2),$$

où k_2 est le deuxième indice de $\mathcal{B}_x(\hat{v})$, ainsi de suite jusqu'à $\psi_x(\hat{u}) \wedge \psi_x(\hat{v})$.

Ces notations sont compliquées, mais elles décrivent l'idée simple que l'on peut pas à pas, en partant des dernières composantes positives, transformer les deux suites u et v pour en obtenir deux autres, \hat{u} et \hat{v} , dont les deuxièmes composantes sont inchangées et dont les premières composantes, correspondant aux indices successifs de $\mathcal{B}_x(\hat{u})$ et $\mathcal{B}_x(\hat{v})$, sont égales jusqu'à $\psi_x(\hat{u}) \wedge \psi_x(\hat{v})$.

Nous renvoyons maintenant le lecteur à la définition de « < » dans l'ensemble \mathcal{S} (définition A.22). On a le résultat suivant.

LEMME 4.38.– Si $u, v \in \mathcal{S}^2$ sont telles que :

$$u^2 < v^2 \text{ dans } \mathcal{S}, \quad (4.100)$$

alors pour tout $x \in \mathbf{R}^+$:

$$(H^3(\hat{u}, x))^2 < (H^3(\hat{v}, x))^2 \text{ dans } \mathcal{S},$$

où H^3 est définie par (4.87).

Démonstration. Remarquons déjà que par définition de \mathcal{S}^2 , en vertu de (4.100) on a nécessairement $N(\hat{u}) \leq N(\hat{v})$.

Ensuite, on montre par récurrence qu'il existe, pour tout $j \in \llbracket 0, \psi_x(\hat{u}) \rrbracket$, une bijection :

$$F_j : \begin{cases} \mathcal{B}_x^j(\hat{u}) & \rightarrow \mathcal{B}_x^j(\hat{v}) \\ i & \mapsto F_j(i) \leq i. \end{cases}$$

Cette propriété est évidente pour $j = 0$ et si on la suppose vraie à l'indice $j - 1$, on a $\text{Card } \mathcal{B}_x^{j-1}(\hat{u}) = \text{Card } \mathcal{B}_x^{j-1}(\hat{v})$, donc d'après (4.87) et la construction de \hat{u} et \hat{v} :

$$\sum_{k \in \mathcal{B}_x^{j-1}(\hat{u})} \hat{u}^1(N(\hat{u}) - k) = \sum_{k \in \mathcal{B}_x^{j-1}(\hat{v})} v^1(N(\hat{v}) - k). \quad (4.101)$$

Par ailleurs, $j \in \mathcal{B}_x^j(\hat{u})$ signifie que le terme de gauche est inférieur à :

$$x \wedge u^2(N(\hat{u}) - j) \leq x \wedge v^2(N(\hat{u}) - j),$$

d'après (4.100). C'est aussi le cas du terme de droite de (4.101), donc il existe un entier $k \leq j + N(\hat{u}) - N(\hat{v}) \leq j$ tel que $k \in \mathcal{B}_x^k(\hat{v}) \subset \mathcal{B}_j^k(\hat{v})$. La récurrence est donc démontrée, en prenant $F_j(j) = i$.

En conséquence, $\psi_x(\hat{u}) \leq \psi_x(\hat{v})$ et donc il existe une injection :

$$F : \begin{cases} \mathcal{B}_x(\hat{u}) & \rightarrow \mathcal{B}_x(\hat{v}) \\ i & \mapsto F(i) \leq i. \end{cases}$$

Le résultat est finalement une conséquence du fait que $\psi_x(\hat{u}) \leq \psi_x(\hat{v})$ et que pour tout $j > \psi_x(\hat{u})$, $j \in \mathcal{B}_x(\hat{v})$ implique que :

$$\begin{aligned} \hat{u}^2(N(\hat{u}) - j) &\leq \sum_{k \in \mathcal{B}_x(\hat{u})} \hat{u}^1(N(\hat{u}) - k) \\ &= \sum_{k \in \mathcal{B}_x^{j-1}(\hat{v})} \hat{v}^1(N(\hat{v}) - k) \leq \hat{v}^2(N(\hat{v}) - j). \end{aligned}$$

Ainsi, tout terme non égal à $(0, 0)$ de $H^3(\hat{u}, x)$ a une deuxième composante plus petite que le terme correspondant de $H^3(\hat{v}, x)$: ces deux termes sont respectivement le terme de même indice de \hat{u} et de \hat{v} , laissés à l'identique. \square

Soit Φ , une discipline indépendante des temps de service des clients. Dans ce paragraphe, nous ajoutons un exposant Φ ou EDF pour souligner la dépendance en l'une ou l'autre des disciplines de service. Plaçons-nous sous la condition de stabilité (4.90). Il existe donc deux profils stationnaires ν^Φ et ν^{EDF} , respectivement sous Φ et EDF. Plaçons-nous sur l'événement :

$$\{\nu^{2,\Phi} < \nu^{2,\text{EDF}}\} = \{\hat{\nu}^{2,\Phi} < \hat{\nu}^{2,\text{EDF}}\}.$$

Alors, il est facile de voir par définition des applications $H^1(\cdot, \sigma)$, $H^2(\cdot)$ et par définition de la discipline EDF que :

$$\left(H^{2,\Phi} \circ H^1(\cdot, \sigma) (\hat{\nu}^\Phi) \right)^2 < \left(H^{2,\text{EDF}} \circ H^1(\cdot, \sigma) (\hat{\nu}^{\text{EDF}}) \right)^2,$$

puisque EDF range les termes des suites par ordre décroissant suivant leurs deuxièmes coordonnées.

Alors, le lemme 4.38 et la définition de l'application H^4 nous permettent de conclure que :

$$H^\Phi (\hat{\nu}^\Phi)^2 < H^{\text{EDF}} (\hat{\nu}^{\text{EDF}})^2.$$

Donc, l'événement $\{\hat{\nu}^{2,\Phi} < \hat{\nu}^{2,\text{EDF}}\}$ est θ -contractant. Comme il inclut l'événement A et que ce dernier est non négligeable, il est presque-sûr.

En particulier, les congestions stationnaires respectives $N(\hat{\nu}^\Phi)$ et $N(\hat{\nu}^{\text{EDF}})$ satisfont :

$$N(\hat{\nu}^\Phi) \leq N(\hat{\nu}^{\text{EDF}}) \text{ presque sûrement,}$$

et donc pour tout $x \in \mathbf{R}$:

$$\mathbf{P}(N(\hat{\nu}^\Phi) \geq x) \leq \mathbf{P}(N(\hat{\nu}^{\text{EDF}}) \geq x).$$

On conclut en remarquant que la transformation aléatoire passant des suites u à leur transformée \hat{u} reviennent à :

- (1) synchroniser le temps de service résiduel courant à chaque arrivée ;
- (2) synchroniser les temps de service suivant en échangeant éventuellement les demandes de services des clients.

Comme on l'a vu précédemment, l'argument d'échange (voir par exemple (4.26)) permet de conclure que l'opération (2) ne change pas la loi des caractéristiques du système, puisque les temps de service sont indépendants et identiquement distribués et que Φ ne dépend pas des temps de service. D'autre part, d'après le paradoxe de l'autobus, l'hypothèse exponentielle implique de surcroît que l'opération de tirer au sort un temps de service courant identique aux deux systèmes à l'arrivée d'un client (et donc faire comme si ce temps de service était remis à zéro), ne change pas non plus la loi des caractéristiques du système. On en conclut donc que X^Φ et X^{EDF} ont respectivement même loi que $N(\hat{\nu}^\Phi)$ et $N(\hat{\nu}^{\text{EDF}})$. On a donc le résultat suivant.

THÉORÈME 4.39.— EDF maximise stochastiquement la congestion stationnaire pour les files $G/M/1/1/G(b)$: pour tout $x \in \mathbf{R}$:

$$\mathbf{P}(X^{\text{EDF}} \geq x) \geq \mathbf{P}(X^\Phi \geq x).$$

⌚ Cette propriété d'optimalité est cruciale : elle signifie en particulier que la discipline EDF perd moins de clients que toute autre discipline indépendante des temps de service.

4.6.4. Files FIFO à un serveur

Considérons le cas particulier où le serveur traite les demandes dans l'ordre des arrivées. Comme nous allons le voir, dans ce cas le système peut être directement représenté par la charge de travail du serveur, comme dans une file G/G/1 classique.

4.6.4.1. File (b)

Supposons tout d'abord que le système est G/G/1/1+G(b)-FIFO : les délais courent seulement jusqu'au début du service. Ainsi, le client C_n est servi jusqu'au bout de son service, sans interruption, dès lors qu'il a atteint le serveur avant sa date buttoir $T_n + D_n$. Comme en section 4.1, notons pour tout n , W_n la quantité de travail (comptée en unité de temps) soumise au serveur juste avant l'arrivée de C_n . Cette quantité de travail représente donc le temps d'attente proposé à C_n avant d'attendre le serveur.

Nous cherchons à établir la dynamique de la suite $(W_n, n \in \mathbf{N})$ partant d'un état initial donné à l'arrivée de C_0 . A l'arrivée de C_n , on est dans l'alternative suivante :

- si la patience de C_n est supérieure au temps d'attente proposé (c'est-à-dire $D \circ \theta^n \geq W_n$), le client C_n atteindra le serveur et apporte donc une contribution de $\sigma \circ \theta^n$ à la quantité de travail de ce dernier ;
- sinon, $D \circ \theta^n < W_n$ et C_n , même s'il reste dans le système pendant un temps $D \circ \theta^n$, ne contribue pas à la charge de travail puisqu'il n'atteindra jamais le serveur.

Donc, partant d'une charge de travail arbitraire W juste avant l'arrivée de C_0 , on a pour tout $n \in \mathbf{N}$:

$$W_{n+1} = [W_n + \sigma \circ \theta^n \mathbf{1}_{\{[0, D \circ \theta^n]\}}(W_n) - \xi \circ \theta^n]^+. \quad (4.102)$$

La suite $(W_n^W, n \in \mathbf{N})$ est donc une SRS portée par l'application aléatoire :

$$\varphi : x \mapsto [x + \sigma \mathbf{1}_{\{[0, D]\}}(x) - \xi]^+, \text{ presque sûrement,}$$

et une charge de travail stationnaire W résout l'équation :

$$W \circ \theta = \varphi(W) \text{ presque sûrement.} \quad (4.103)$$

L'application φ admet presque sûrement une discontinuité en D et n'est pas monotone : le théorème de Loynes ne s'applique pas ici. Nous donnons ci-après, en utilisant la théorie des événements de renouvellement, une condition suffisante pour l'existence d'une solution à (4.103).

THÉORÈME 4.40.– Si la condition (4.90) est vérifiée, (4.103) admet une unique solution finie W , qui vérifie :

$$W \leq Y^{\sigma+D, \xi}, \text{ presque sûrement,} \quad (4.104)$$

où $Y^{\sigma+D, \xi}$ est définie comme en (4.35). De plus, pour toute variable aléatoire Z telle que $Z \leq Y^{\sigma+D, \xi}$ presque sûrement, les suites $(W_n^Z, n \in \mathbf{N})$ et $(W \circ \theta^n, n \in \mathbf{N})$ couplent pour le couplage arrière fort.

Démonstration. Existence. Remarquons tout d'abord que pour tout $x \in \mathbf{R}^+$:

$$\begin{aligned} & \varphi(x) \\ &= [x + \sigma \mathbf{1}_{\{(-\infty, D]\}}(x) + x \mathbf{1}_{\{(D, D+\sigma]\}}(x) + x \mathbf{1}_{\{(D+\sigma, \infty)\}}(x) - \xi]^+ \\ &\leq [(D + \sigma) \mathbf{1}_{\{(-\infty, D]\}}(x) + (D + \sigma) \mathbf{1}_{\{(D, D+\sigma]\}}(x) + x \mathbf{1}_{\{(D+\sigma, \infty)\}}(x) - \xi]^+ \\ &= [x \vee (\sigma + D) - \xi]^+ = F^{\sigma+D, \xi}(x), \text{ presque sûrement,} \end{aligned} \tag{4.105}$$

où $F^{\sigma+D, \xi}$ est définie comme en (4.33). Comme cette dernière application est presque sûrement croissante, pour tous $x \leq y$:

$$\varphi(x) \leq F^{D+\sigma, \xi}(y), \text{ presque sûrement} \tag{4.106}$$

et une récurrence immédiate montre que $Z \leq Y^{\sigma+D, \xi}$ presque sûrement implique :

$$W_n^Z \leq Y^{\sigma+D, \xi} \circ \theta^n, n \geq 0, \text{ presque sûrement}$$

Donc, si l'on définit pour tout n l'événement :

$$\mathcal{A}_n = \{Y^{\sigma+D, \xi} \circ \theta^n = 0\},$$

on a $\mathcal{A}_n \subseteq \{W_n^Z = 0\}$, $n \geq 0$, et donc $(\mathcal{A}_n, n \in \mathbf{N})$ est une suite θ -compatible d'événements de renouvellement de longueur 1 pour $(W_n^Z, n \in \mathbf{N})$. Par conséquent, comme (4.90) revient à $\mathbf{P}(\mathcal{A}_0) > 0$, c'est d'après le théorème 2.11 une condition suffisante pour l'existence d'une solution W à (4.103) et pour le couplage fort dans le passé de $(W_n^Z, n \in \mathbf{N})$ avec $(W \circ \theta^n, n \in \mathbf{N})$.

Unicité. Soit W , une solution de (4.103). Tout d'abord, si l'on avait $W > D$ presque sûrement (ce qui implique en particulier que $W \circ \theta > 0$ presque sûrement), on aurait :

$$W \circ \theta = W - \xi, \text{ presque sûrement,}$$

en contradiction avec le lemme 2.2. On a donc :

$$\mathbf{P}(W \leq D) > 0. \tag{4.107}$$

D'après (4.106), sur l'événement $\{W \leq Y^{\sigma+D, \xi}\}$:

$$W \circ \theta = \varphi(W) \leq F^{\sigma+D, \xi}(Y^{\sigma+D, \xi}) = Y^{\sigma+D, \xi} \circ \theta.$$

Donc, $\{W \leq Y^{\sigma+D, \xi}\}$ est θ -contractant. Par ailleurs, sur $\{W \leq D\}$:

$$\begin{aligned} W \circ \theta &= [W + \sigma - \xi]^+ \leq [D + \sigma - \xi]^+ \\ &\leq [Y^{\sigma+D, \xi} \vee (D + \sigma) - \xi]^+ = Y^{\sigma+D, \xi} \circ \theta \end{aligned}$$

et (4.107) implique que $\mathbf{P}(W \leq Y^{\sigma+D, \xi}) > 0$, soit (4.104).

Par conséquent, pour toute solution $W : (W_n^W, n \in \mathbf{N}) = (W \circ \theta^n, n \in \mathbf{N})$ admet $(\mathcal{A}_n, n \in \mathbf{N})$ comme suite d'événements de renouvellement de longueur 1. D'après le corollaire 2.13, $\mathbf{P}(\mathcal{A}_0) > 0$ implique l'unicité de W . \square

LEMME 4.41.– On a :

$$Y^{\sigma \wedge D, \xi} \leq W \text{ presque sûrement,} \quad (4.108)$$

où $Y^{\sigma \wedge D, \xi}$ est définie comme en (4.35).

Démonstration. Laissée au lecteur : on utilise une minoration comme en (4.105). \square

4.6.4.2. File à rejet

Considérons une file d'attente G/G/1/1 : le système a un serveur et pas de file d'attente. Les clients, qui entrent suivant un flux G/G, sont donc :

- immédiatement servis si le système est vide à leur arrivée ;
- immédiatement perdus si le serveur est occupé à leur arrivée.

Ce système classique, largement étudié dans la littérature (voir les remarques bibliographiques), est souvent appelé file à rejet. Il facile de voir qu'il est un cas particulier de la file G/G/1/1+G(b)-FIFO, où l'on suppose que les délais sont identiquement nuls, presque sûrement. Alors, en adoptant les mêmes notations que dans la section précédente, la suite des charges de travail aux arrivées des clients vérifie la dynamique :

$$W_{n+1} = [W_n + \sigma \circ \theta^n \mathbf{1}_{\{0\}}(W_n) - \xi \circ \theta^n]^+, \quad n \in \mathbf{N}, \quad \mathbf{P} \text{ – presque sûrement.}$$

Sur l'espace de Palm des arrivées et des services, une charge de travail stationnaire vérifie donc l'équation :

$$W \circ \theta = [W + \sigma \mathbf{1}_{\{0\}}(W) - \xi]^+, \text{ presque sûrement.} \quad (4.109)$$

Le résultat suivant découle directement du théorème 4.40.

COROLLAIRE 4.42.– Si :

$$\mathbf{P} \left(\sup_{j \in \mathbf{N}^*} \left(\sigma \circ \theta^{-j} - \sum_{i=1}^j \xi \circ \theta^{-i} \right) \leq 0 \right) > 0, \quad (4.110)$$

(4.109) admet une unique solution finie W^0 telle que $W^0 \leq Y^{\sigma, \xi}$, où $Y^{\sigma, \xi}$ est définie comme en (4.35).

4.6.4.3. File à plafond de charge

Un autre exemple classique de file à perte peut être traité dans ce cadre : supposons que l'unique serveur du système n'accepte les clients qui arrivent que si sa charge de travail à cet instant n'excède pas un seuil donné, noté $d > 0$. Le serveur sert alors les clients acceptés suivant une discipline de service quelconque X (X préemptive ou non), jusqu'à la fin de leur service. Sous les hypothèses courantes et en utilisant les mêmes notations, la suite des charges de travail vérifie donc la dynamique :

$$W_{n+1} = [W_n + \sigma \circ \theta^n \mathbf{1}_{\{[0, d]\}}(W_n) - \xi \circ \theta^n]^+, \quad n \in \mathbf{N}, \mathbf{P} \text{ – presque sûrement.}$$

La suite $(W_n, n \in \mathbf{N})$ pour ce système égale celle du système G/G/1/1+G(b)-FIFO avec le même flux, même si les deux systèmes sont différents en général puisque l'on ne sert pas les mêmes clients si $X \neq FIFO$.

Une charge stationnaire pour ce modèle est une solution \mathbf{P} -presque sûrement finie de :

$$W \circ \theta = [W + \sigma \mathbf{1}_{\{[0, d]\}}(W) - \xi]^+, \quad \text{presque sûrement.} \quad (4.111)$$

On a ainsi le corollaire suivant.

COROLLAIRE 4.43.– Si :

$$\mathbf{P} \left(\sup_{j \in \mathbf{N}^*} \left(\sigma \circ \theta^{-j} - \sum_{i=1}^j \xi \circ \theta^{-i} \right) \leq -d \right) > 0, \quad (4.112)$$

(4.111) admet une unique solution finie W^d telle que :

$$Y^{\sigma \wedge d, \xi} W^d \leq Y^{\sigma + d, \xi}, \quad \text{presque sûrement.}$$

4.6.4.4. File (e)

Nous considérons maintenant une file G/G/1/1+G(e)-FIFO. Dans ce cas, un client peut arriver au bout de sa patience et donc être perdu, alors qu'il est en service. Un tel client contribue alors à la charge de travail du serveur, en lui apportant une quantité de

travail correspondant au temps qu'il passera en service avant la fin de sa patience. Plus précisément, la charge de travail apportée par le client C_n à son arrivée est donnée par :

$$\left\{ \begin{array}{l} \sigma \circ \theta^n, \quad \text{si } W_n \leq (D \circ \theta^n - \sigma \circ \theta^n)^+; \\ \sigma \circ \theta^n - (W_n + \sigma \circ \theta^n - D \circ \theta^n) = D \circ \theta^n - W_n, \\ \quad \text{si } (D \circ \theta^n - \sigma \circ \theta^n)^+ < W_n \leq D \circ \theta^n; \\ 0, \quad \text{si } W_n > D \circ \theta^n. \end{array} \right.$$

Dans le premier cas, la patience de C_n dure au-delà de la fin de son service. Dans le deuxième, elle se termine alors que C_n est en service et celui-ci reste en service pour une durée égale à $D \circ \theta^n - W_n$. Finalement, dans le troisième cas C_n n'a pas le temps d'atteindre le serveur avant la fin de sa patience. L'équation dynamique précédente peut être reformulée dans une forme plus compacte. Pour tout $n \in \mathbf{N}$, partant d'une charge W_0 à l'arrivée de C_0 :

$$W_{n+1} = \left[W_n + \left(\sigma \circ \theta^n - (W_n + \sigma \circ \theta^n - D \circ \theta^n)^+ \right) - \xi \circ \theta^n \right]^+. \quad (4.113)$$

Par conséquent, une charge de travail stationnaire sur l'espace de Palm est une variable aléatoire S à valeurs dans \mathbf{R}_+ qui résout l'équation :

$$S \circ \theta = \psi(S) = \left[S + \left(\sigma - (S + \sigma - D)^+ \right) - \xi \right]^+. \quad (4.114)$$

THÉORÈME 4.44.– Il existe une solution presque sûrement finie S à (4.114), telle que presque sûrement :

$$S \leq Y^{D, \sigma}, \quad (4.115)$$

où $Y^{D, \sigma}$ est défini comme en (4.35). De plus, cette solution est unique dès que :

$$\mathbf{P} \left(\sup_{j \in \mathbf{N}^*} \left(D \circ \theta^{-j} - \sum_{i=1}^j \xi \circ \theta^{-i} \right) \leq 0 \right) > 0. \quad (4.116)$$

Démonstration. Il est facile de vérifier que l'application ψ est presque sûrement croissante et continue. On peut donc appliquer le théorème de Loynes et conclure à l'existence d'une solution minimale S à (4.114), limite presque sûre de la suite de Loynes correspondante.

D'après le théorème 4.19, l'unique solution $Y^{D, \xi}$ presque sûrement finie à l'équation :

$$Y \circ \theta = F^{D, \xi}(Y)$$

est donnée par :

$$Y^{D,\xi} = \left[\sup_{j \in \mathbf{N}^*} \left(D \circ \theta^{-j} - \sum_{i=1}^j \xi \circ \theta^{-i} \right) \right]^+.$$

En notant maintenant que pour tout $x \in \mathbf{R}^+$, presque sûrement :

$$\begin{aligned} \psi(x) &= [((x + \sigma) \wedge D) \mathbf{1}_{\{[0, D]\}}(x) + x \mathbf{1}_{\{(D, \infty)\}}(x) - \xi]^+ \\ &\leq [(x \vee D) \wedge (x + \sigma \mathbf{1}_{\{[0, D]\}}(x)) - \xi]^+ \\ &= \varphi(x) \wedge F^{D,\xi}(x), \end{aligned} \quad (4.117)$$

on voit clairement que l'événement $\{S \leq Y_{D,\xi}\}$ est θ -contractant. D'autre part, on a $\mathbf{P}(S \leq D) > 0$, puisque le contraire impliquerait que $S \circ \theta = S - \xi$ presque sûrement, une contradiction avec le lemme 2.2. Mais sur l'événement $\{S \leq D\}$:

$$S \circ \theta = [((S + \sigma) \wedge D) - \xi]^+ \leq [D \vee Y_{D,\xi} - \xi]^+ = Y_{D,\xi} \circ \theta.$$

Ceci implique donc que presque sûrement :

$$S \leq Y_{D,\xi} < \infty.$$

Finalement, pour toute solution S' de (4.114), $\{S = S'\}$ est θ -invariant. En vertu de la minimalité de S et de (4.115), cet événement inclut $\{Y_{D,\xi} = 0\}$. Il est donc presque-sûr dès que (4.116) est vraie. \square

Les preuves des deux lemmes suivants suivent les mêmes principes et sont laissées au lecteur.

LEMME 4.45.– Sous la condition (4.116), si une variable aléatoire Z satisfait presque sûrement l'inégalité $Z \leq Y^{D,\xi}$, alors il y a couplage arrière fort entre $(W_n^Z, n \in \mathbf{N})$ et $(S \circ \theta^n, n \in \mathbf{N})$. Par ailleurs, si la condition (4.90) est satisfaite, on a presque sûrement :

$$S \leq W, \quad (4.118)$$

où W désigne l'unique solution de (4.103) et :

$$S \geq Y^{\sigma \wedge D, \xi}, \quad (4.119)$$

où $Y^{\sigma \wedge D, \xi}$ est définie comme en (4.35).

4.6.5. Probabilité de perte

La probabilité de perte à l'équilibre est une caractéristique cruciale des performances du système. Celle-ci peut être définie intuitivement sur l'espace de Palm des arrivées, services et délais comme la proportion asymptotique de clients perdus dans un système stable.

Considérons tout d'abord la file $G/G/1/G(b) - FIFO$ et supposons que la condition de stabilité (4.90) est vérifiée. Pour tout $n \geq 0$, le client C_n a une patience de $D \circ \theta^n$ et se voit proposer un temps d'attente $W \circ \theta^n$ si C_0 avait trouvé à son entrée une charge de travail W . Donc C_n est perdu si et seulement si $W \circ \theta^n > D \circ \theta^n$. Par conséquent, si l'on note $\pi(b)$ la probabilité de perte pour ce système et :

$$B = \{(x, y) \in (\mathbf{R}^+)^2; x > y\},$$

on a \mathbf{P} -presque sûrement :

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^n \mathbf{1}_{\{B\}}((W, D) \circ \theta^i) = \mathbf{E}[\mathbf{1}_{\{B\}}((W, D))] = \mathbf{P}(W > D). \quad (4.120)$$

D'après (4.35), (4.104) et (4.108), on a donc :

$$\begin{aligned} \mathbf{P} \left(\sup_{j \in \mathbf{N}^*} \left((\sigma \wedge D) \circ \theta^{-j} - \sum_{i=1}^j \xi \circ \theta^{-i} \right) > D \right) \\ \leq \pi(b) \leq \mathbf{P} \left(\sup_{j \in \mathbf{N}^*} \left((\sigma + D) \circ \theta^{-j} - \sum_{i=1}^j \xi \circ \theta^{-i} \right) > D \right). \end{aligned} \quad (4.121)$$

Revenons maintenant au système $G/G/1/G(e) - FIFO$. Le client C_n est perdu (son service n'arrive pas à son terme) si son temps de séjour proposé $W_n + \sigma \circ \theta^n$ dépasse sa patience $D \circ \theta^n$. La probabilité de perte stationnaire $\pi(e)$ est donc donnée, pour la même raison que précédemment, par :

$$\pi(e) = \mathbf{P}(S > D - \sigma).$$

D'après (4.115) et (4.119), on a donc :

$$\begin{aligned} \mathbf{P} \left(\sup_{j \in \mathbf{N}^*} \left((\sigma \wedge D) \circ \theta^{-j} - \sum_{i=1}^j \xi \circ \theta^{-i} \right) > D - \sigma \right) \\ \leq \pi(e) \leq \mathbf{P} \left(\sup_{j \in \mathbf{N}^*} \left(D \circ \theta^{-j} - \sum_{i=1}^j \xi \circ \theta^{-i} \right) > D - \sigma \right). \end{aligned} \quad (4.122)$$

D'autre part, la probabilité stationnaire $\hat{\pi}(e)$ qu'un client ne puisse même pas atteindre le serveur est donnée par :

$$\hat{\pi}(e) = \mathbf{P}(S > D).$$

D'après (4.118), on a donc :

$$\hat{\pi}(e) \leq \pi(b).$$

4.7. Notes et commentaires

L'approche que nous proposons pour l'étude de stabilité, reposant sur l'étude des suites récurrentes stochastiques, rejoint en de nombreux points celle introduite et développée dans [BAC 02].

Le théorème de Loynes pour la file à un serveur a été introduit par Loynes dans [LOY 62]. La preuve de l'optimalité de SRPT qui est présentée ici est due à Flipo [FLI 81]. L'optimalité de FIFO a été traitée sous des angles différents par plusieurs auteurs. Nous présentons ici une preuve proche de celle de [FOS 81]. L'optimalité de EDF sans perte a été démontré dans le même cadre dans [MOY 08a]. L'argument d'échange est démontré page 267 de [BAC 02]. La représentation des systèmes *Processor sharing* et à une infinité de serveurs par des suites de profils est dûe à [MOY 08b]. La construction de l'état stationnaire du système à S files en parallèle est dûe à Neveu [NEV 84]. La construction de la charge de travail stationnaire de la file FIFO avec clients impatientes suit la représentation de [BAC 84]. Elle peut être trouvée sous cette forme dans [MOY 10]. L'optimalité d'EDF pour la file non conservative a été formulée dans des termes comparables dans [PAN 88]. La preuve présentée ici est originale.

Résumé

- Dans une file à un seul serveur, la condition de stabilité est $\rho < 1$ où ρ est la charge.
- Dans une file à S serveurs, cette condition devient $\rho < S$.
- Un système à S files en parallèle, où chaque client rejoint la file la plus courte du point de vue de la charge de travail, est équivalente à une file FIFO avec S serveurs. Elle a donc la même condition de stabilité.
- Des paquets de longueur déterministes minimisent le délai moyen d'attente.
- La discipline de service SRPT est celle qui minimise les temps d'attente parmi toutes les disciplines conservatives.
- La discipline FIFO minimise dans le cas indépendant les temps d'attente et de séjour parmi les disciplines indépendantes des temps de service.
- Le système GI/GI à une infinité de serveurs est stable si $\mathbf{P}(\sigma \leq \xi) > 0$.
- La file GI/GI/1 avec clients impatientes est stable si $\mathbf{P}(\sigma + D \leq \xi) > 0$, où D désigne le temps de patience.
- La discipline EDF est optimale pour le retard dans le cas conservatif et pour la perte dans le cas non conservatif.

Chapitre 5

La file M/GI/1

Au rayon des files à un serveur que l'on sait étudier totalement, après la file M/M/1 que nous étudierons dans le chapitre 8, les suivantes en terme de généralités sont les files GI/M/1 et M/GI/1, c'est-à-dire que les temps inter-arrivées (respectivement les temps de service) sont indépendants et identiquement distribués mais pas de loi exponentielle. Dans ce cas malheureusement, le processus « nombre de clients dans le système à l'instant t » n'est plus markovien. En effet, à un instant donné, on ne peut pas reproduire le raisonnement de l'exemple 7.1 puisque seule la loi exponentielle satisfait le paradoxe de l'autobus (voir les théorèmes [6.7] et [7.3]).

Pour contourner la difficulté, le système est observé à temps discret à des instants convenablement choisis, en l'occurrence pour la file M/GI/1, juste après le départ de chaque client. Pour garder ce chapitre assez simple, nous nous contentons d'étudier la chaîne incluse qui donne les résultats les plus importants.

5.1. Le nombre de clients dans la file

Soit X_n comme étant le nombre de clients dans le système immédiatement après le départ du client n . Résumons les principaux résultats. On rappelle que les arrivées forment un processus de Poisson d'intensité λ et que les temps de service sont indépendants et de même loi \mathbf{P}_σ , de moyenne $1/\mu$. On a posé $\rho = \lambda/\mu$ et σ_n est le temps de service du client n . On a déjà vu dans l'introduction que X_n le nombre de clients dans le système après le départ du client n satisfait l'équation

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

où A_n est le nombre de clients qui arrivent pendant le service du client n .

THÉORÈME 5.1.— La suite $(A_n, n \geq 0)$ est une suite de variables aléatoires indépendantes et identiquement distribuées de loi donnée par

$$\mathbf{P}(A_n = k) = \int_0^\infty \exp(-\lambda t) \frac{(\lambda t)^k}{k!} d\mathbf{P}_\sigma(t).$$

En particulier,

$$\mathbf{E}[A_n] = \rho = \lambda/\mu.$$

Démonstration. Posons T'_n l'instant de départ du client n et X le processus à temps continu comptant le nombre de clients présents dans le système. Le processus X est donc adapté par rapport à la filtration

$$\mathcal{G}_t = \sigma\{N(u), u \leq t\} \vee \sigma\{\sigma_n, n \geq 1\}.$$

Comme le processus de Poisson N est indépendant des temps de service, N reste un processus de Poisson d'intensité λ par rapport à \mathcal{G} , i.e., le processus $(t \mapsto N(t) - \lambda t)$ est une \mathcal{G} martingale. Comme on peut écrire,

$$T'_1 = \inf\{t > 0, \Delta X(t) = -1\} \text{ et } T'_n = \inf\{t > T'_{n-1}, \Delta X(t) = -1\},$$

les variables aléatoires T'_n sont des \mathcal{G} -temps d'arrêt. De plus, ils sont tous construits « de la même manière » :

$$T'_n = T'_{n-1} + T'_1 \circ \theta^{T'_{n-1}}.$$

D'après la propriété de Markov forte, la suite

$$(T'_1, T'_2 - T'_1, \dots, T'_n - T'_{n-1})$$

est une suite de variables aléatoires indépendantes, identiquement distribuées. Par construction, A_1 est le nombre d'arrivées pendant le premier temps de service, qui dure par définition σ_1 unités de temps. Pour calculer A_1 , on conditionne par la valeur de σ_1 ; comme N indépendant de σ_1 , on a

$$\begin{aligned} \mathbf{P}(A_1 = k) &= \mathbf{E}[\mathbf{E}[\mathbf{1}_k(A_1) | \sigma_1]] \\ &= \mathbf{E}[\mathbf{E}[\mathbf{1}_k(N(\sigma_1)) | \sigma_1]] \\ &= \mathbf{E}[\mathbf{P}(N(\sigma_1) = k) | \sigma_1] \\ &= \mathbf{E}\left[\exp(-\lambda\sigma_1) \frac{(\lambda\sigma_1)^k}{k!}\right] \\ &= \int_0^\infty \exp(-\lambda t) \frac{(\lambda t)^k}{k!} d\mathbf{P}_\sigma(t). \end{aligned}$$

Pour calculer l'espérance de A_1 , on utilise sa définition, d'où

$$\begin{aligned} \mathbf{E}[A_1] &= \sum_{k=0}^{\infty} k \mathbf{P}(A_1 = k) \\ &= \int_0^{\infty} \exp(-\lambda t) \sum_{k=0}^{\infty} k \frac{(\lambda t)^k}{k!} d\mathbf{P}_{\sigma}(t) \\ &= \lambda \int_0^{\infty} t d\mathbf{P}_{\sigma}(t) \\ &= \rho. \end{aligned}$$

□

D'après le théorème 4.2, on connaît déjà les conditions de stabilité du système : il est stable, dans le sens où il existe une charge de travail et un nombre de clients stationnaire, si et seulement si la charge $\rho = \lambda/\mu$ est strictement inférieure à 1. Comme X est une chaîne de Markov, on peut préciser le mode de récurrence.

THÉORÈME 5.2.– Posons $a_k = \mathbf{P}(A_n = k)$ et $\rho = \mathbf{E}[A_n]$. Supposons que $0 < a_0 \leq a_0 + a_1 < 1$.

- 1) La chaîne X est transiente si et seulement si $\rho > 1$,
- 2) X est récurrente nulle si et seulement si $\rho = 1$,
- 3) X est récurrente positive si et seulement si $\rho < 1$.

Démonstration. Il est évident que X est irréductible puisque il décrit l'ensemble \mathbf{N} par incréments de 1 et de -1, grâce aux conditions sur a_0 et a_1 qui garantissent qu'il y ait de temps en temps au moins un départ.

Supposons que $\rho > 1$. Soit $X_0 = l > 0$. Tant que X n'a pas décré de l unités, c'est-à-dire tant que n est inférieur à τ_0^1 , l'évolution de X est donnée par

$$X_n = l + (A_1 - 1) + \dots + (A_n - 1).$$

Posons $\hat{X}_n = \sum_{j=1}^n (A_j - 1)$ et $\hat{\tau}_0 = \inf\{n > 0, \hat{X}_n = 0\} = \tau_1^1$. Comme $\rho > 1$, d'après la loi forte des grands nombres,

$$\frac{\hat{X}_n}{n} \longrightarrow \rho - 1 \quad \mathbf{P}_i \text{ p.s. pour tout } i.$$

Par conséquent, pour presque toutes les trajectoires ω , il existe $N(\omega)$ tel que

$$n \geq N(\omega) \implies 0 < (\rho - 1)/2 \leq \frac{\hat{X}_n}{n} \leq 3(\rho - 1)/2. \quad (5.1)$$

Par conséquent, l'atteinte de 0 doit se faire avant $N(\omega)$. Pour conclure, montrons qu'il existe une valeur l_0 de l pour laquelle

$$0 < \alpha = \mathbf{P}(\omega : N(\omega) < l_0) < 1. \quad (5.2)$$

Comme X ne peut décroître au maximum que d'une unité par pas de temps, si $X_0 = l_0$ et $N(\omega) < l_0$ alors $X_1(\omega) > 0, \dots, X_{N(\omega)}(\omega) > 0$ et d'après (5.1), $X_n(\omega) > 0$ pour tout $n > N(\omega)$. Par conséquent, partant de 0, comme la chaîne est irréductible, il y a une probabilité non nulle que l'on atteigne X_0 , partant de ce point avec probabilité α on ne retournera jamais en 0 donc 0 est transient, ce qui par irréductibilité implique que la chaîne est transiente.

Montrons maintenant (5.2). Si pour tous les $i \geq 0$, $N(\omega)$ est p.s. inférieur à i , cela signifie que $N(\omega)$ est p.s. nul donc que $X_n \geq X_0$ pour toute valeur de n , ce qui est contraire à l'hypothèse d'irréductibilité. Si pour tous les $i \geq 0$, $N(\omega)$ est p.s. supérieur à i , cela signifie que $N(\omega)$ est p.s. infini, ce qui est contraire à la convergence de \hat{X}_n/n .

Supposons $\rho \leq 1$. Prenons pour h la fonction identité et $F = \{0\}$ dans les critères de Foster. Comme $X_n = f(X_0, A_1, \dots, A_n)$ est indépendant de A_{n+1} , pour $i > 0$,

$$\begin{aligned} \mathbf{E}[X_{n+1} | X_n = i] &= \mathbf{E}[i + A_{n+1} - 1 | X_n = i] \\ &= i + \mathbf{E}[A_{n+1}] - 1 \\ &= i + \rho - 1, \end{aligned}$$

donc d'après les critères de Foster, la chaîne est récurrente nulle ssi $\rho = 1$ et récurrente positive si et seulement si $\rho < 1$. \square

5.2. Formules de Pollacek-Khinchin

Le calcul de la probabilité stationnaire dans le cas $\rho < 1$ se fait par une méthode subtile mais fréquemment utilisée, celle de la transformée en z ou probabilistiquement parlant, par le calcul de la fonction génératrice. On sait déjà que la probabilité stationnaire existe puisque d'après les critères de Foster la chaîne est récurrente positive. Soit π cette probabilité stationnaire et soit X_∞ , une variable aléatoire de loi π .

THÉORÈME 5.3.— Soit π la probabilité invariante d'une file M/GI/1 de charge $\rho < 1$. On a l'identité suivante, appelée formule de Pollacek-Khinchin :

$$\sum_k z^k \pi(k) = \frac{(z-1)Q_A(z)}{z - Q_A(z)}(1 - \rho), \quad (5.3)$$

où \mathcal{Q}_A est la fonction génératrice de A_1 , i.e. pour tout $\mathcal{Q}_A(z) = \mathbf{E}[z^{A_1}]$ pour tout z .

Démonstration. Posons

$$\mathcal{Q}_{X_\infty}(z) = \mathbf{E}[z^{X_\infty}] = \sum_k z^k \pi(k), \text{ pour } |z| \leq 1.$$

Si $\rho < 1$, on sait que X_n converge en loi vers X_∞ , donc que

$$\lim_{n \rightarrow \infty} \mathbf{E}[z^{X_n}] = \mathcal{Q}_{X_\infty}(z).$$

$$\begin{aligned} \mathcal{Q}_{X_n}(z) &= \mathbf{E}[z^{X_{n+1}}] \\ &= \mathbf{E}[z^{A_{n+1} z^{(X_n-1)^+}}] \\ &= \mathbf{E}[z^{A_{n+1}}] \left(\mathbf{E}[z^{X_n-1} \mathbf{1}_{\{X_n > 0\}}] + \mathbf{P}(X_n = 0) \right) \text{ et} \\ \mathbf{E}[z^{X_n-1} \mathbf{1}_{\{X_n > 0\}}] &= \frac{1}{z} \left(\mathbf{E}[z^{X_n}] - \mathbf{E}[z^{X_n} \mathbf{1}_{\{X_n=0\}}] \right) \\ &= \frac{1}{z} \mathbf{E}[z^{X_n}] - \frac{1}{z} \mathbf{P}(X_n = 0). \end{aligned}$$

Notons $\mathcal{Q}_A(z) = \mathbf{E}[z^{A_1}]$, en passant à la limite on obtient

$$\mathcal{Q}_{X_\infty}(z) = \frac{(z-1)\mathcal{Q}_A(z)}{z - \mathcal{Q}_A(z)} \pi(0). \quad (5.4)$$

Au voisinage de 1,

$$\begin{aligned} \mathcal{Q}_A(z) &= 1 + (z-1)\mathbf{E}[A_1] + o(z-1) \\ &= 1 + \rho(z-1) + o(z-1) \text{ et} \\ \mathcal{Q}_{X_\infty}(z) &= 1 + O(z-1), \end{aligned}$$

donc le terme de droite de (5.4) doit d'une part tendre vers 1 et est d'autre part équivalent à $(1-\rho)\pi(0)$, soit $\pi(0) = 1-\rho$. Au final,

$$\mathcal{Q}_{X_\infty}(z) = \frac{(z-1)\mathcal{Q}_A(z)}{z - \mathcal{Q}_A(z)} (1-\rho).$$

Comme $\mathcal{Q}_{X_\infty}(z)$ est une fonction développable en série entière, on a

$$\mathcal{Q}_{X_\infty}(z) = 1 + \sum_{k=1}^{\infty} \frac{\mathcal{Q}_{X_\infty}^{(k)}(0)}{k!} z^k, \quad (5.5)$$

il suffit donc en théorie de dériver k fois \mathcal{Q}_{X_∞} en 0 pour obtenir $\pi(k)$. \square

Nous souhaitons calculer \mathcal{Q}_A en fonction de la loi des temps de service et de l'intensité du processus de Poisson. Pour $z \geq 0$, posons

$$\mathcal{L}_\sigma(z) = \int_0^\infty e^{-zt} d\mathbf{P}_\sigma(t),$$

c'est-à-dire que \mathcal{L}_σ est la transformée de Laplace-Stieltjes de la loi des temps de service.

LEMME 5.4.–

$$\mathcal{Q}_A(z) = \mathcal{L}_\sigma(\lambda - \lambda z) \text{ pour tout } z \in [0, 1].$$

Démonstration. Par définition, $\mathcal{Q}_A(z) = \mathbf{E}[z^{N_\sigma}]$, d'où l'on tire :

$$\begin{aligned} \mathcal{Q}_A(z) &= \mathbf{E}[\mathbf{E}[z^{N_\sigma} | \sigma]] \\ &= \int_0^\infty \mathbf{E}[z^{N_u} | \sigma = u] d\mathbf{P}_\sigma(u) \\ &= \int_0^\infty \mathbf{E}[z^{N_u}] d\mathbf{P}_\sigma(u) \\ &= \int_0^\infty \exp(-(1-z)\lambda u) d\mathbf{P}_\sigma(u) \\ &= \mathcal{L}_\sigma(\lambda - \lambda z). \end{aligned}$$

□

LEMME 5.5 (Deuxième équation de Pollacek-Khinchin).– Si $\rho < 1$, le nombre moyen de clients à l'état stationnaire est donné par :

$$\mathbf{E}[X_\infty] = \rho + \frac{\rho^2(1 + C_b^2)}{2(1 - \rho)} \text{ où } C_b^2 = \frac{\text{Var}[\sigma]}{\mathbf{E}[\sigma]^2}, \quad (\text{P-K } 2)$$

avec σ une variable aléatoire de loi \mathbf{P}_σ .

Démonstration. Il est supposé su que $S'(1) = \mathbf{E}[X_\infty]$. Par ailleurs, d'après le théorème de dérivation sous le signe intégrale,

$$\frac{d\mathcal{L}_\sigma}{ds}(s) = - \int_0^\infty x e^{-sx} d\mathbf{P}_\sigma(x),$$

d'où il apparaît que

$$\frac{d\mathcal{Q}_A}{dz}(1) = \rho.$$

Le calcul complet de la dérivée donne (P-K 2).

□

REMARQUE.– Dans la file M/M/1, on a $C_b = 1$ et l'on retrouve ainsi le résultat connu : $\mathbf{E}[X_\infty] = \rho(1 - \rho)^{-1}$.

EXEMPLE (Optimalité du déterminisme).– On retrouve clairement avec (P-K 2) qu'à charge ρ fixée, le nombre moyen de clients dans le système à l'équilibre est minimisé pour $\text{Var}[\sigma] = 0$, i.e. pour un temps de service déterministe.

D'après la formule de Little, on minimise également le délai moyen d'attente à l'équilibre, et c'est essentiellement pour cette raison que l'on privilégiera, dans les réseaux de paquets, les messages de taille fixe.

EXEMPLE (La file M/G/1).– On note ainsi la file dans laquelle les temps de service suivent une loi Γ , c'est-à-dire lorsque l'on a

$$d\mathbf{P}_\sigma(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} \mathbf{1}_{\mathbf{R}^+}(x) dx$$

où α et β sont strictement positifs. Un rapide calcul montre que

$$\mathbf{E}[\sigma] = \alpha/\beta \text{ et } \text{Var}[\sigma] = \alpha/\beta^2 \text{ donc } C_b^2 = 1/\alpha.$$

On s'aperçoit alors que lorsque α devient petit, β choisi de manière telle que le rapport α/β soit constant, la charge est constante mais le nombre moyen de clients tend vers l'infini. Comme $\text{Var}[\sigma]$ peut s'écrire aussi $\mathbf{E}[\sigma]/\beta$, α petit induit β petit donc une variance des temps de service très grande. Compte-tenu de ce résultat et du précédent sur la file M/D/1, on voit que l'on a tout intérêt à la limiter les fluctuations des temps de service.

Les lois pour lesquelles C_b^2 est supérieur à 1 sont dites sur-variantes. Dans le cas de la loi Gamma, cette sur-variance est due au fait que la densité tendant vers 0 à l'infini moins vite qu'une exponentielle, la probabilité d'avoir des temps de service longs est bien plus élevé que dans la file M/M/1. Ce phénomène s'avère crucial et assez fréquent puisqu'il a été observé statistiquement que sur la toile, la longueur des fichiers suivait une loi de Pareto, c'est-à-dire de la forme :

$$d\mathbf{P}_\sigma(x) = x^{-\alpha} \mathbf{1}_{[K, +\infty[}(x) dx$$

où $K > 0$ représente la longueur minimale des fichiers et $\alpha > 0$ le taux de décroissance à l'infini. La situation est encore plus dramatique dans ce cas que dans celui de la loi Gamma, puisque si $\alpha < 2$, la variance de σ est infini.

REMARQUE.– En utilisant la formule de Little, on montre aisément que

$$\frac{\mathbf{E}[\text{TA}]}{\mathbf{E}[X]} = \frac{\rho(1 + C_b^2)}{2(1 - \rho)},$$

formule où apparaît une dépendance linéaire entre le temps d'attente moyen et le coefficient de variation pour une charge fixée du serveur.

5.3. Temps de séjour

Pour calculer la loi du temps de séjour dans le système, nous partons de l'observation suivante : étant donnée que la politique de service est de type FIFO, au moment où le client quitte le système, il laisse derrière lui tous les clients arrivés pendant son séjour dans celui-ci. Par conséquent la loi du nombre de clients aux instants de départ coïncide avec celle du nombre d'arrivées pendant les temps de de séjour :

$$N(\text{TS}_n) = X_n, \quad (5.6)$$

où TS_n est le temps de séjour du client n . De même, comme le temps de séjour ne dépend que des clients qui sont arrivés avant le client n , TS_n est indépendant du nombre des arrivées qui se produisent pendant le séjour du client n . Connaissant la distribution aux instants de départ par la formule de Pollaczek-Khinchin, nous pouvons en déduire celle du temps de séjour. Notons que l'on peut décomposer TS_n en la somme du temps d'attente TA_n dans la file plus le temps de service σ_n .

THÉORÈME 5.6.— Si $\rho < 1$ alors les variables aléatoires TS_n et TA_n convergent en loi vers des variables aléatoires notées respectivement TS et TA . De plus, pour tout $s \geq 0$,

$$\begin{aligned} \mathcal{L}_{\text{TS}}(s) &= \mathbf{E} [e^{-s\text{TS}}] = \mathcal{L}_\sigma(s) \frac{s(1-\rho)}{s-\lambda + \lambda\mathcal{L}_\sigma(s)} \\ \mathcal{L}_{\text{TA}}(s) &= \mathbf{E} [e^{-s\text{TA}}] = \frac{s(1-\rho)}{s-\lambda + \lambda\mathcal{L}_\sigma(s)}. \end{aligned}$$

Démonstration. En utilisant la relation (5.6), on a

$$\begin{aligned} \mathcal{Q}_{X_n}(z) &= \mathbf{E} [\mathbf{E} [z^{N_{[T_n, T_n + \text{TS}_n[}} | \text{TS}_n]]] \\ &= \int_0^\infty \mathbf{E} [z^{N_{[T_n, T_n + u[}} | \text{TS}_n = u] d\mathbf{P}_{\text{TS}_n}(u) \\ &= \int_0^\infty \mathbf{E} [z^{N_{[T_n, T_n + u[}}] d\mathbf{P}_{\text{TS}_n}(u) \\ &= \int_0^\infty \mathbf{E} [z^{N^u}] d\mathbf{P}_{\text{TS}_n}(u) \\ &= \int_0^\infty \exp(-(1-z)\lambda u) d\mathbf{P}_{\text{TS}_n}(u) \\ &= \mathcal{L}_{\text{TS}_n}(\lambda - \lambda z). \end{aligned}$$

Comme X_n converge en loi vers X_∞ , \mathcal{Q}_{X_n} converge simplement vers $\mathcal{Q}_{X_\infty}(z)$ donc $\mathcal{L}_{\text{TS}_n}$ a aussi une limite simple ce qui est équivalent à dire que TS_n converge en loi vers TS . Comme $\text{TA}_n = \text{TS}_n - \sigma_n$, et que la loi de σ_n est constante, on en déduit que

TA_n converge en loi, vers une variable aléatoire notée TA. Compte-tenu de (5.3), on en déduit que

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

Comme TA_n ne dépend que de ce qui s'est passé strictement avant le client n , TA_n et σ_n sont indépendants donc

$$\mathbf{E} [e^{-sT_S n}] = \mathbf{E} [e^{-sTA_n}] \mathbf{E} [e^{-s\sigma_n}],$$

d'où l'on tire la formule désirée pour \mathcal{L}_{TA} . □

5.4. Distribution de queue du temps d'attente

Lorsque l'on veut dimensionner un buffer, on a a priori besoin de connaître les probabilités de perte pour chaque valeur possible du dit buffer. Dans les cas autres que celui de la M/M/1, ces quantités sont inaccessibles analytiquement, en revanche dans le cas de la file M/GI/1 on peut accéder la queue de la distribution du temps d'attente virtuel, c'est-à-dire $\mathbf{P}(TA > x)$ pour x grand. De là, on en déduit un dimensionnement en choisissant x de telle sorte que cette probabilité de dépassement de seuil soit inférieure au taux de perte toléré. Il est possible qu'en pareil cas, on surdimensionne (parfois de beaucoup) le buffer.

LEMME 5.7.– Pour tout $s > 0$,

$$\mathbf{E} [e^{-sTA}] = 1 - \int_0^\infty \mathbf{P}(TA > x) s e^{-sx} dx.$$

Démonstration. D'après le théorème de Fubini, on a :

$$\begin{aligned} \int_0^\infty \mathbf{P}(TA > x) s e^{-sx} dx &= \int_0^\infty \left(\int_x^\infty d\mathbf{P}_{TA}(u) \right) s e^{-sx} dx \\ &= \int_0^\infty \left(\int_0^u s e^{-sx} dx \right) d\mathbf{P}_{TA}(u) \\ &= \int_0^\infty (1 - e^{-su}) d\mathbf{P}_{TA}(u) \\ &= 1 - \mathbf{E} [e^{-sTA}]. \end{aligned}$$

□

Autrement dit, si l'on note $TA^c(x) = \mathbf{P}(TA > x)$, on a

$$\mathcal{L}_{TA^c}(s) = \frac{1}{s} (1 - \mathcal{L}_{TA}(s)).$$

L'étude asymptotique de $\mathcal{L}_{\text{TA}^c}$ permet d'étudier le comportement de la queue de distribution de TA.

EXEMPLE (File M/PH/1).— Considérons qu'une proportion p des requêtes puissent être traitées localement en un temps de loi exponentielle de paramètre μ_1 et que les autres requêtes aient besoin d'un traitement distant qui s'effectue en un temps exponentiel de paramètre μ_2 . On dit alors que le temps de service est de type « phase », d'où le nom de la file M/PH/1. On a alors :

$$1/\mu = \mathbf{E}[\sigma_1] = p/\mu_1 + (1-p)/\mu_2 \text{ et } \text{Var}[G_1] = 2p/\mu_1^2 + 2(1-p)/\mu_2^2.$$

De même,

$$\mathcal{L}_\sigma(s) = \mathbf{E}[e^{-s\sigma_1}] = p \frac{\mu_1}{\mu_1 + s} + (1-p) \frac{\mu_2}{\mu_2 + s}.$$

Tous calculs faits, on trouve

$$\mathcal{L}_{\text{TA}_\infty^c}(s) = \frac{(\rho \mu_2 + s\rho - 1 + \rho \mu_1) \lambda}{\mu_1 \mu_2 + \mu_2 s - \mu_2 \lambda \mu_1 \rho + s^2 - \lambda s + \mu_1 s}.$$

Le dénominateur s'annule en deux points réels :

$$\alpha_+ = -\frac{1}{2}(\mu_1 + \mu_2 - \lambda + \sqrt{\Delta})$$

$$\alpha_- = -\frac{1}{2}(\mu_1 + \mu_2 - \lambda - \sqrt{\Delta}) \text{ avec}$$

$$\Delta = (\mu_1 + \mu_2 - \lambda)^2 - 4\mu_1\mu_2(1-\rho).$$

On a donc la décomposition en éléments simples suivante :

$$\mathcal{L}_{\text{TA}_\infty^c}(s) = \frac{\rho}{2} \left(\frac{\mu_1 + \mu_2 - \lambda + \sqrt{\Delta}}{s - \alpha_-} + \frac{\mu_1 + \mu_2 - \lambda - \sqrt{\Delta}}{s - \alpha_+} \right).$$

REMARQUE.— Il n'est pas absolument clair que si $\mu_1 + \mu_2 - \lambda$ est négatif alors α_+ est négatif. En fait en utilisant la condition $\rho = \lambda(p/\mu_1 + (1-p)/\mu_2) < 1$, on montre qu'automatiquement $\mu_1 + \mu_2 - \lambda$ est positif.

En utilisant les formules d'inversion de Laplace on trouve :

$$\mathbf{P}(\text{TA}_\infty > x) = \frac{\rho}{2} \left((\mu_1 + \mu_2 - \lambda + \sqrt{\Delta}) e^{\alpha_- x} + (\mu_1 + \mu_2 - \lambda - \sqrt{\Delta}) e^{\alpha_+ x} \right).$$

En fait quand x devient grand, seul compte le terme qui a la décroissance la plus lente, c'est-à-dire le terme avec l'exponentielle contenant α_- . On a donc

$$\mathbf{P}(\text{TA}_\infty > x) \simeq \frac{\rho}{2} (\mu_1 + \mu_2 - \lambda + \sqrt{\Delta}) e^{\alpha_- x}.$$

Ce résultat est à comparer avec celui de la file M/M/1 qui donne

$$\mathbf{P}(T_{A_\infty} > x) \simeq \rho e^{-\mu(1-\rho)x}.$$

Fixons l'unité de temps de telle sorte que $\lambda = 1$ et fixons ρ , ce qui revient à fixer le temps moyen de service $1/\mu$. Dans ce cas pour garantir une probabilité de dépassement de seuil inférieure à ϵ , il faut dans la file M/M/1 un seuil supérieur à

$$-\frac{1}{\mu-1} \ln(\mu\epsilon);$$

dans le cas de la file M/PH/1 il faut prendre

$$x \geq \frac{1}{\alpha_-} \ln \left(\mu\epsilon \cdot \frac{2}{\mu_1 + \mu_2 - 1 + \sqrt{\Delta}} \right).$$

Fixer ρ revient à lier les 3 paramètres μ_1 , μ_2 et p mais il reste tout de même deux degrés de liberté que l'on choisit comme étant μ_1 et μ_2 . On voit que le dimensionnement de la file M/PH/1 est déterminé non seulement par ρ mais aussi par le produit et la somme de μ_1 et μ_2 . En d'autres termes, la seule connaissance de la charge ne suffit pas pour dimensionner le buffer et garantir une perte donnée. C'est une situation très différente de celle de la file M/M/1 où la seule connaissance de ρ permet de calculer le seuil.

En supposant toujours que la file d'attente est stable ($\rho < 1$), en régime stationnaire, un cycle de fonctionnement de la file se compose d'une période d'inactivité I qui se termine par l'arrivée d'un client, suivi que d'une période d'activité U qui se termine par le départ du dernier client du système laissant derrière lui un système vide.

5.5. Périodes d'activité

La période I d'inactivité est celle qui sépare l'instant de départ du dernier client de celui de la prochaine arrivée. Compte tenu du caractère sans mémoire de la loi exponentielle, I suit la même distribution que celle de l'inter-arrivée :

$$\mathbf{P}(I \leq t) = 1 - e^{-\lambda t}.$$

Pour analyser la période d'activité U , nous allons prendre comme origine des temps $t = 0$ l'instant d'arrivée du premier client dans un système vide, ou de manière équivalente, le départ d'un client en laissant derrière lui un seul client dans le système. On note dans la suite, $(X(t), t \geq 0)$ le processus à temps continu comptant le nombre de clients dans le système à tout instant.

DÉFINITION 5.1.– On appelle période d'activité élémentaire U la variable aléatoire suivante :

$$U = \begin{cases} \inf \{t > 0, X_t = 0 \mid X_0 = 1\} & \text{si } \rho < 1 \\ \infty & \text{si } \rho \geq 1 \end{cases}$$

Il est possible de généraliser la définition de la période d'activité au cas où l'instant initial coïncide avec le départ d'un client laissant derrière lui n autres paquets.

DÉFINITION 5.2.– On appelle période d'activité de condition initiale n , U_n la variable aléatoire suivante :

$$U_n = \begin{cases} \inf \{t > 0, Y_t = 0 \mid Y_0 = n\} & \text{si } \rho < 1 \\ \infty & \text{si } \rho \geq 1 \end{cases} .$$

Dans ce cas $U_1 \equiv U$.

THÉORÈME 5.8.– La distribution de probabilité de U_n est le produit de convolution d'ordre n de la distribution de U .

Démonstration. La durée de la période d'activité est insensible à la politique de service, pourvu que celle-ci soit *conservative*, i.e. le serveur sert jusqu'au bout tous les clients qui se présentent : elle s'écrit comme la somme des temps de service de tous les clients présents dans le système durant sa durée. Ainsi, pour calculer U_n , nous allons la décomposer de la manière suivante : les n clients présents à l'origine du temps seront appelés les pères.

- 1) Servir le client en tête de file (le premier père),
- 2) servir tous les clients qui sont arrivés pendant le service du père (les fils),
- 3) servir tous les clients qui sont arrivés pendant le service des fils (les petits fils),
- 4) répéter l'étape précédente jusqu'à ce qu'il n'y ait plus de descendants à servir,
- 5) répéter les étapes 1,2,3 et 4 précédentes pour le deuxième père, puis le troisième père, . . . , jusqu'au n ème père.

La durée des étapes 1,2,3 et 4 n'est autre que U . En vertu des hypothèses, les variables aléatoires représentant le temps de service d'un des pères suivi du service de tous ses descendants sont indépendantes et de même loi que U . Donc la loi de la somme

$$U_n = U + U + \dots + U \text{ } n \text{ fois}$$

est la convolution d'ordre n de celle de U . □

THÉORÈME 5.9.– Pour tout $s \geq 0$,

$$\mathcal{L}_U(s) = \mathbf{E} [e^{-sU}] = \mathcal{L}_\sigma [s + \lambda - \lambda \mathcal{L}_U(s)]. \quad (5.7)$$

Démonstration. Nous allons décomposer U de manière similaire à U_n :

- 1) Servir le père, pendant ce service, V fils sont arrivés,
- 2) servir le premier fils suivi de tous ses descendants,
- 3) répéter l'étape précédente pour les $V - 1$ fils restants.

Ainsi

$$U = \sigma_1 + \sum_{i=1}^V \Phi_i,$$

où chaque Φ_i est distribuée suivant la loi de U . Pour calculer la loi de U , on commence par calculer la loi conditionnelle en V et σ_1 , puis nous déconditionnons successivement sur V puis sur σ_1 . Pour tous x et k ,

$$\begin{aligned} \mathbf{E}[e^{-sU} \mid \sigma_1 = x, V = k] &= \mathbf{E}[e^{-s(x + \sum_{i=1}^k \Phi_i)}] \\ &= e^{-sx} \prod_{i=1}^k \mathbf{E}[e^{-s\Phi_i}] \\ &= e^{-sx} [\mathcal{L}_U(s)]^k. \end{aligned}$$

En déconditionnant sur V ,

$$\begin{aligned} \mathbf{E}[e^{-sU} \mid \sigma_1 = t] &= \sum_{k=0}^{\infty} \mathbf{E}[e^{-sU} \mid \sigma_1 = t, V = k] \mathbf{P}(V = k \mid \sigma_1 = x) \\ &= e^{-sx} \sum_{k=0}^{\infty} [\mathcal{L}_U(s)]^k \frac{(\lambda x)^k}{k!} e^{-\lambda x} \\ &= e^{-x[s + \lambda - \lambda \mathcal{L}_U(s)]}. \end{aligned}$$

Enfin en déconditionnant sur σ_1 ,

$$\begin{aligned} \mathcal{L}_U(s) &= \int_0^{\infty} \mathbf{E}[e^{-sU} \mid \sigma_1 = x] d\mathbf{P}_{\sigma}(x) \\ &= \int_0^{\infty} e^{-x[s + \lambda - \lambda \mathcal{L}_U(s)]} d\mathbf{P}_{\sigma}(x). \end{aligned}$$

□

EXEMPLE (File M/M/1).– Dans ce cas, tout est calculable et l'on obtient :

THÉORÈME 5.10.– Dans le cas de la file M/M/1,

$$\mathcal{L}_U(s) = \frac{1}{2\lambda} [(\lambda + \mu + s) - \sqrt{(\lambda + \mu + s)^2 - 4\lambda\mu}].$$

En inversant $\mathcal{L}_U(s)$ et en dérivant on obtient la densité de la période d'activité élémentaire

$$g(t) = \frac{d}{dt}G(t) = \frac{1}{\sqrt{\rho t}} e^{-(\lambda+\mu)t} I_1[2t\sqrt{\lambda\mu}],$$

où $I_k(t)$ est la fonction de Bessel modifiée de première espèce d'ordre k définie par

$$I_k(t) = \sum_{m=0}^{\infty} \frac{\left(\frac{t}{2}\right)^{(k+2m)}}{(k+m)!m!}.$$

Démonstration. Comme

$$\mathcal{L}_\sigma(s) = \frac{\mu}{s + \mu},$$

il s'ensuit que

$$\mathcal{L}_U(s) = \frac{\mu}{s + \lambda - \lambda\mathcal{L}_U(s) + \mu}.$$

D'où

$$\lambda[\mathcal{L}_U(s)]^2 - (\lambda + \mu + s)\mathcal{L}_U(s) + \mu = 0.$$

Cette équation a deux racines

$$\alpha(s) = \frac{1}{2\lambda} [(\lambda + \mu + s) - \sqrt{(\lambda + \mu + s)^2 - 4\lambda\mu};$$

$$\beta(s) = \frac{1}{2\lambda} [(\lambda + \mu + s) + \sqrt{(\lambda + \mu + s)^2 - 4\lambda\mu},$$

où $s \in \mathcal{C}$, $\Re e(s) > 0$ et $\lambda\mu > 0$. D'autre part, $\mathcal{L}_U(s) \in \mathbf{C}$. Le système étant supposé stable la distribution de U n'est pas dégénérée, donc $|\mathcal{L}_U(s)| \leq 1$. Par conséquent, seule les racines de module inférieur ou égal à l'unité conviennent. Nous allons voir maintenant si les deux racines remplissent cette condition.

Notons d'abord que $|\alpha(s)| < |\beta(s)|$ car $\lambda\mu\Re e(s) > 0$, ceci pour tout s dans le demi plan droit du plan complexe. Quand $|\mathcal{L}_U(s)| = 1$ (sur la frontière du disque unité), on a

$$|(\lambda + \mu + s)\mathcal{L}_U(s)| = |(\lambda + \mu + s)| > \lambda + \mu \geq |\mu + \lambda\mathcal{L}_U(s)^2|.$$

Nous avons donc deux fonctions complexes

$$f(z) = -(\lambda + \mu + s)z;$$

$$h(z) = \mu + \lambda z^2,$$

analytiques à l'intérieur et sur la frontière d'un domaine fermé du plan complexe défini par son contour

$$\mathcal{C} = \{z \in \mathcal{C}; |z| = 1\} \cap \{s; \Re e(s) > 0\}$$

et telles que $|f(z)| > |h(z)|$ sur le contour. Le théorème de Rouché permet de dire que la plus grande (en module) des deux fonctions (c'est-à-dire $f(z)$) et la somme des deux fonctions ont le même nombre de zéros dans le domaine défini par le contour \mathcal{C} . Or, dans ce domaine, $f(z)$ a un seul zéro. Donc seule une des deux racines $\alpha(s), \beta(s)$ se trouve dans le domaine délimité par \mathcal{C} . C'est obligatoirement la plus petite des deux $\alpha(s)$, qui convient. Le résultat est démontré. \square

En théorie, il est possible d'obtenir numériquement à partir de l'expression analytique de $g(\cdot)$, la distribution de $G(\cdot)$. Seulement, l'opération n'est pas simple compte tenue de la complexité de la forme de la série ci-dessus. Il est également théoriquement possible de calculer la distribution de U_n en inversant la fonction $(\mathcal{L}_U(s))^n$. Paratiquement, il est très facile de calculer le premier moment de U :

$$\mathbf{E}[U] = \int_0^\infty t \, d\mathbf{P}_\sigma(t) = -\frac{d}{ds} \mathcal{L}_U(s) \Big|_{\{s=0\}},$$

soit

$$\mathbf{E}[U] = \begin{cases} \frac{1}{\mu - \lambda} & \text{si } \rho < 1 \\ \infty & \text{si } \rho \geq 1 \end{cases}.$$

THÉORÈME 5.11.– Si $\rho > 1$,

$$\mathbf{P}(U < \infty) = \frac{1}{\rho} < 1 \text{ et } \mathbf{P}(U = \infty) = 1 - \frac{1}{\rho} > 0$$

Démonstration. Remarquons que

$$\lim_{s \rightarrow 0} \mathcal{L}_U(s) = \lim_{s \rightarrow 0} \int_0^\infty e^{-sx} \, d\mathbf{P}_\sigma(x) = \lim_{x \rightarrow \infty} G(x).$$

Or le second membre tend vers $\mathbf{P}(U < \infty)$. Si la file est stable ($\rho < 1$), l'état 0 est récurrent positif : on y revient presque sûrement au bout d'un temps fini, partant de 0 avec un temps inter-visite moyen fini. Donc

$$\rho < 1 \implies G(\infty) = \mathbf{P}(U < \infty) = 1.$$

Si $\rho = 1$, alors 0 est récurrent nul. La chaîne y revient presque sûrement, mais le temps moyen séparant deux visites est infini. Nous avons toujours $G(\infty) = 1$. Enfin, si $\rho > 1$, alors 0 est transient, et il y a une probabilité non nulle de ne jamais revenir à 0 partant de cet état. D'où

$$\rho > 1 \implies G(\infty) = \mathcal{L}_U(0) = \frac{1}{2\lambda} [\lambda + \mu - \sqrt{(\lambda + \mu)^2 - 4\lambda\mu}] = \frac{\mu}{\lambda},$$

ce qui permet de conclure. \square

DEUXIÈME PARTIE

Modélisation à temps continu

Chapitre 6

Processus de Poisson

La modélisation d'un système physique doit obéir à deux contraintes. D'une part, elle doit refléter aussi fidèlement que possible la « réalité » et d'autre part elle doit avoir un rôle prédictif; autrement dit on doit pouvoir faire des calculs grâce à cette modélisation. Au-delà de la difficulté à déterminer quantitativement et qualitativement les paramètres pertinents d'un système physique, l'expérience prouve que plus l'on voudra un modèle précis moins celui-ci sera utilisable pour les calculs.

Dans le cadre des files d'attente, il nous faut modéliser en premier lieu le processus des arrivées des requêtes. Le processus de Poisson, que nous étudions dans cette section, est le modèle le plus fréquemment utilisé essentiellement parce que c'est l'un des rares avec lequel on peut faire des calculs. Cette modélisation se trouve être très pertinente pour les arrivées des appels téléphoniques à un commutateur, malheureusement il n'en est pas de même pour les autres types de réseaux dans lesquels le trafic est beaucoup plus versatile. Néanmoins, comme nous le verrons à la fin de ce chapitre, le processus de Poisson peut être modifié de façon à refléter dans une certaine mesure cette versatilité.

La définition d'un processus ponctuel et les notations associées sont définies dans A.5.2. Rappelons qu'un processus ponctuel intégrable est une suite strictement croissante de variables aléatoires positives (T_1, T_2, \dots) telle que $T_n \rightarrow \infty$ presque sûrement. Par convention, on adjoint à cette suite la variable aléatoire $T_0 = 0$ presque sûrement. Ces variables aléatoires vont représenter les instants d'arrivées des requêtes au système. On peut tout aussi bien décrire la suite par l'écart de temps qui s'écoule entre deux arrivées successives : $\xi_n = T_{n+1} - T_n$ est la n^{e} interarrivée. La donnée de la suite $(\xi_n, n \in \mathbf{N})$ définit aussi un processus ponctuel par la relation $T_n = \sum_{i \leq n-1} \xi_i$. On notera finalement $N(t)$, le nombre de points (d'arrivées), entre les instants 0 et t .

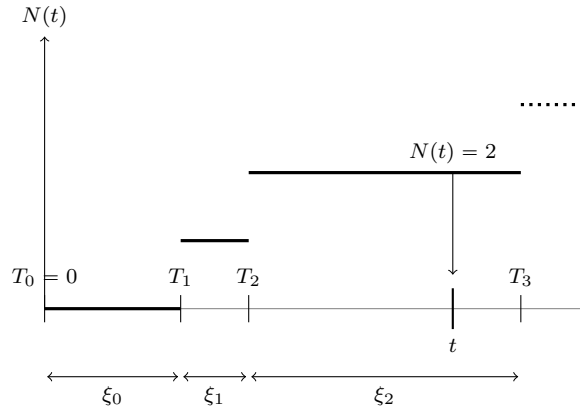


Figure 6.1 – Notations relatives aux processus ponctuels

6.1. Définitions

Le processus de Poisson admet plusieurs caractérisations. Comme n'importe laquelle peut être considérée comme une définition et les autres comme des propriétés, nous donnons à toutes le statut de définition et nous démontrons qu'elles sont équivalentes.

DÉFINITION 6.1.– Le processus ponctuel N est un processus de Poisson d'intensité λ si et seulement si les variables aléatoires $(\xi_n, n \in \mathbf{N})$ sont indépendantes et de même loi exponentielle de paramètre λ .

DÉFINITION 6.2.– Le processus ponctuel N est un processus de Poisson d'intensité λ si et seulement si les deux conditions suivantes sont vérifiées :

- pour tout $t \geq 0$, la variable aléatoire $N(t)$ suit une loi de Poisson de paramètre λt ;
- conditionnellement à $\{N(t) = n\}$, la famille (T_1, \dots, T_n) est uniformément distribuée sur $[0, t]$.

DÉFINITION 6.3.– Le processus ponctuel N est un processus de Poisson d'intensité λ si et seulement si les deux conditions suivantes sont vérifiées :

- pour tout $0 = t_0 < t_1 < \dots < t_n$, les variables aléatoires $(N(t_i + 1) - N(t_i), 1 \leq i \leq n - 1)$ sont indépendantes ;
- pour tout t, s , la variable aléatoire $N(t + s) - N(t)$ suit une loi de Poisson de paramètre λs :

$$\mathbf{P}(N(t + s) - N(t) = k) = \exp(-\lambda s) \frac{(\lambda s)^k}{k!}.$$

DÉFINITION 6.4.– Le processus ponctuel N est un processus de Poisson d'intensité λ si et seulement si pour toute fonction $f : \mathbf{R}^+ \rightarrow \mathbf{R}^+$ (ou pour toute fonction f à support compact dans \mathbf{R}^+), on a l'identité suivante :

$$\mathbf{E} \left[\exp \left(- \sum_{n \geq 1} f(T_n) \right) \right] = \exp \left(- \int_0^\infty (1 - e^{-f(s)}) \lambda \, ds \right).$$

DÉFINITION 6.5.– Le processus ponctuel N est un processus de Poisson d'intensité λ si et seulement si le processus $(N(t) - \lambda t, t \geq 0)$ est une martingale par rapport à la filtration \mathcal{F} engendrée par N , c'est-à-dire $\mathcal{F}_t = \sigma\{N(s), s \leq t\}$.

Afin de montrer l'équivalence entre ces définitions, nous devons introduire trois résultats techniques.

LEMME 6.1.– La densité de la loi de (T_1, \dots, T_n) est donnée par :

$$d\mathbf{P}_{(T_1, \dots, T_n)}(x_1, \dots, x_n) = \lambda^n \exp(-\lambda x_n) \mathbf{1}_C(x_1, \dots, x_n) \, dx_1 \dots dx_n, \quad (6.1)$$

où :

$$C = \{(y_1, \dots, y_n) \in (\mathbf{R}^+)^n, 0 \leq y_1 \leq \dots \leq y_n\}.$$

En particulier, T_n suit une loi gamma de paramètres n et λ , définie par :

$$d\mathbf{P}_{T_n}(x) = \lambda^n \exp(-\lambda x) \frac{x^{n-1}}{(n-1)!} \mathbf{1}_{\mathbf{R}^+}(x) \, dx \quad (6.2)$$

Démonstration. On procède par identification :

$$\begin{aligned} & \mathbf{E}[f(T_1, \dots, T_n)] \\ &= \int \dots \int_{(\mathbf{R}^+)^n} f(x_0, x_0 + x_1, \dots, x_0 + \dots + x_{n-1}) \, d\mathbf{P}_{\xi_0}(x_0) \dots d\mathbf{P}_{\xi_{n-1}}(x_{n-1}). \end{aligned}$$

On considère le changement de variables :

$$u_1 = x_0, \quad u_2 = x_0 + x_1, \quad \dots, \quad u_n = x_0 + \dots + x_{n-1},$$

dont le jacobien vaut clairement 1. Les conditions $x_0 \geq 0, \dots, x_{n-1} \geq 0$ reviennent à $0 \leq u_1 \leq u_2 \leq \dots \leq u_n$. On a donc :

$$\mathbf{E}[f(T_1, \dots, T_n)] = \int \dots \int_{(\mathbf{R}^+)^n} f(u_n) \lambda^n e^{-\lambda u_n} \mathbf{1}_C(u_1, \dots, u_n) \, du_1 \dots du_n.$$

La densité de la loi jointe s'en déduit. Si f ne dépend que de T_n , on obtient :

$$\begin{aligned} \mathbf{E}[f(T_n)] &= \int \dots \int_{0 \leq u_1 \dots \leq u_n} f(u_n) \lambda^n \exp(-\lambda u_n) \, d u_1 \dots d u_n \\ &= \int_0^\infty f(u_n) \lambda^n \exp(-\lambda u_n) \left(\int_0^{u_n} d u_{n-1} \int \dots \int_0^{u_2} d u_1 \right) d u_n \\ &= \int_0^\infty f(u_n) \lambda^n \exp(-\lambda u_n) \frac{u_n^{n-1}}{(n-1)!} \, d u_n. \end{aligned}$$

Le résultat en découle. \square

LEMME 6.2.– Soit X une variable aléatoire de loi de Poisson de paramètre λ . On a :

$$\mathbf{E}[e^{-sX}] = \exp(-\lambda(1 - e^{-s})).$$

Démonstration. Par définition de la loi de Poisson, on a l'identité suivante :

$$\mathbf{E}[e^{-sX}] = \sum_{k=0}^{\infty} \mathbf{E}[e^{-sk}] e^{-\lambda} \frac{\lambda^k}{k!} = \exp(-\lambda + \lambda e^{-s}),$$

d'où le résultat. \square

LEMME 6.3.– Soit (U_1, \dots, U_n) des variables aléatoires indépendantes de loi uniforme sur $[0, t]$. Soit \bar{U} , le réordonnement de ce n -uplet par ordre croissant, c'est-à-dire :

$$\bar{U}_1(\omega) \leq \bar{U}_2(\omega) \leq \dots \leq \bar{U}_n(\omega), \text{ presque sûrement.}$$

La loi de \bar{U} est donnée par :

$$d\mathbf{P}_{(\bar{U}_1, \dots, \bar{U}_n)}(x_1, \dots, x_n) = \frac{n!}{t^n} \mathbf{1}_{\mathcal{C}}(x_1, \dots, x_n) \, d x_1 \dots d x_n$$

Démonstration. On appelle σ la variable aléatoire à valeurs dans \mathfrak{S}_n , qui représente le réarrangement à faire pour mettre dans l'ordre les différentes valeurs des $U_i(\omega)$, c'est-à-dire que $\sigma(\omega)$ est défini de la manière suivante : si l'on a :

$$U_2(\omega) \leq U_3(\omega) \leq U_1(\omega),$$

alors :

$$\sigma(\omega) = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}.$$

L'image de i par $\sigma(\omega)$ est l'indice de la variable aléatoire qui est en i^e position pour cet ω . On a donc par définition $\bar{U}_i(\omega) = U_{\sigma(\omega)(i)}(\omega)$. Comme les variables aléatoires $U_i, i \in \llbracket 1, n \rrbracket$ sont indépendantes et de même loi, pour tout $\tau \in \mathfrak{S}_n$, on a :

$$d\mathbf{P}_{(U_{\tau(1)}, \dots, U_{\tau(n)})}(u_1, \dots, u_n) = \otimes_{i=1}^n \frac{1}{t} \mathbf{1}_{[0,t]}(u_i) \, d u_i.$$

On remarque en particulier que cette loi ne dépend pas de τ . Par conséquent, il vient :

$$\begin{aligned} \mathbf{P}(\sigma = \tau) &= \mathbf{P}(U_{\tau(1)} \leq \dots \leq U_{\tau(n)}) \\ &= \int \dots \int \mathbf{1}_{\mathcal{C}}(u_1, \dots, u_n) \, d\mathbf{P}_{(U_{\tau(1)}, \dots, U_{\tau(n)})}(u_1, \dots, u_n) \\ &= \mathbf{P}(\sigma = \text{Id}). \end{aligned}$$

Ainsi, σ suit une loi uniforme sur \mathfrak{S}_n , c'est-à-dire que $\mathbf{P}(\sigma = \tau) = 1/n!$. Pour calculer la loi du n -uplet \bar{U} on décompose l'espace d'état en la partition $\cup_{\tau \in \mathfrak{S}_n} (\sigma = \tau)$. Pour toute fonction continue bornée, on a l'identité suivante :

$$\begin{aligned} \mathbf{E}[f(\bar{U}_1, \dots, \bar{U}_n)] &= \sum_{\tau \in \mathfrak{S}_n} \mathbf{E}[f(\bar{U}_1, \dots, \bar{U}_n); \sigma = \tau] \\ &= \sum_{\tau \in \mathfrak{S}_n} \mathbf{E}[f(U_{\tau(1)}, \dots, U_{\tau(n)}) \mathbf{1}_{\mathcal{C}}(U_{\tau(1)}, \dots, U_{\tau(n)})] \\ &= \sum_{\tau \in \mathfrak{S}_n} \int \dots \int f(u_1, \dots, u_n) \mathbf{1}_{\mathcal{C}}(u_1, \dots, u_n) \, d\mathbf{P}_{(U_{\tau(1)}, \dots, U_{\tau(n)})}(u_1, \dots, u_n) \\ &= \sum_{\tau \in \mathfrak{S}_n} \int \dots \int f(u_1, \dots, u_n) \mathbf{1}_{\mathcal{C}}(u_1, \dots, u_n) \otimes_{i=1}^n \frac{1}{t} \mathbf{1}_{[0,t]}(u_i) \, du_i \\ &= \frac{n!}{t^n} \int \dots \int f(u_1, \dots, u_n) \mathbf{1}_{\mathcal{C}}(u_1, \dots, u_n) \, du_1 \dots du_n, \end{aligned}$$

d'où le résultat. □

Démonstration des équivalences entre les définitions. Nous allons montrer la chaîne d'implications : $6.1 \implies 6.2 \implies 6.3 \implies 6.4 \implies 6.5 \implies 6.1$.

$6.1 \implies 6.2$.

Montrons d'abord que $N(t)$ suit une loi de Poisson. Comme il est évident que les événements $\{N(t) = k\}$ et $\{T_k \leq t < T_{k+1}\}$ coïncident, on a :

$$\begin{aligned} \mathbf{P}(N(t) = k) &= \mathbf{P}(T_k \leq t < T_k + \xi_{k+1}) \\ &= \iint \mathbf{1}_{\{x \leq t\}} \mathbf{1}_{\{x+y > t\}} \, d\mathbf{P}_{T_k}(x) \, d\mathbf{P}_{\xi_k}(y) \\ &= \int_0^t \left(\int_{t-x}^{\infty} \lambda e^{-\lambda y} \, dy \right) \lambda^k \frac{x^{k-1}}{(k-1)!} \exp(-\lambda x) \, dx \\ &= e^{-\lambda t} \int_0^t \lambda^k \frac{x^{k-1}}{(k-1)!} \, dx = e^{-\lambda t} \frac{(\lambda x)^k}{k!}. \end{aligned}$$

Pour la loi conditionnelle, on procède de façon analogue :

$$\begin{aligned}
& \mathbf{P}((T_1, \dots, T_n) \in A \mid N(t) = n) \mathbf{P}(N(t) = n) \\
&= \mathbf{P}((T_1, \dots, T_n) \in A, T_n \leq t < T_{n+1}) \\
&= \int_{0 \leq u_1 \leq \dots \leq u_{n+1}} \mathbf{1}_A(u_1, \dots, u_n) \mathbf{1}_{[u_n, u_{n+1}[}(t) \lambda^{n+1} e^{-\lambda u_{n+1}} \, du_1 \dots du_{n+1} \\
&= \lambda^n \int_{0 \leq u_1 \leq \dots \leq u_{n+1}} \mathbf{1}_A(u_1, \dots, u_n) \mathbf{1}_{[u_n, \infty[}(t) \left(\int_t^\infty \lambda e^{-\lambda u_{n+1}} \right) \, du_1 \dots du_n \\
&= \lambda^n e^{-\lambda t} \int_{0 \leq u_1 \leq \dots \leq u_{n+1}} \mathbf{1}_A(u_1, \dots, u_n) \mathbf{1}_{[u_n, \infty[}(t) \, du_1 \dots du_n.
\end{aligned}$$

En divisant le terme de droite par $e^{-\lambda t}(\lambda t)^n/n!$, on obtient :

$$\begin{aligned}
& \mathbf{P}((T_1, \dots, T_n) \in A \mid N(t) = n) \\
&= \frac{n!}{t^n} \int_{0 \leq u_1 \leq \dots \leq u_{n+1}} \mathbf{1}_A(u_1, \dots, u_n) \mathbf{1}_{[u_n, \infty[}(t) \, du_1 \dots du_n,
\end{aligned}$$

ce qui en vertu du lemme 6.3 signifie que (T_1, \dots, T_n) a conditionnellement à $\{N(t) = n\}$ la même loi que $(\bar{U}_1, \dots, \bar{U}_n)$ définie dans ce lemme. On résume ceci en disant que, conditionnellement à $\{N(t) = n\}$, le n -uplet (T_1, \dots, T_n) est uniformément réparti sur $[0, t]$.

6.2 \implies 6.3.

Soit $t_0 = 0 < t_1 < \dots < t_n$ des réels et i_0, \dots, i_{n-1} des entiers. On veut prouver que :

$$\mathbf{P} \left(\bigcap_{l=0}^{n-1} \{N(t_{l+1}) - N(t_l) = i_l\} \right) = \prod_{l=1}^n \mathbf{P}(N(t_{l+1}) - N(t_l) = i_l).$$

On peut toujours écrire :

$$\begin{aligned}
& \mathbf{P} \left(\bigcap_{l=0}^{n-1} \{N(t_{l+1}) - N(t_l) = i_l\} \right) \\
&= \sum_{k \in \mathbf{N}} \mathbf{P} \left(\bigcap_{l=0}^{n-1} \{N(t_{l+1}) - N(t_l) = i_l\} \mid N(t_n) = k \right) \mathbf{P}(N(t_n) = k).
\end{aligned}$$

La seule valeur de k pour laquelle les probabilités conditionnelles de cette somme sont non nulles est $k_0 = \sum_l i_l$. Pour calculer la probabilité conditionnelle correspondante on sait que les impacts entre 0 et t_n sont uniformément répartis. Cette quantité égale donc la probabilité que k points uniformément répartis sur un intervalle se répartissent en i_1 points dans l'intervalle $[0, t_1]$, i_2 points dans l'intervalle $]t_1, t_2]$, etc. Chaque point se trouve dans un intervalle de longueur x avec probabilité x/t_n , par conséquent on a :

$$\begin{aligned} \mathbf{P} \left(\bigcap_{l=0}^{n-1} \{N(t_{l+1}) - N(t_l) = i_l\} \mid N(t_n) = k_0 \right) \\ = \frac{k_0!}{i_1! \dots i_n!} \prod_{l=0}^{n-1} \left(\frac{t_{l+1} - t_l}{t_n} \right)^{i_l}. \end{aligned}$$

Comme $N(t_n)$ suit une loi de Poisson de paramètre λt_n et $k_0 = \sum_{l=0}^{n-1} i_l$, on en déduit que :

$$\begin{aligned} \mathbf{P} \left(\bigcap_{l=0}^{n-1} \{N(t_{l+1}) - N(t_l) = i_l\} \right) \\ = e^{-\lambda t_n} \frac{(\lambda t_n)^{k_0}}{k_0!} \frac{k_0!}{i_1! \dots i_n!} \prod_{l=0}^{n-1} \left(\frac{t_{l+1} - t_l}{t_n} \right)^{i_l} \\ = \prod_{l=0}^{n-1} e^{-\lambda(t_{l+1} - t_l)} \frac{(\lambda(t_{l+1} - t_l))^{i_l}}{i_l!}. \end{aligned}$$

On a pu mettre la probabilité de l'intersection comme un produit de probabilité, par conséquent les variables aléatoires sont indépendantes. En prenant $n = 2$, $t_1 = t$, $t_2 = t + s$, $i_0 = i$ et $i_1 = j$, on obtient :

$$\mathbf{P}(N(t) = i, N(t+s) - N(t) = j) = e^{-\lambda t} \frac{(\lambda t)^i}{i!} e^{-\lambda s} \frac{(\lambda s)^j}{j!}.$$

En sommant sur toutes les valeurs de i , on obtient le résultat voulu.

6.3 \implies 6.4.

Remarquons que si l'on prend $f(s) = \mathbf{1}_{]a, b]}(s)$, on a :

$$\sum_n f(T_n) = N(b) - N(a).$$

Du lemme 6.2, on déduit que le résultat est vrai pour les fonctions indicatrices et par linéarité pour les fonctions étagées (c'est-à-dire les combinaisons linéaires de fonctions

indicatrices). Par convergence monotone, on en déduit que le résultat est vrai pour toute fonction mesurable positive.

6.4 \implies 6.5.

Il suffit d'écrire $N(t+s) = (N(t+s) - N(t)) + N(t)$ et d'utiliser l'indépendance de ces variables aléatoires pour prouver que :

$$\mathbf{E}[N(t+s) | \mathcal{F}_t] = \mathbf{E}[N(t+s) - N(t)] + N(t) = \lambda s + N(t),$$

d'où le résultat.

6.5 \implies 6.1

Comme N n'évolue pas entre ses instants de sauts, pour f bornée, on a :

$$\begin{aligned} f(N(t)) - f(N(s)) &= \sum_{s < r \leq t, \Delta N(r)=1} f(N(r)) - f(N(r_-)) \\ &= \int_s^t (f(N(r_-) + 1) - f(N(r_-))) dN(r) \\ &= \int_s^t (f(N(r_-) + 1) - f(N(r_-))) (dN(r) - \lambda dr) \\ &\quad + \int_s^t (f(N(r_-) + 1) - f(N(r_-))) \lambda dr. \end{aligned}$$

Comme le processus $(r \mapsto f(N(r_-)))$ est prévisible, le théorème A.34 implique que l'intégrale stochastique définit bien une martingale donc on a :

$$t \mapsto f(N(t)) - \int_0^t (f(N(r_-) + 1) - f(N(r_-))) \lambda dr$$

est une martingale. D'après le théorème 7.15, N est un processus de Markov de générateur infinitésimal :

$$Af(x) = \lambda(f(x+1) - f(x)) \text{ pour } f : \mathbf{N} \rightarrow \mathbf{R} \text{ bornée.}$$

D'après le théorème 7.9, la définition 6.1 est vérifiée. \square

6.2. Propriétés

Il ne faut pas se laisser abuser par la définition 6.2 qui stipule que, conditionnellement au nombre de points sur un intervalle, les impacts sont uniformément répartis sur cet intervalle. Certes, lorsque l'on observe une trajectoire, on connaît t et le nombre

d'impacts dans cet intervalle, on devrait donc observer des points uniformément répartis. Pourtant, on observe une répartition similaire à celle de la figure 6.2. C'est le phénomène de « clusterisation » : sur une trajectoire, les arrivées donnent l'impression d'être groupées. C'est d'ailleurs ce que l'on constate dans les magasins où après une période d'inactivité, plusieurs clients arrivent presque en même temps.

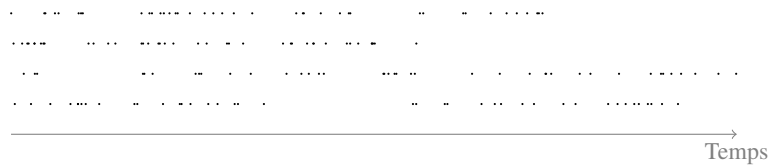


Figure 6.2 – Quatre trajectoires d'un processus de Poisson

⚡ En fait, la notion même de répartition uniforme est floue, il ne faut pas la confondre avec la constance de l'écart entre les arrivées. Comme le montre la figure 6.2, l'uniformité dans la répartition se « voit » sur plusieurs trajectoires : il n'y a ici presque plus de parties de $[0, 1]$ qui ne comporte pas d'impact dans l'une ou l'autre des trajectoires. Il y a une différence essentielle entre le phénomène de « clusterisation » apparente du processus de Poisson et le régime de « rafales » (ou *burst*) que l'on observe dans les grands réseaux haut débit : dans le premier cas, le débit moyen instantané (calculé en moyennant sur un grand nombre de trajectoires) ne dépend pas du temps (puisque'il est égal à λ) tandis que dans l'autre cas, il va grandement varier avec le temps (penser à de la vidéo à débit variable). On ne peut donc pas représenter de tels trafics par un processus de Poisson.

THÉORÈME 6.4. – Soit N un processus de Poisson d'intensité λ . Pour toute fonction f à support compact, on a :

$$\mathbf{E} \left[\exp \left(- \sum_{n \geq 1} f(T_n) \right) \right] = \exp \left(- \int_{\mathbf{R}^+} (1 - e^{-f(s)}) \lambda \, ds \right) \quad (6.3)$$

Démonstration. Si l'on prend $f(s) = \alpha \mathbf{1}_{]a,b]}(s)$, on a :

$$\sum_n f(T_n) = \alpha(N(b) - N(a)).$$

On sait que $N(b) - N(a)$ suit une loi de Poisson de paramètre $\lambda(b - a)$. Par conséquent, on a :

$$\begin{aligned} \mathbf{E} \left[\exp \left(- \sum_{n \geq 1} f(T_n) \right) \right] &= \sum_{n=0}^{\infty} e^{-\alpha n} e^{-\lambda(b-a)} \frac{(\lambda(b-a))^n}{n!} \\ &= \exp(-\lambda(b-a) + \lambda(b-a)e^{-\alpha}) \\ &= \exp \left(- \int (1 - e^{-f(s)}) \lambda \, ds \right). \end{aligned}$$

Par indépendance des incréments, l'équation (6.3) est donc vraie pour les fonctions étagées. Par convergence dominée, elle l'est aussi pour les fonctions à support compact. \square

6.2.1. Superposition, amincissement

Lorsque l'on dispose de deux processus ponctuels N^1 et N^2 , la superposition de ces processus est le processus ponctuel, noté $N = N^1 + N^2$, dont les impacts sont ceux de N^1 et ceux de N^2 .

En vertu du théorème 6.4, le résultat suivant est immédiat.

THÉORÈME 6.5.— La superposition de deux processus de Poisson indépendants est un processus de Poisson d'intensité somme.

REMARQUE.— En particulier, lors de la superposition de deux processus de Poisson indépendants, deux clients ne peuvent arriver simultanément. Ce résultat reste vrai pour n'importe quels processus ponctuels indépendants.

Supposons maintenant qu'un processus de Poisson N d'intensité λ soit éclaté en deux processus N^1 et N^2 selon un tirage de Bernoulli de paramètre p , c'est-à-dire qu'à chaque point de N , on décide qu'il appartient à N^1 avec probabilité p et à N^2 avec probabilité $1 - p$, ce tirage au sort étant indépendant de tout le reste et en particulier des précédents tirages au sort (voir la figure 6.3 dans laquelle le chiffre au-dessus ou au-dessous de chaque point représente le numéro du flux auquel ce point est attribué). On dit que le processus de Poisson est aminci.



Figure 6.3 – Amincissement d'un processus de Poisson

Le théorème 6.4 assure qu'un processus de Poisson sur \mathbf{R}^+ est un cas particulier de processus de Poisson spatial (voir section 10.3). Par conséquent, le théorème 10.6 assure le résultat qui suit.

THÉORÈME 6.6.— Les processus N^1 et N^2 résultants de l'amincissement de N , sont deux processus de Poisson indépendants d'intensités respectives λp et $\lambda(1 - p)$.

6.2.2. Paradoxe de l'autobus

Il est en revanche un résultat spécifique de la dimension 1, connu sous le nom de paradoxe de l'autobus (ou paradoxe de l'inspection). Interprétons les points d'un processus de Poisson comme les instants d'arrivée des bus à un arrêt donné. Soit :

$$\begin{cases} W(t) &= T_{N(t)+1} - t, \\ Z(t) &= t - T_{N(t)}, \\ \xi(t) &= W(t) + Z(t) = T_{N(t)+1} - T_{N(t)}, \end{cases}$$

respectivement le temps d'attente du bus si l'on arrive en t à l'arrêt, le temps écoulé depuis le passage du dernier bus et l'intervalle de temps entre le bus que l'on prend et celui que l'on a raté.

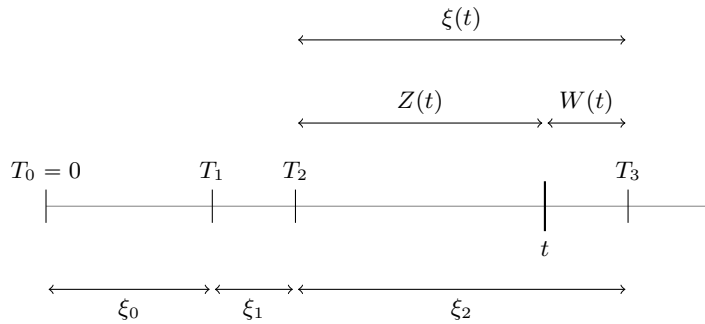
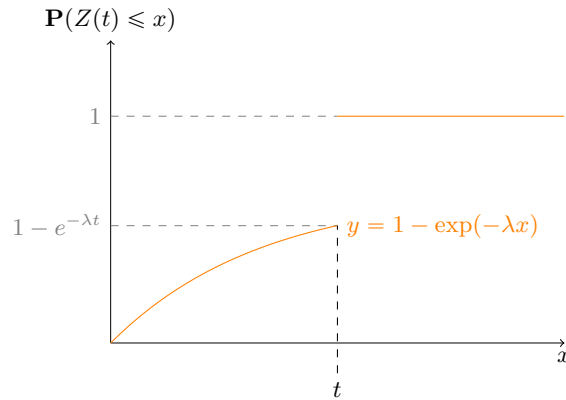


Figure 6.4 – Notations

THÉORÈME 6.7 (Paradoxe de l'autobus).— $W(t)$ suit une loi exponentielle de paramètre λ et est indépendante de $Z(t)$, dont la loi est donnée par :

$$\mathbf{P}(Z(t) \leq x) = \begin{cases} 1 - e^{-\lambda x} & \text{si } 0 \leq x < t; \\ 1 & \text{si } x \geq t. \end{cases}$$

Figure 6.5 – Fonction de répartition de $Z(t)$

⚡ Ceci peut sembler paradoxal puisque la valeur moyenne de W_t , c'est-à-dire le temps moyen d'attente, est donc $1/\lambda$ alors que le temps moyen qui s'écoule entre les passages de deux bus vaut aussi $1/\lambda$. Cette propriété est une manifestation de ce qu'il est convenu d'appeler l'absence de mémoire de la loi exponentielle, sur laquelle nous reviendrons dans le chapitre suivant. En fait, tout se passe comme si, au moment où l'on arrive à l'arrêt de bus, le compteur qui égrène le temps qui s'écoule entre deux bus était remis à zéro et que l'on recomptait un temps de loi exponentielle jusqu'à la prochaine arrivée.

A proprement parler, cette approche est mathématiquement fautive car si ξ_n suit bien une loi exponentielle de paramètre λ , $\xi(t)$ n'a pas la loi de ξ_n puisqu'en fonction de ω le numéro du bus que l'on vient de rater change. Conditionnellement à $\{N(t) = n\}$, $\xi(t)$ a effectivement une loi exponentielle de paramètre λ mais contrairement à ce que l'on pourrait croire, lorsque l'on déconditionne, la loi de $\xi(t)$ n'est plus une loi exponentielle.

Démonstration du théorème 6.7. Notons que $Z(t)$, par définition même, ne peut pas être plus grand que t . Pour $0 \leq x < t$ et $y \geq 0$, en utilisant au passage l'indépendance de T_n et ξ_n pour tout $n \in \mathbf{N}$ on a :

$$\begin{aligned} \mathbf{P}(Z(t) \leq x, W(t) > y) &= \sum_{n \in \mathbf{N}} \mathbf{P}(Z(t) \leq x, W(t) > y, N(t) = n) \\ &= \sum_{n \in \mathbf{N}} \mathbf{P}(t - T_n \leq x, T_n + \xi_n - t > y, T_n \leq t < T_n + \xi_n). \end{aligned}$$

Pour $n = 0$, cette dernière probabilité vaut :

$$\mathbf{P}(t \leq x, \xi_0 > t + y) = 0$$

puisque l'on a supposé que $x < t$. La somme précédente s'écrit donc :

$$\begin{aligned}
& \sum_{n>1} \iint \mathbf{1}_{\{t-u \leq x\}} \mathbf{1}_{\{u+v-t > y\}} \mathbf{1}_{\{u \leq t\}} d\mathbf{P}_{T_n}(u) d\mathbf{P}_{\xi_{n+1}}(v) \\
&= \sum_{n>1} \int_{t-x}^t \lambda^n e^{-\lambda u} \frac{u^{n-1}}{(n-1)!} \left(\int_{t+y-u}^{\infty} \lambda e^{-\lambda v} dv \right) du \\
&= \sum_{n>1} \lambda^n e^{-\lambda(t+y)} \int_{t-x}^t \frac{u^{n-1}}{(n-1)!} du \\
&= \sum_{n>1} \lambda^n e^{-\lambda(t+y)} \left[\frac{t^n}{n!} - \frac{(t-x)^n}{n!} \right] \\
&= e^{-\lambda(t+y)} \left(\sum_{n>1} \frac{(\lambda t)^n}{n!} - \sum_{n>1} \frac{(\lambda(t-x))^n}{n!} \right) \\
&= e^{-\lambda(t+y)} \left(e^{\lambda t} - 1 - \left(e^{\lambda(t-x)} - 1 \right) \right) \\
&= e^{-\lambda y} (1 - e^{-\lambda x}).
\end{aligned}$$

Comme $\lim_{x \nearrow t} (1 - e^{-\lambda x}) = 1 - e^{-\lambda t} < 1$, on en déduit qu'il y a un saut dans la fonction de répartition de $Z(t)$ et par conséquent que $\mathbf{P}(Z(t) = t) = e^{-\lambda t} > 0$. \square

REMARQUE.— Le calcul de l'espérance de $\xi(t)$ se fait de la manière suivante :

$$\begin{aligned}
\mathbf{E}[\xi(t)] &= \mathbf{E}[W(t)] + \mathbf{E}[Z(t)] \\
&= \frac{1}{\lambda} + t \cdot e^{-\lambda t} + \int_0^t \lambda s e^{-\lambda s} ds \\
&= \frac{1}{\lambda} (2 - e^{-\lambda t}).
\end{aligned}$$

Par conséquent, lorsque t grandit, la moyenne de $\xi(t)$ tend vers $2/\lambda$ et le temps moyen d'attente, qui vaut $1/\lambda$, en représente la moitié, conformément à l'intuition.

6.3. Analogie discret : le processus de Bernoulli

L'analogie du processus de Poisson à temps discret est défini de la manière suivante.

DÉFINITION 6.6.— Soit $p \in]0, 1[$ et $(\tilde{\xi}_n, n \in \mathbf{N})$, une suite de variables aléatoires indépendantes et identiquement distribuées, de loi géométrique de paramètre p , c'est-à-dire :

$$\mathbf{P}(\tilde{\xi}_0 = k) = p(1-p)^{k-1}; \quad k > 1.$$

On pose alors $\tilde{T}_0 = 0$ et $\tilde{T}_{n+1} - \tilde{T}_n = \tilde{\xi}_n$ pour tout $n \in \mathbf{N}$. Le processus ponctuel $(T_0, T_1, \dots, T_n, \dots)$ ainsi défini est appelé processus de Bernouilli.

REMARQUE.— Comme pour le processus de Poisson, on peut définir le processus aléatoire \tilde{N} , à trajectoires cadlag, défini pour tout $t \geq 0$ par :

$$\tilde{N}(t) = \sum_{n \in \mathbf{N}} \mathbf{1}_{\{\tilde{T}_n \leq t\}},$$

qui compte le nombre de points du processus jusqu'à t . Il est clair que $\tilde{N}(t)$ suit pour tout $t \geq 0$, la loi binomiale $\mathcal{B}([t], p)$, où $[\cdot]$ désigne la partie entière.

Nous pouvons vérifier que le processus de Bernouilli satisfait à l'analogie du paradoxe de l'autobus. Pour tout $k \in \mathbf{N}$, notons par analogie avec la section 6.2.2 :

$$\begin{cases} \tilde{W}_k &= k - \tilde{T}_{\tilde{N}(k)}, \text{ le temps écoulé depuis le dernier point avant } k; \\ \tilde{Z}_k &= \tilde{T}_{\tilde{N}(k)+1}, \text{ le temps à attendre avant le premier point après } k. \end{cases}$$

Par définition, \tilde{W}_k est nul s'il y a précisément un point en k , alors que \tilde{Z}_k est nécessairement strictement positif. On a le résultat suivant.

THÉORÈME 6.8.— Pour tout $k \geq 0$, \tilde{W}_k suit la loi géométrique de paramètre p et est indépendante de \tilde{Z}_k , dont la loi est donnée par :

$$\mathbf{P}(\tilde{Z}_k = i) = \begin{cases} p(1-p)^i & \text{si } i < k; \\ (1-p)^i & \text{si } i = k; \\ 0 & \text{si } i > k. \end{cases}$$

Démonstration. Le schéma de la preuve est le même que celui de la preuve du théorème 6.7. Fixons $j > 1$. Il est clair par définition que \tilde{Z}_k ne peut pas être supérieur à k . Par ailleurs, pour tout $i < k$, $\tilde{Z}_k = i$ si $\tilde{N}(k)$ est strictement positif (sinon $k = i$) et strictement inférieur à $k - i$, car il n'y a aucun point entre l'instant $k - i$ et l'instant k . Par conséquent :

$$\begin{aligned} \mathbf{P}(\tilde{Z}_k = i, \tilde{W}_k = j) &= \sum_{1 \leq n \leq k-i} \mathbf{P}(\tilde{Z}_k = i, \tilde{W}_k = j, \tilde{N}(k) = n) \\ &= \sum_{1 \leq n \leq k-i} \mathbf{P}(\tilde{T}_n = k - i, \tilde{\xi}_n = i + j) \\ &= \sum_{1 \leq n \leq k-i} \mathbf{P}(\tilde{T}_n = k - i) p(1-p)^{i+j-1}, \end{aligned}$$

par indépendance. Or, $T_n = k - i$ revient à dire qu'il y a un point à $k - i$ (et donc un succès à une expérience de Bernouilli de paramètre p) et si $k - i - 1 > 0$, qu'il

y a eu $n - 1$ succès lors de $k - i - 1$ expériences de Bernouilli précédentes, toutes indépendantes. Autrement dit :

$$\mathbf{P}(\tilde{T}_n = k - i) = p\mathbf{P}(B = n - 1) = pC_{k-i-1}^{n-1}p^{n-1}(1-p)^{k-i-1-(n-1)},$$

où B est une variable aléatoire de loi binomiale $\mathcal{B}(k - i - 1, p)$ et en comprenant éventuellement $C_0^0 = 1$. Pour tout $i < k$, on a donc :

$$\begin{aligned} \mathbf{P}(\tilde{Z}_k = i, \tilde{W}_k = j) &= \sum_{1 \leq n \leq k-i} pC_{k-i-1}^{n-1}p^{n-1}(1-p)^{k-i-1-(n-1)}p(1-p)^{i+j-1} \\ &= p^2(1-p)^{i+j-1} \sum_{0 \leq n \leq k-i-1} C_{k-i-1}^n p^n (1-p)^{k-i-1-n} \\ &= p(1-p)^{j-1}p(1-p)^i, \end{aligned} \quad (6.4)$$

d'après la formule du binôme. Par ailleurs, $\tilde{Z}_k = k$ revient à dire que $\tilde{N}(k) = 0$ et donc :

$$\begin{aligned} \mathbf{P}(\tilde{Z}_k = k, \tilde{W}_k = j) &= \mathbf{P}(\tilde{\xi}_0 = i + j) \\ &= p(1-p)^{i+j-1} = p(1-p)^{j-1}(1-p)^i. \end{aligned} \quad (6.5)$$

De (6.4) et (6.5), on déduit que \tilde{Z}_k et \tilde{W}_k sont indépendantes et suivent les lois annoncées. \square

⌘ Cette propriété d'oubli se comprend aisément : une variable aléatoire $\tilde{\xi}_i$ compte le nombre d'essais de Bernouilli indépendants, nécessaires pour obtenir un succès. La variable aléatoire \tilde{W}_k compte le nombre d'essais encore nécessaires à partir de k pour obtenir le premier succès après k . Les essais étant indépendants, on voit bien là encore que l'attente « capitalisée » entre le dernier succès jusqu'à k n'augmente pas la chance de succès à chaque tentative après k et donc \tilde{W}_k a la même loi que n'importe lequel des $\tilde{\xi}_i$.

Le dernier résultat est donc intuitivement clair et il peut nous aider à comprendre le paradoxe de l'autobus en temps continu. En effet, le processus de Poisson n'est autre qu'une version macroscopique de processus de Bernouilli. Plus précisément, on fixe $\lambda > 0$ et pour tout entier $n > 1$ tel que $\lambda/n < 1$, on note \tilde{N}^n , un processus de Bernouilli de paramètre λ/n , de variables associées $\tilde{\xi}_0^n, \tilde{\xi}_1^n, \dots$. Finalement, on définit pour tout n le processus ponctuel \bar{N}^n comme suit :

$$\left\{ \begin{array}{l} \bar{\xi}_i^n = \tilde{\xi}_i^n/n, \quad i \in \mathbf{N}; \\ \bar{T}_0^n = 0, \quad \bar{T}_{i+1}^n - \bar{T}_i^n = \bar{\xi}_i^n, \quad i \in \mathbf{N}; \\ \bar{N}^n(t) = \sum_{i \in \mathbf{N}} \mathbf{1}_{\{\bar{T}_i^n \leq t\}}. \end{array} \right.$$

A partir du processus de Bernoulli de paramètre λ , pour obtenir \bar{N}^n , on divise donc par n la probabilité d'occurrence d'un point à chaque instant et l'on compense par une « accélération du temps » de facteur n , en divisant les temps interpoints par n . On a alors le résultat suivant.

THÉORÈME 6.9.— La suite de processus $\{\bar{N}^n, n > [\lambda]\}$ converge en loi vers le processus de Poisson de paramètre λ .

Démonstration. La notion de convergence en loi pour les processus est assez lourde à définir. Nous n'entrerons pas ici dans ces détails. En fait, il suffit dans notre cas de vérifier que les interpoints de \bar{N}^n tendent en loi vers ceux du processus de Poisson de paramètre λ . C'est clairement le cas puisque pour tout $i \in \mathbf{N}$ et tout t , on a :

$$\mathbf{P}(\bar{\xi}_i^n > t) = \mathbf{P}(\tilde{\xi}_i^n > nt) = \left(1 - \frac{\lambda}{n}\right)^{[nt]},$$

et la dernière quantité tend vers $e^{-\lambda t}$ quand n tend vers l'infini. \square

6.4. Simulation du processus de Poisson

La définition 6.1 permet de simuler des trajectoires d'un processus de Poisson d'intensité λ simplement en effectuant des tirages successifs de variables aléatoires de loi exponentielle de paramètre λ .

Algorithme 6.1 . Réalisation d'une trajectoire d'un processus de Poisson (méthode 1)

Données : λT

Résultat : une trajectoire $(t_n, n \geq 1)$ sur $[0, T]$ d'un processus de Poisson d'intensité λ

$t \leftarrow 0;$

$n \leftarrow 0;$

tant que $t \leq T$ **faire**

$x \leftarrow$ réalisation d'une $\varepsilon(\lambda);$

$t \leftarrow t + x;$

$t_n \leftarrow t;$

$n \leftarrow n + 1$

fin

retourner t_1, t_2, \dots, t_n

La définition 6.2 permet de simuler une trajectoire sur $[0, t]$ en effectuant un tirage d'une loi de Poisson de paramètre λt , dont on note le résultat par k , puis d'effectuer

k tirages uniformes sur $[0, t]$. Une fois ordonné, le plus petit de ces tirages peut être assimilé au premier impact, le deuxième plus petit à T_2 , etc.

Algorithme 6.2 . Réalisation d'une trajectoire d'un processus de Poisson (méthode 2)

Données : λT

Résultat : une trajectoire $(t_n, n \geq 1)$ sur $[0, T]$ d'un processus de Poisson d'intensité λ

$n \leftarrow$ réalisation d'une variable aléatoire de loi de Poisson de paramètre λT ;

pour $i = 1, \dots, n$ **faire**

 | $u_i \leftarrow$ réalisation d'une $U([0, 1])$;

fin

$(t_1, \dots, t_n) \leftarrow$ tri en croissant de (u_1, \dots, u_n) ;

retourner t_1, t_2, \dots, t_n

Dans les deux méthodes, si l'on veut ensuite prolonger la trajectoire sur $[t, t + s]$, on recommence la même chose sur $[t, t + s]$ seulement, puisque la définition 6.3 garantit que la trajectoire sur $[t, t + s]$ est indépendante de celle sur $[0, t]$.

La méthode 1 est la plus avantageuse pour les « grandes » valeurs de λ , parce que simuler une loi de Poisson pour λ grand se révèle impossible dès lors que $\exp(-\lambda)$ devient plus petit que la précision numérique de l'ordinateur. En revanche, comme nous le voyons dans le chapitre 10, la méthode 2 est la seule qui s'adapte à la simulation d'un processus de Poisson spatial.

6.5. Processus de Poisson non homogène

DÉFINITION 6.7.— N est un processus de Poisson non homogène d'intensité $\lambda(s)$ où s varie dans \mathbf{R}^+ si et seulement si pour toute fonction $f : \mathbf{R}^+ \rightarrow \mathbf{R}^+$, on a l'identité suivante :

$$\exp\left(-\sum_{n \geq 1} f(T_n)\right) = \exp\left(-\int (1 - e^{-f(s)}) \lambda(s) ds\right). \quad (6.6)$$

Cette définition doit être rapprochée de la définition 6.4. Si λ est une fonction constante alors on retrouve la définition d'un processus de Poisson d'intensité λ . Cette classe de processus modélise des trafics dans lesquels le débit varie au cours du temps mais de façon connue (λ est une fonction déterministe). On suppose toujours que λ est

continue à droite pourvue de limites à gauche en tout point. Posons :

$$a(t) = \int_0^t \lambda(s) \, ds;$$

$$\tau(t) = \inf\{s \geq 0, a(s) \geq t\}.$$

Le graphe de τ s'obtient en prenant le symétrique de celui de a par rapport à la première bissectrice. Par définition, $a(\tau(t)) = t$. La relation inverse, $\tau(a(t)) = t$ n'est vraie que si λ ne s'annule pas sur un intervalle. En effet, dans ce cas, a présente un intervalle de constance et τ n'est plus continue.

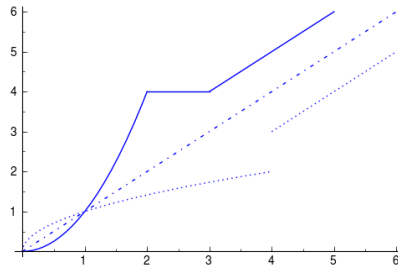


Figure 6.6 – Graphe de a et de son inverse τ (en pointillés)

THÉORÈME 6.10.– Soit N un processus de Poisson non homogène d'intensité $\lambda(s) \, ds$ alors le processus \tilde{N} défini par :

$$\tilde{N}(t) = \sum_{s: \Delta N(s)=1} \mathbf{1}_{\{a(s) \leq t\}}, \quad (6.7)$$

est un processus de Poisson homogène d'intensité ds .

☞ En d'autres termes, quand N saute à l'instant t , \tilde{N} a un saut à l'instant $a(t)$.

Démonstration. Il suffit de montrer que $M(t) = N(\tau_t)$ vérifie les propriétés de la définition 6.3. Or, pour tout $0 = t_0 < t_1 < \dots < t_n$, on l'identité suivante :

$$M(t_{i+1}) - M(t_i) = \sum_n \mathbf{1}_{] \tau_{t_i}, \tau_{t_{i+1}}]}(T_n).$$

Comme τ est déterministe, par définition du processus de Poisson non homogène, on a :

$$\begin{aligned}
\mathbf{E} \left[\exp \left(- \sum_j \alpha_j (M(t_{j+1}) - M(t_j)) \right) \right] &= \\
&= \mathbf{E} \left[\exp \left(- \sum_n \sum_j \alpha_j \mathbf{1}_{\{\tau_{t_j}, \tau_{t_{j+1}}\}}(T_n) \right) \right] \\
&= \exp \left(- \sum_j \int_{\tau_{t_j}}^{\tau_{t_{j+1}}} (1 - e^{-\alpha_j}) \lambda(s) \, ds \right) \\
&= \exp \left(- \sum_j (1 - e^{-\alpha_j}) [a(\tau_{t_{j+1}}) - a(\tau_{t_j})] \right) \\
&= \exp \left(- \sum_j \alpha_j (t_{j+1} - t_j) \right),
\end{aligned}$$

d'où le résultat. \square

Avec les notations précédentes, on en déduit l'algorithme de simulation d'un processus de Poisson non homogène.

Algorithme 6.3 . Réalisation d'une trajectoire d'un processus de Poisson non homogène

Données : a, T

Résultat : une trajectoire $(t_n, n \geq 1)$ sur $[0, T]$ d'un processus de Poisson d'intensité $\lambda(s) \, ds$

$s_1, \dots, s_n \leftarrow$ réalisation d'un processus de Poisson d'intensité 1 sur $a(T)$;

retourner $t_i = a(s_i), i = 1, \dots, n$

THÉORÈME 6.11.— Le processus ponctuel N est un processus de Poisson non homogène d'intensité $\lambda(s) \, ds$ si et seulement si le processus \tilde{N} est une martingale où :

$$\tilde{N} : t \longmapsto N(t) - \int_0^t \lambda(s) \, ds.$$

Démonstration. Soit f à support compact. D'après la formule d'Itô (A.14), on a :

$$\begin{aligned}
\exp \left(\int_0^t f(r) \, d\tilde{N}(r) \right) &= 1 + \int_0^t \exp \left(\int_0^s f(r) \, d\tilde{N}(r) \right) f(s) \, d\tilde{N}(s) \\
&\quad + \int_0^t \exp \left(\int_0^s f(r) \, d\tilde{N}(r) \right) (e^{f(s)} - 1 - f(s)) \, dN(s).
\end{aligned}$$

En vertu du théorème A.34, l'intégrale stochastique du terme de droite est une martingale, donc en prenant les espérances, il reste :

$$\begin{aligned} \mathbf{E} \left[\exp \left(\int_0^t f(r) d\tilde{N}(r) \right) \right] \\ &= 1 + \mathbf{E} \left[\int_0^t \exp \left(\int_0^s f(r) d\tilde{N}(r) \right) (e^{f(s)} - 1 - f(s)) dN(s) \right] \\ &= 1 + \mathbf{E} \left[\int_0^t \exp \left(\int_0^s f(r) d\tilde{N}(r) \right) (e^{f(s)} - 1 - f(s)) \lambda(s) ds \right]. \end{aligned}$$

En posant $\phi(t) = \exp(\int_0^t f(r) d\tilde{N}(r))$, il vient alors :

$$\phi(t) = 1 + \int_0^t \phi(s) u(s) ds,$$

où $u(s) = (e^{f(s)} - 1 - f(s))\lambda(s)$. En résolvant l'équation différentielle, on obtient :

$$\phi(t) = \exp \left(\int_0^t (e^{f(s)} - 1 - f(s)) \lambda(s) ds \right),$$

soit en simplifiant les deux côtés par $\exp \left(- \int_0^t f(s) \lambda(s) ds \right)$:

$$\mathbf{E} \left[\exp \left(\int_0^t f(r) dN(r) \right) \right] = \exp \left(\int_0^t (e^{f(s)} - 1) \lambda(s) ds \right).$$

Réciproquement, en appliquant (6.6) à :

$$f = \sum_{i=0}^{n-1} \alpha_i \mathbf{1}_{]t_i, t_{i+1}[},$$

on obtient :

$$\begin{aligned} \mathbf{E} \left[\exp \left(- \sum_i \alpha_i (N(t_{i+1}) - N(t_i)) \right) \right] \\ &= \exp \left(- \lambda \sum_i \int_0^\infty (1 - e^{-\alpha_i \mathbf{1}_{]t_i, t_{i+1}[}(s)}) ds \right). \end{aligned}$$

Remarquons alors que la fonction $s \rightarrow (1 - e^{-\alpha_i \mathbf{1}_{]t_i, t_{i+1}[}(s)})$ s'annule en dehors de l'intervalle $]t_i, t_{i+1}[$ et vaut $1 - e^{-\alpha_i}$ sur cet intervalle. On a donc :

$$\mathbf{E} \left[\exp \left(- \sum_i \alpha_i (N(t_{i+1}) - N(t_i)) \right) \right] = \exp \left(- \sum_i (1 - e^{-\alpha_i}) \int_{t_i}^{t_{i+1}} \lambda(s) ds \right).$$

On obtient donc que la transformée de Laplace du vecteur aléatoire $(N(t_i + 1) - N(t_i), 1 \leq i \leq n-1)$ est le produit des transformées de Laplace de chaque composante (puisque qu'elle s'écrit comme produit de fonctions dépendant chacune d'un seul des α_i), donc les variables aléatoires sont indépendantes. Par classe monotone, on en déduit que $N(t + s) - N(t)$ est indépendante de $\mathcal{F}_t = \sigma(N(r), r \leq t)$. Dans le cas où $n = 2, t_1 = a, t_2 = b$ on obtient de la formule précédente que :

$$\mathbf{E} [\exp(-\alpha (N(b) - N(a)))] = \exp(-(1 - e^{-\alpha}) \int_a^b \lambda(s) \, ds),$$

donc $N(b) - N(a)$ suit une loi de Poisson de paramètre $\int_a^b \lambda(s) \, ds$. Par conséquent :

$$\mathbf{E} [N(t + s) - N(t) | \mathcal{F}_t] = \mathbf{E} [N(t + s) - N(t)] = \int_t^{t+s} \lambda(s) \, ds,$$

donc \tilde{N} est une martingale. \square

6.6. Processus de Cox

Les processus de Cox sont des processus de Poisson pour lesquels on choisit aléatoirement l'intensité. De ce qui précède, on voit que cela revient à mettre une mesure de probabilité sur les fonctions positives (de préférence cadlag).

DÉFINITION 6.8.— Soit M une variable aléatoire à valeurs dans \mathbb{D} . Le processus ponctuel N est un processus de Cox d'intensité M si et seulement si pour toute fonction f à support compact, on a :

$$\mathbf{E} \left[\exp \left(- \sum_n f(T_n) \right) \mid M \right] = \exp \left(- \int (1 - e^{-f(s)}) M(s) \, ds \right).$$

Par conséquent, pour calculer la transformée de Laplace, il faut déconditionner par rapport à M , d'où :

$$\mathbf{E} \left[\exp \left(- \sum_n f(T_n) \right) \right] = \mathbf{E} \left[\exp \left(- \int (1 - e^{-f(s)}) M(s) \, ds \right) \right].$$

Les résultats précédents de martingales s'étendent sans changement autre que celui de la filtration, qui doit ici être prise égale à $\mathcal{F}_0 = \sigma(M)$, $\mathcal{F}_t = \sigma(M) \vee \sigma(N(r), r \leq t)$.

L'exemple de processus de Cox qui nous sera le plus utile est celui des MMPP, voir la section 7.6.

6.7. Problèmes

EXERCICE 10.— Soit N un processus de Poisson d'intensité λ , on note T_n le n^{e} instant de saut. Par convention, $T_0 = 0$. Soit $(Z_n, n \geq 1)$, une suite de variables aléatoires de même loi telles que pour tout n , T_n et Z_n sont indépendantes. Soit g la densité de la loi commune aux Z_n .

1) Montrer que pour toute fonction f :

$$E[f(T_n, Z_n)] = \int_0^{+\infty} \int f(t, z)g(z)\lambda e^{-\lambda t} \frac{(\lambda t)^{n-1}}{(n-1)!} dz dt.$$

2) En déduire que :

$$E\left[\sum_{n \geq 1} f(T_n, Z_n)\right] = \lambda \int_0^{+\infty} \int f(t, z)g(z) dz dt.$$

3) On suppose que les communications téléphoniques d'un abonné durent un temps aléatoire de loi exponentielle de moyenne trois minutes. Ces durées sont indépendantes entre elles. Au siècle dernier, le coût d'une communication était fonction de sa durée t selon la formule suivante :

$$c(t) = \alpha \text{ si } t \leq t_0 \text{ et } c(t) = \alpha + \beta(t - t_0) \text{ si } t \geq t_0.$$

Déduire de ce qui précède que le coût moyen d'une heure totale de communication est donné par :

$$\lambda \int_0^{+\infty} c(t)\lambda e^{-\lambda t} dt$$

avec $\lambda = 20$. (Indication : considérer $Z_n = T_{n+1} - T_n$ et expliquer pourquoi on peut appliquer le résultat précédent.)

Application numérique : pour les communications locales, en 1999, on avait les paramètres suivants : $t_0 = 3$ minutes, $\alpha = 0,11$ euro et $\beta = 0,043$ euro par minute. Pour les communications nationales, $t_0 = 39$ secondes et $\beta = 0,17$ euro par minute. α était le même. En tarif réduit, diviser β par 2. En mettant $t_0 = 1$ minute et $\alpha = 0,15$ euro, à combien s'élevait le prix de la seconde supplémentaire en téléphonie mobile dans un forfait dont le montant pour 1 heure de communication était de 23,62 euros ?

EXERCICE 11.— Un distributeur automatique de billets enregistre les instants de début et de fin de requêtes de ces clients mais évidemment par leur heure d'arrivée dans la file. Un nouveau cycle d'activité ayant commencé à 7h30, on a relevé les données du tableau 11.

Supposons que les arrivées aient lieu selon processus de Poisson, que peut-on dire de l'instant d'arrivée T_1 du client numéro 1? En particulier, calculer son espérance.

Client numéro	Début de service	Fin de service
0	7h30	7h34
1	7h34	7h40
2	7h40	7h42
3	7h45	7h50

Tableau 6.1 –

EXERCICE 12.– Une compagnie d'assurance doit payer pour des sinistres au rythme de cinq par jour. On suppose que les instants d'occurrence des sinistres suivent un processus de Poisson et que les montants des dégâts sont indépendants les uns des autres, de loi exponentielle, de moyenne cinq cents euros. On introduit $(X_i, i \geq 1)$ une suite de variables aléatoires indépendantes équidistribuées, de loi exponentielle de moyenne $1/\mu = 3\,000$ euros et N un processus de Poisson d'intensité $\lambda = 5$ jours⁻¹, indépendant des X_i .

- 1) Que représente $Z = \sum_{i=1}^{N(365)} X_i$?
- 2) Calculer le montant total moyen dépensé par an par la compagnie d'assurance.
- 3) Calculer $E[e^{-sZ}]$.
- 4) En déduire la variance de Z .

6.8. Notes et commentaires

Pour des résultats plus approfondis sur les processus ponctuels en dimension quelconque, on pourra consulter [BRE 81, DAL 03, LAS 95]. Pour les convergences en loi de processus ponctuels, on pourra consulter l'annexe D de [ROB 03].

Résumé

- Un processus de Poisson est un processus qui représente un flux aléatoire à débit moyen constant dans le temps.
- Son intensité λ représente le nombre moyen d'arrivées par unité de temps.
- La superposition de deux processus de Poisson est un processus de Poisson. L'amincissement d'un processus de Poisson est un processus de Poisson.
- Un processus de Poisson représente bien les événements issus de l'activité humaine (téléphone, instants d'arrivée dans un magasin, ouverture de session mail, ouverture de page *web*, etc.) pas ceux de l'activité d'une machine (envoi de paquets, messages de signalisation, etc.).

Chapitre 7

Processus de Markov

La modélisation markovienne d'un système dynamique se fait naturellement en utilisant une chaîne de Markov dont les durées de séjour dans chaque état deviennent aléatoires. Cette description est insuffisante pour établir les propriétés mathématiques intéressantes. On a, pour cela, recours au formalisme des processus markoviens avec leur semi-groupe et leur générateur infinitésimal. Pour aller plus loin et prouver notamment des identités comme PASTA ou ses avatars, on a besoin de caractériser les processus de Markov comme solutions de problèmes de martingales. Dans ce chapitre, nous passons en revue ces différentes caractérisations et montrons qu'elles sont équivalentes.

Dans tout ce chapitre, E désigne un espace au plus dénombrable muni de la topologie discrète. On rappelle qu'il est conseillé de lire le petit paragraphe sur les notations de la remarque A.1.1.

7.1. Préliminaires

Nous commençons par énoncer deux lemmes techniques sur la loi exponentielle, qui nous seront utiles dans les calculs qui suivent.

LEMME 7.1.– Soit U et V , deux variables aléatoires indépendantes, de lois respectives $\varepsilon(\lambda)$ et $\varepsilon(\mu)$, où $\lambda, \mu > 0$. Les deux propriétés suivantes sont satisfaites :

- (i) $\mathbf{P}(U \leq V) = \lambda/(\lambda + \mu)$;
- (ii) $U \wedge V \sim \varepsilon(\lambda + \mu)$.

Démonstration. (i) La densité du couple de variables aléatoires (U, V) est donnée pour tout (u, v) par :

$$f_{(U,V)}(u, v) = \lambda e^{-\lambda u} \mu e^{-\mu v} \mathbf{1}_{\mathbf{R}^+}(u) \mathbf{1}_{\mathbf{R}^+}(v).$$

On notant le sous-ensemble $A = \{(u, v) \in \mathbf{R}^2; u \leq v\}$, on peut donc écrire :

$$\begin{aligned} \mathbf{P}(U \leq V) &= \int \int_A f_{(U,V)}(u, v) du dv \\ &= \int_0^{+\infty} \lambda e^{-\lambda u} \left\{ \int_u^{+\infty} \mu e^{-\mu v} dv \right\} du \\ &= \int_0^{+\infty} \lambda e^{-(\lambda+\mu)u} du = \frac{\lambda}{\lambda + \mu}. \end{aligned}$$

(ii) Il suffit de voir que, pour tout $x \geq 0$, on a :

$$\mathbf{P}(U \wedge V \geq x) = \mathbf{P}(\{U \geq x\} \cap \{V \geq x\}) = e^{-\lambda x} e^{-\mu x} = e^{-(\lambda+\mu)x}.$$

D'où le résultat. \square

LEMME 7.2 (Propriété d'oubli de la loi exponentielle).— Soit U , une variable aléatoire de loi $\varepsilon(\lambda)$, où $\lambda > 0$. Soit $t \geq 0$. Conditionnellement à $\{U \geq t\}$, la variable aléatoire $U - t$ suit aussi la loi $\varepsilon(\lambda)$.

Démonstration. Il suffit de calculer, pour tout $x \geq 0$, la probabilité conditionnelle :

$$\mathbf{P}(U \geq t + x \mid U \geq t) = \frac{\mathbf{P}(U \geq t + x)}{\mathbf{P}(U \geq t)} = \frac{e^{-\lambda(t+x)}}{e^{-\lambda t}} = e^{-\lambda x}.$$

D'où le résultat. \square

☞ Si la durée de vie de chaque humain suivait une loi exponentielle, le résultat précédent signifierait que la probabilité d'atteindre 90 ans quand on a déjà plus de 80 ans est la même que celle d'atteindre l'âge de 10 ans au moment de sa naissance. Comme il y a sensiblement moins d'octogénaires qui deviennent nonagénaires que de bébés qui atteignent le collège, on ne peut raisonnablement pas représenter la durée de vie d'un humain par une telle loi.

7.2. Construction trajectorielle

DÉFINITION 7.1.— Soit ν une mesure de probabilité sur E et $Q = (q(x, y), x, y \in E)$ une famille de réels tels que pour tout $(x, y) \in E$, les propriétés suivantes soient satisfaites :

$$q(x, y) \geq 0 \text{ et } \sum_{y \neq x} q(x, y) = 1. \quad (7.1)$$

Le processus de Markov X de paramètres (ν, Q) est ainsi construit :

- $X(0)$ est une variable aléatoire de loi ν ;
- si $X(0) = x$, on considère $\xi_1 = T_1 \sim \varepsilon(q(x, x))$ indépendante de $X(0)$;
 $X(t) = x$ pour $t < T_1$; ensuite, on considère \hat{X}_1 une variable aléatoire indépendante de $(X(0), T_1)$ telle que :

$$\mathbf{P}(\hat{X}_1 = y) = q(x, y);$$

- soit x_1 la valeur de \hat{X}_1 ; on considère $\xi_2 \sim \varepsilon(q(x_1, x_1))$ indépendante de $(X(0), \xi_1, \hat{X}_1)$. $X(t) = \hat{X}_1$ pour $T_1 \leq t < T_2$;
- on continue cette construction jusqu'à l'infini.

REMARQUE.- Une variable aléatoire de loi exponentielle de paramètre 0 doit être comprise comme égale presque sûrement à l'infini. Pour tout point $x \in E$ tel que $q(x, x) = 0$, (7.1) impose que $q(x, y) = 0$ pour tout y . Le point x est donc un point « cimetièrre », puisque l'on n'en sort jamais.

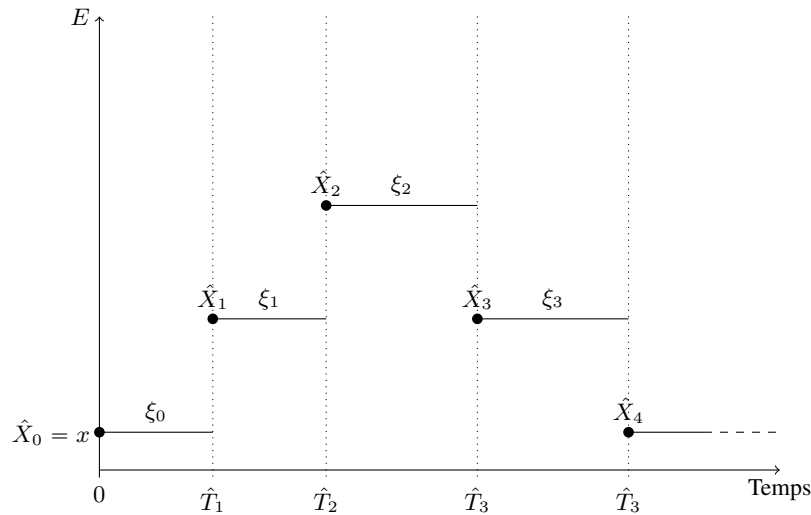


Figure 7.1 – Réalisation d’une trajectoire d’un processus de Markov de paramètres (ν, Q)

REMARQUE.- Compte tenu des hypothèses d’indépendance, il est évident que la suite $(\hat{X}_n, n \geq 0)$ est une chaîne de Markov de matrice de transition \hat{Q} définie par :

$$\hat{Q}(x, y) = \begin{cases} q(x, y) & \text{si } x \neq y \\ 0 & \text{si } x = y. \end{cases}$$

Cette chaîne de Markov s'appelle la chaîne incluse. Il ne peut pas y avoir de transition d'un état vers lui-même puisque l'on observe les positions de X seulement lors de ses changements d'état.

EXEMPLE 7.1 (File M/M/1).— Dans ce modèle, les arrivées ont lieu selon un processus de Poisson d'intensité λ et les temps de séjour sont indépendants et de loi exponentielle de paramètre μ . Le processus étudié $X = (X(t), t \geq 0)$ est le processus qui compte le nombre de clients dans le système à chaque instant. L'espace d'états E est égal à l'espace des entiers naturels. S'il y a $i \neq 0$ clients dans le système, le prochain événement est un départ ou une arrivée. En vertu du paradoxe de l'autobus (théorème 6.7), la prochaine arrivée aura lieu après une durée de loi exponentielle de paramètre λ . Par ailleurs, le théorème 7.3 ci-après nous apprend que le prochain départ aura lieu après un temps exponentiel de paramètre μ . En vertu du lemme 7.1, le temps de séjour dans l'état i suit donc une loi exponentielle de paramètre $\lambda + \mu$ donc $q(i, i) = \lambda + \mu$ pour $i \neq 0$. Si $i = 0$, il ne peut pas y avoir de départ donc $q(0, 0) = \lambda$. Toujours en vertu du lemme 7.1, la probabilité de passer de l'état i à l'état $(i + 1)$ correspond à la probabilité qu'une variable aléatoire de loi exponentielle de paramètre λ se termine avant une variable aléatoire de loi exponentielle de paramètre μ donc :

$$q(i, i + 1) = \frac{\lambda}{\lambda + \mu} \text{ et de même } q(i, i - 1) = \frac{\mu}{\lambda + \mu}.$$

Si $i = 0$, on obtient $q(0, 1) = 1$.

Il nous reste donc à démontrer notre assertion sur la loi du temps de service résiduel. Le processus des départs n'est pas de façon évidente un processus de Poisson puisqu'il n'y a pas de départs quand la file est vide. On ne peut donc pas *a priori* appliquer le paradoxe de l'autobus (voir le théorème 6.7). Néanmoins, si l'on conditionne par le fait d'être dans une période d'activité alors les départs forment un processus de Poisson. Nous formalisons cette idée dans le théorème suivant.

THÉORÈME 7.3.— Soit $R(t)$ le temps à venir entre t et le prochain départ. Pour tout $t \geq 0$, $R(t)$ suit, conditionnellement à $\{X(t) > 0\}$, la loi $\varepsilon(\mu)$.

Démonstration. On note $T'_0 < T'_1 < \dots$ les instants successifs de sortie du système et pour tout $t \geq 0$, $B_t \in \mathbf{N}^*$, l'indice du dernier client entré dans un système vide avant t . Autrement dit, T_{B_t} représente l'instant de début de la dernière période d'activité débutée avant t . A partir de là, on marque les instants de sortie du serveur par un processus ponctuel de la façon suivante :

$$\tilde{T}_0^t = T_{B_t}; \quad \tilde{T}_k^t = T'_{B_t+k-1}; \quad k > 1,$$

le k^e instant de sortie du serveur à partir du début de la dernière période d'activité débutée avant t . Pour tout $s \geq T_{B_t}$, on note également :

$$M_s^t = \sum_{k \in \mathbb{N}^*} \mathbf{1}_{\{\tilde{T}_k^t \leq s\}}$$

et pour tout $u \geq 0$:

$$\tilde{M}_u^t = M_{T_{B_t} + u}^t.$$

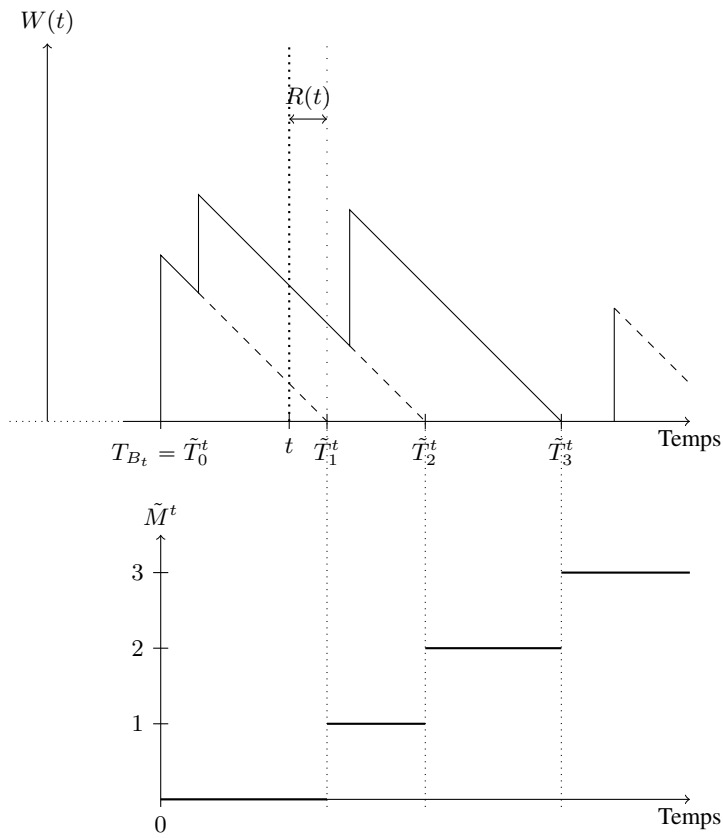


Figure 7.2 – Notations

Les notations sont un peu compliquées, mais l'idée est simple : le processus ponctuel $(\tilde{M}_u^t, u \geq 0)$ marque les instants de sortie du serveur à partir du début de la dernière période d'activité débutée avant t , instant auquel on initialise l'échelle de temps. Le

nombre de sorties du serveur entre le début de la période d'activité et t est donné par $\tilde{M}_{t-T_{B_t}}^t$. Sur $\{X(t) > 0\}$, le système n'est jamais vide entre T_{B_t} et t et pour tout $1 \leq k \leq \tilde{M}_{t-T_{B_t}}^t + 1$, on a donc :

$$\tilde{T}_k^t - \tilde{T}_{k-1}^t = \sigma_{B_t+k-1},$$

le temps de service du $k - 1^e$ client depuis le début de la période d'activité. Ces temps de service sont indépendants les uns des autres, indépendants du passé avant T_{B_t} et tous de loi $\varepsilon(\mu)$ pour $k \leq \tilde{M}_{t-T_{B_t}}^t$. Donc, conditionnellement à $\{X(t) > 0\}$, le processus $(\tilde{M}_u^t, u \geq 0)$ est égal en loi sur l'intervalle $[0, \tilde{M}_{t-T_{B_t}}^t]$ à un processus de Poisson de paramètre μ , dont $\tilde{T}_0 < \tilde{T}_1 < \dots$ représentent les points et $\tilde{\xi}_0, \tilde{\xi}_1, \dots$ les tailles des intervalles entre les points. Pour tout x , on peut donc écrire que :

$$\begin{aligned} \mathbf{P}(R(t) \geq x \mid X(t) > 0) \\ = \sum_{k \in \mathbf{N}} \mathbf{P}\left(\left\{\tilde{T}_k + \tilde{\xi}_k - (t - T_{B_t}) \geq x\right\} \cap \left\{\tilde{T}_k \leq t - T_{B_t} < \tilde{T}_{k+1}\right\}\right), \end{aligned}$$

et l'on peut poursuivre le calcul comme dans la preuve du théorème 6.7 pour conclure la preuve. \square

DÉFINITION 7.2.— Un processus de Markov X est dit irréductible (respectivement récurrent, transient) si et seulement si la chaîne incluse \hat{X} est irréductible (respectivement récurrente, transiente).

Une autre construction équivalente est possible. Elle est plus artificielle mais permet de démontrer plus facilement certaines propriétés mathématiques.

DÉFINITION 7.3.— Un processus X de paramètres (ν, Q) est régulier lorsque :

$$\|q\|_\infty = \sup_{x \in E} q(x, x) < \infty.$$

Soit X un processus de Markov régulier de paramètres (ν, Q) . On pose :

$$\tilde{q}(x, y) = \begin{cases} \frac{q(x, x)q(x, y)}{\|q\|_\infty} & \text{si } x \neq y, \\ 1 - \frac{q(x, x)}{\|q\|_\infty} & \text{si } x = y. \end{cases}$$

On considère \tilde{X} la chaîne de Markov de loi initiale ν et de matrice de transition \tilde{Q} . On considère aussi N un processus de Poisson d'intensité $\|q\|_\infty$ indépendant de \tilde{X} .

Le processus $Y(t) = \tilde{X}_{N(t)}$ a même loi que X . En effet, dans les trajectoires de ce processus, on peut rester pendant plusieurs transitions dans le même état. Soit $x \in E$ et :

$$\tilde{\tau}_{x^c} = \inf\{n > 0, \tilde{X}_n \neq x\}.$$

Conditionnellement à $\tilde{X}(0) = x$, $\tilde{\tau}_{x^c}$ est indépendante de N et suit une loi géométrique de paramètre $q(x, x)/\|q\|_\infty$. Soit $\tau_{x^c} = \inf\{t > 0, Y(t) \neq x\}$. On a alors :

$$\mathbf{P}(\tau_{x^c} \geq t | Y(0) = x) = \mathbf{P}\left(\sum_{j=1}^{\tilde{\tau}_{x^c}} \xi_j \geq t\right),$$

où les ξ_j sont les interpoints de N , donc des variables aléatoires indépendantes entre elles et indépendantes de $\tilde{\tau}_{x^c}$, de loi exponentielle de paramètre $\|q\|_\infty$. Du lemme 7.1, on déduit que le temps de séjour dans l'état x suit bien une loi exponentielle de paramètre $q(x, x)$. De plus, au moment d'un saut de \tilde{X} , on remarque que :

$$\mathbf{P}(\tilde{X}_1 = y | \tilde{X}_1 \neq \tilde{X}_0 = x) = \frac{\tilde{q}(x, y)}{1 - \tilde{q}(x, x)} = q(x, y) = \mathbf{P}(\hat{X}_1 = y | \hat{X}_0 = x).$$

REMARQUE.— On déduit aussi de cette construction que pour s fixé, il n'y a presque sûrement pas de saut : $\mathbf{P}(\Delta X(s) > 0) = 0$. En effet, on a :

$$\begin{aligned} \mathbf{P}(\Delta X(s) > 0 = 0) &\leq \mathbf{P}(\Delta N(s) = 1) \\ &= \mathbf{E}\left[\int \mathbf{1}_s(x) dN(x)\right] = \int \mathbf{1}_s(x) \|q\|_\infty dx = 0. \end{aligned}$$

Par conséquent, $\mathbf{P} \otimes d$ -presque sûrement, $X(s) = X(s^-)$. En effet, la mesure de Lebesgue ne voit pas les sauts puisqu'ils sont en nombre au plus dénombrable.

7.3. Semi-groupe markovien et générateur infinitésimal

DÉFINITION 7.4.— Soit X , un processus à valeurs dans E et à trajectoires cadlag et soit $\mathcal{F}_t = \sigma\{X_u, u \leq t\}$. Le processus X vérifie la propriété de Markov simple lorsque pour tout $t, s \geq 0$, on a :

$$\mathbf{E}[f(X(t+s)) | \mathcal{F}_t] = \mathbf{E}[f(X(t+s)) | X(t)]. \quad (7.2)$$

Le processus X est dit homogène lorsque pour tout $t \geq 0$, pour tout $x \in E$, on a :

$$\mathbf{E}[f(X(t+s)) | X(t) = x] = \mathbf{E}[f(X(s)) | X(0) = x]. \quad (7.3)$$

Soit X un processus à trajectoires cadlag satisfaisant (7.2) et (7.3). Pour $f \in l^\infty(E)$, pour $x, y \in E$, on pose :

$$P_t(x, y) = \mathbf{P}(X(t) = y \mid X(0) = x);$$

$$P_t f(x) = \sum_{y \in E} f(y) P_t(x, y) = \mathbf{E}[f(X(t)) \mid X(0) = x].$$

THÉORÈME 7.4 (Equation de Kolmogorov).— Pour tout $t \geq 0$, P_t est continu de $l^\infty(E)$ dans lui-même. Par ailleurs, pour tout $t, s \geq 0$, $P_{t+s} = P_t P_s = P_s P_t$.

REMARQUE.— Cette identité s'entend soit comme une identité entre opérateurs, c'est-à-dire que pour toute $f \in l^\infty(E)$, pour tout $x \in E$, on a :

$$P_{t+s} f(x) = P_t(P_s f)(x)$$

où comme une identité entre matrices (même si la notion de matrice de taille infinie reste floue) :

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

La famille $(P_t, t \geq 0)$ un semi-groupe d'opérateurs : tous ces opérateurs satisfont une propriété de stabilité analogue à celle d'un groupe mais chaque élément de la famille n'a pas nécessairement d'inverse pour le produit de composition.

Preuve du théorème 7.4. On remarque d'abord que d'après la définition de P_t , on a :

$$P_t \mathbf{1} = \mathbf{1} \text{ et } |P_t f| \leq P_t |f|.$$

De plus, d'après les propriétés de l'espérance conditionnelle, si $f \geq 0$ alors $P_t f \geq 0$ donc en particulier, si $f \leq g$, on a $P_t f \leq P_t g$. Par conséquent si f est bornée, $P_t f$ l'est aussi :

$$|P_t f(x)| \leq P_t |f|(x) \leq P_t(\|f\|_\infty \mathbf{1})(x) = \|f\|_\infty P_t \mathbf{1}(x) = \|f\|_\infty.$$

Pour tout $t \geq 0$, P_t est donc continu de $l^\infty(E)$ dans lui-même. Comme il est vrai que :

$$\sigma(X(0)) \vee \mathcal{F}_t = \mathcal{F}_t.$$

D'après la propriété d'emboîtement des espérances conditionnelles, on a :

$$\begin{aligned} P_{t+s} f(x) &= \mathbf{E}[f(X(t+s)) \mid X(0) = x] \\ &= \mathbf{E}[\mathbf{E}[f(X(t+s)) \mid \mathcal{F}_t] \mid X(0) = x] \\ &= \mathbf{E}[(P_s f)(X(t)) \mid X(0) = x] \\ &= P_t(P_s f)(x). \end{aligned}$$

On obtient la deuxième identité en conditionnant par \mathcal{F}_s . □

Comme le processus X a des trajectoires cadlag, le théorème de convergence dominée assure que pour tout $x \in E$, $P_t f(x)$ tend vers $f(x)$.

DÉFINITION 7.5.– On dit qu'un semi-groupe P est fortement continu lorsque pour tout $f \in l^\infty(E)$, on a :

$$\lim_{t \rightarrow 0} \sup_{x \in E} |P_t f(x) - f(x)| = 0.$$

DÉFINITION 7.6.– Soit P un semi-groupe fortement continu. Soit $\text{Dom } A$ l'ensemble des fonctions f pour lesquelles $t^{-1}(P_t f(x) - f(x))$ a une limite quand t tend vers 0. Le générateur infinitésimal A du semi-groupe P est défini sur $\text{Dom } A$ par :

$$Af(x) = \lim_{t \rightarrow 0} \frac{1}{t} (P_t f(x) - f(x)). \quad (7.4)$$

THÉORÈME 7.5.– Soit P un semi-groupe fortement continu de générateur infinitésimal A .

(i) Pour $f \in \text{Dom } A$ et $t \geq 0$, la fonction :

$$x \mapsto \int_0^t P_s f(x) \, ds$$

appartient à $\text{Dom } A$ et :

$$P_t f - f = A \int_0^t P_s f \, ds. \quad (7.5)$$

(ii) Pour $f \in \text{Dom } A$ et $t \geq 0$, la fonction $P_t f$ appartient à $\text{Dom } A$ et :

$$\frac{d}{dt} P_t f = A P_t f = P_t A f. \quad (7.6)$$

(iii) Pour $f \in \text{Dom } A$ et $t \geq 0$, l'identité suivante est satisfaite :

$$P_t f - f = \int_0^t A P_s f \, ds = \int_0^t P_s A f \, ds. \quad (7.7)$$

Démonstration. Par définition de A , il faut montrer que :

$$\frac{1}{h} (P_h - \text{Id}) \left(\int_0^t P_s f(x) \, ds \right) \text{ converge quand } h \rightarrow 0.$$

Or :

$$\begin{aligned}
 (P_h - \text{Id}) \left(\int_0^t P_s f(x) \, ds \right) &= \int_0^t P_h P_s f(x) \, ds - \int_0^t P_s f(x) \, ds \\
 &= \int_0^t P_{s+h} f(x) \, ds - \int_0^t P_s f(x) \, ds \\
 &= \int_h^{t+h} P_s f(x) \, ds - \int_0^t P_s f(x) \, ds \\
 &= \int_t^{t+h} P_s f(x) \, ds - \int_0^h P_s f(x) \, ds.
 \end{aligned}$$

Comme P est un semi-groupe fortement continu, la continuité en 0 implique la continuité en s , c'est-à-dire que pour $f \in l^\infty(E)$, l'application $(s \mapsto P_s f)$ est continue de \mathbf{R}^+ dans $l^\infty(E)$. Par conséquent, on a la limite suivante :

$$\lim_{h \rightarrow 0} \frac{1}{h} \left(\int_t^{t+h} P_s f(x) \, ds - \int_0^h P_s f(x) \, ds \right) = P_t f(x) - f(x).$$

Pour $h > 0$, on pose $A_h f = h^{-1}(P_h f - f)$. On a facilement :

$$A_h P_t f = h^{-1}(P_{t+h} f - P_t f) = P_t (h^{-1}(P_h f - f)) = P_t A_h f.$$

Comme $f \in \text{Dom } A$, $A_h f$ tend vers Af quand h tend vers 0 donc $A_h P_t f$ converge aussi, ce qui revient à dire que $P_t f$ appartient à $\text{Dom } A$ et que (7.6) est vraie.

L'identité (7.7) est une conséquence immédiate de (7.5) et (7.6). □

☞ En fait, tout cela s'interprète très facilement. Formellement, (7.6) implique que $P_t f = \exp(tA)f$. Cette écriture a un sens bien connu si A est une matrice. Le fait que A puisse être un opérateur (une matrice de « taille infinie ») oblige à quelques circonvolutions mathématiques, mais l'essentiel est là. Il est, par exemple, évident sur cette écriture que A et P_t commutent (voir (7.6)) et que $A^{-1}f = \int_0^\infty P_t f \, dt$ (penser à ce que vaut $\int_0^\infty \exp(at) \, dt$ quand a est réel négatif).

THÉORÈME 7.6.— Soit X un processus à valeurs dans E au plus dénombrable, à trajectoires cadlag et satisfaisant les propriétés (7.2) et (7.3). Le couple (ν, A) détermine complètement la loi de X . En particulier, pour tout entier $n \geq 1$, pour toutes

$f_j, j = 1, \dots, n$, fonctions bornées définies sur E et pour tout $t_1 < \dots < t_n$, on a :

$$\begin{aligned} & \mathbf{E} \left[\prod_{j=1}^n f_j(X(t_j)) \right] \\ &= \int_E P_{t_1} (f_1 P_{t_2-t_1} (f_2 \dots P_{t_{n-1}-t_{n-2}} (f_{n-1} P_{t_n-t_{n-1}} f_n)) \dots) (x) d\nu(x). \end{aligned} \quad (7.8)$$

Démonstration. En vertu du théorème 7.5, connaître P est équivalent à connaître A . Tenons pour acquise l'identité (7.8), en l'appliquant à $f_j = \mathbf{1}_{A_j}$ où A_j est un sous-ensemble quelconque de E , on voit que le terme de gauche est :

$$\mathbf{P}(X(t_1) \in A_1, \dots, X(t_n) \in A_n).$$

Les lois fini-dimensionnelles de X sont donc complètement caractérisées par (ν, A) . D'après le théorème de prolongement, on en déduit que la loi de X est bien complètement caractérisée par (ν, A) . Reste à démontrer (7.8). Pour $n = 1$, c'est la définition de P . Supposons que le résultat soit vrai pour n . Par conditionnement et d'après la définition de P , on a :

$$\begin{aligned} & \mathbf{E} \left[\prod_{j=1}^{n+1} f_j(X(t_j)) \right] \\ &= \mathbf{E} \left[\prod_{j=1}^n f_j(X(t_j)) \mathbf{E} [f_{n+1}(X(t_{n+1})) | \mathcal{F}_{t_n}] \right] \\ &= \mathbf{E} [f_1(X(t_1)) \dots f_{n-1}(X(t_{n-1})) (f_n P_{t_{n+1}-t_n} f_{n+1})(X(t_n))]. \end{aligned}$$

Comme $f_n P_{t_{n+1}-t_n} f_{n+1}$ est bornée, le résultat vient par récurrence. \square

THÉORÈME 7.7.— Soit X un processus à valeurs dans E au plus dénombrable, à trajectoires cadlag et satisfaisant les propriétés (7.2) et (7.3). Le processus X satisfait la propriété de Markov forte : pour toute fonction $F : D(\mathbf{R}^+, E) \rightarrow \mathbf{R}$ bornée, pour tout temps d'arrêt τ , on a :

$$\mathbf{E}[F(\theta_\tau X) | \mathcal{F}_\tau] = \mathbf{E}[F(X) | X(0) = X(\tau)]. \quad (7.9)$$

Démonstration. Supposons dans un premier temps que τ prenne ses valeurs dans $\mathfrak{T} = \{t_j, j \geq 1\}$ dénombrable. Les ensembles $\{\tau \leq t\}$ et $\{\tau > t\}$ appartiennent, par définition d'un temps d'arrêt, à \mathcal{F}_t , donc :

$$\{\tau = t\} = \left(\{\tau \leq t\} \cap \bigcap_{s < t, s \in \mathfrak{T}} \{\tau > s\} \right) \in \mathcal{F}_t.$$

On peut alors suivre la preuve de Markov fort (voir (3.6)) établie pour les chaînes de Markov (en remplaçant n par t_n), pour obtenir que pour toute fonction $f \in l^\infty(E)$:

$$\mathbf{E}[f(X(\tau + s)) | \mathcal{F}_\tau] = \mathbf{E}[f(X(\tau + s) | X(\tau))] = P_s f(X(\tau)). \quad (7.10)$$

Pour τ temps d'arrêt quelconque, on considère la suite de temps d'arrêt $\{\tau_n, n \geq 1\}$ définie par :

$$\tau_n = \sum_{k=0}^{\infty} \frac{k+1}{2^n} \mathbf{1}_{]k2^{-n}, (k+1)2^{-n}]}(\tau).$$

Cette suite converge en décroissant vers τ . Comme X a des trajectoires continues à droite $X(\tau_n + s)$ tend presque sûrement vers $X(\tau + s)$. Par conséquent, pour tout $A \in \mathcal{F}_\tau \subset \mathcal{F}_{\tau_n}$, par convergence dominée, on a :

$$\begin{aligned} \mathbf{E}[f(X(\tau + s)) \mathbf{1}_A] &= \lim_{n \rightarrow \infty} \mathbf{E}[f(X(\tau_n + s)) \mathbf{1}_A] \\ &= \lim_{n \rightarrow \infty} \mathbf{E}[P_s f(X(\tau_n)) \mathbf{1}_A] = \mathbf{E}[P_s f(X(\tau)) \mathbf{1}_A], \end{aligned}$$

donc (7.10) reste vraie pour τ quelconque.

Par conditionnements successifs, pour $0 < s_1 < \dots < s_k$, on a donc :

$$\begin{aligned} \mathbf{E} \left[\prod_{j=1}^k f_j(X(\tau + s_j)) | \mathcal{F}_\tau \right] &= (P_{s_k - s_{k-1}} f_k \dots P_{s_1} f_1)(X(\tau)) \\ &= \mathbf{E} \left[\prod_{j=1}^k f_j(X(\tau + s_j)) | X(\tau) \right]. \end{aligned}$$

Or, on a aussi (voir (7.8)) :

$$\mathbf{E} \left[\prod_{j=1}^k f_j(X(s_j)) | X(0) = x \right] = (P_{s_k - s_{k-1}} f_k \dots P_{s_1} f_1)(x).$$

Par conséquent :

$$\mathbf{E}[F \circ \theta_\tau | X(\tau) = x] = \mathbf{E}[F | X(0) = x],$$

pour les fonctions F de la forme $\prod_j f_j$. Par classe monotone, ce résultat reste vrai pour toutes les fonctions bornées de $D(\mathbf{R}^+, E)$ dans \mathbf{R} , d'où (7.9). \square

DÉFINITION 7.7.— Soit X un processus de Markov régulier de paramètres (ν, Q) . Pour $f \in l^\infty(E)$, on pose :

$$A_Q f(x) = q(x, x) \sum_{y \neq x} (f(y) - f(x)) q(x, y).$$

Si l'on identifie f au vecteur colonne $(f(x), x \in E)$ (après avoir ordonné les éléments de E ce qui est possible puisque E s'injecte dans \mathbf{N}), on peut réécrire l'identité précédente sous la forme d'un produit matriciel en introduisant la matrice, notée également A_Q et définie par :

$$A_Q(x, y) = \begin{cases} -q(x, x) & \text{si } x = y, \\ q(x, y)q(x, x) & \text{si } x \neq y. \end{cases} \quad (7.11)$$

On notera que $A_Q(x, y) = (A_Q \mathbf{1}_y)(x)$ pour tous x, y .

THÉORÈME 7.8.— Soit X un processus de Markov régulier de paramètres (ν, Q) . Le processus X satisfait la propriété de Markov simple (7.2) et d'homogénéité (7.3). Le semi-groupe associé est fortement continu. Son générateur infinitésimal est A_Q et son domaine est $l^\infty(E)$.

Démonstration. Partons de la deuxième construction des trajectoires de X . Dans ce cas, connaître \mathcal{F}_t revient à connaître en particulier le nombre de sauts de N avant t et la valeur de X après le dernier saut de N avant t . De par la construction même des trajectoires, ces deux quantités sont les deux seules utiles pour déterminer la suite de la trajectoire. On a donc :

$$\begin{aligned} \mathbf{E}[f(X(t+s)) | \mathcal{F}_t] &= \mathbf{E}\left[f(X(t+s)) | N(t), \tilde{X}_{N(t)}\right] \\ &= \sum_{n \geq 0} \mathbf{E}\left[f(X(t+s)) \mathbf{1}_{\{N(t+s)-N(t)=n\}} | N(t), \tilde{X}_{N(t)}\right] \\ &= \sum_{n \geq 0} \mathbf{E}\left[f(\tilde{X}_{N(t)+n}) \mathbf{1}_{\{N(t+s)-N(t)=n\}} | N(t), \tilde{X}_{N(t)}\right]. \end{aligned}$$

Comme \tilde{X} et N sont indépendants et N a des accroissements indépendants, on a :

$$\mathbf{E}[f(X(t+s)) | \mathcal{F}_t] = \sum_{n \geq 0} \tilde{Q}^{(n)} f(\tilde{X}_{N(t)}) e^{-\|q\|_\infty s} \frac{(\|q\|_\infty s)^n}{n!}.$$

Par ailleurs :

$$\begin{aligned}
& \mathbf{E}[f(X(t+s)) | X(t)] \\
&= \sum_{n \geq 0} \mathbf{E} \left[\mathbf{E} [f(X(t+s)) \mathbf{1}_{\{N(t+s)-N(t)=n\}} | N(t), X(t)] | X(t) \right] \\
&= \sum_{n \geq 0} \mathbf{E} \left[\mathbf{E} [f(\tilde{X}_{N(t)+n}) \mathbf{1}_{\{N(t+s)-N(t)=n\}} | N(t), \tilde{X}_{N(t)}] | X(t) \right] \\
&= \mathbf{E} \left[\sum_{n \geq 0} \tilde{Q}^{(n)} f(\tilde{X}_{N(t)}) e^{-\|q\|_\infty s} \frac{(\|q\|_\infty s)^n}{n!} | X(t) \right] \\
&= \sum_{n \geq 0} \tilde{Q}^{(n)} f(\tilde{X}_{N(t)}) e^{-\|q\|_\infty s} \frac{(\|q\|_\infty s)^n}{n!},
\end{aligned}$$

d'après la première partie de la preuve et puisque $X(t) = \tilde{X}_{N(t)}$. La propriété de Markov simple est donc bien vérifiée. On déduit aussi de la dernière équation que :

$$\mathbf{E}[f(X(t+s)) | X(t) = x] = \sum_{n \geq 0} \tilde{Q}^{(n)} f(x) e^{-\|q\|_\infty s} \frac{(\|q\|_\infty s)^n}{n!}.$$

Le terme de droite ne dépend pas de t donc la propriété d'homogénéité est aussi vérifiée. \square

THÉORÈME 7.9.— Soit X un processus à valeurs dans E au plus dénombrable, satisfaisant (7.2) et (7.3). Soit ν la loi de $X(0)$ et A son générateur infinitésimal. Le processus X est un processus de Markov de paramètres (ν, Q_A) où :

$$Q_A(x, y) = \begin{cases} |A(x, x)| & \text{si } y = x, \\ \frac{A(x, y)}{|A(x, x)|} & \text{si } y \neq x. \end{cases} \quad (7.12)$$

Démonstration. Posons $T_0 = 0$ et pour tout entier n :

$$\begin{cases} T_{n+1} &= \inf\{t > T_n, X(t) \neq X(T_n)\} \\ \xi_n &= T_{n+1} - T_n, \\ \hat{X}_n &= X(T_n), \end{cases}$$

avec la convention habituelle : $\inf \emptyset = \infty$.

Soit pour tout $u \geq 0$, $g(u) = \mathbf{P}(T_1 > u | X_0 = x)$. Montrons que g est solution de l'équation fonctionnelle caractéristique de la fonction exponentielle. D'après les

propriétés (7.2) et (7.3), on a :

$$\begin{aligned}
g(u+v) &= \mathbf{P}(T_1 > u+v \mid X_0 = x) \\
&= \mathbf{E} \left[\mathbf{1}_{\{X(s)=x, s \in [0, u]\}} \mathbf{1}_{\{X(t)=x, t \in [u, u+v]\}} \right] \\
&= \mathbf{E} \left[\mathbf{1}_{\{X(s)=x, s \in [0, u]\}} \mathbf{E} \left[\mathbf{1}_{\{X(t)=x, t \in [u, u+v]\}} \mid \mathcal{F}_u \right] \right] \\
&= \mathbf{E} \left[\mathbf{1}_{\{X(s)=x, s \in [0, u]\}} \mathbf{E} \left[\mathbf{1}_{\{X(t)=x, t \in [0, v]\}} \mid X_0 = x \right] \right] \\
&= \mathbf{E} \left[\mathbf{1}_{\{X(s)=x, s \in [0, u]\}} \right] g(v) \\
&= g(u)g(v).
\end{aligned}$$

Comme g est borné, on en déduit l'existence de $q(x) \geq 0$ tel que :

$$g(u) = \exp(-q(x)u).$$

Le temps de séjour dans l'état initial suit donc une loi exponentielle.

Par définition d'un temps d'arrêt, l'événement $\{T_1 > u\}$ appartient à \mathcal{F}_u . De plus, sur $\{T_1 > u\}$, $X(u) = X(0)$ donc :

$$\begin{aligned}
\mathbf{P}(\hat{X}_1 = y, T_1 > u \mid X(0) = x) &= \mathbf{E}_x \left[\mathbf{1}_{[u, \infty[}(T_1) \mathbf{E}_x \left[\mathbf{1}_{\{y\}}(\hat{X}_1) \mid \mathcal{F}_u \right] \right] \\
&= \mathbf{E}_x \left[\mathbf{1}_{[u, \infty[}(T_1) \mathbf{E}_x \left[\mathbf{1}_{\{y\}}(\hat{X}_1) \mid X_0 = x \right] \right].
\end{aligned}$$

Comme la quantité $\mathbf{P}(\hat{X}_1 = y \mid X_0 = x)$ est déterministe, elle sort de l'espérance. Par conséquent, si l'on pose :

$$\mathbf{P}(\hat{X}_1 = y \mid X(0) = x) = q(x, y),$$

on peut écrire :

$$\mathbf{P}(\hat{X}_1 = y, T_1 > u \mid X(0) = x) = q(x, y) \exp(-q(x)u). \quad (7.13)$$

On déduit de (7.13) que conditionnellement à $X(0) = \hat{X}_0$, \hat{X}_1 et T_1 sont indépendantes. On a donc bien prouvé ce que l'on voulait jusqu'au premier saut : temps de séjour de loi exponentielle dépendant de l'état initial, puis choix du nouvel état de façon indépendante de ce temps de séjour.

Supposons prouvé que pour $j \leq n-1$, pour tout $y \in E$, tout $u \geq 0$, on a l'identité :

$$\begin{aligned}
\mathbf{P}(\hat{X}_{j+1} = y, \xi_j > u \mid \hat{X}_0, \dots, \hat{X}_j, T_1, \dots, T_j) \\
= q(\hat{X}_j, y) \exp(-q(\hat{X}_j, \hat{X}_j)u). \quad (7.14)
\end{aligned}$$

Etant donné que la connaissance de $\hat{X}_0, \dots, \hat{X}_n, T_1, \dots, T_n$ est équivalente à celle de tout le passé du processus jusqu'à l'instant T_n , nous avons :

$$\begin{aligned} \mathbf{P}(\hat{X}_{n+1} = y, \xi_n > u \mid \hat{X}_0, \dots, \hat{X}_n, T_1, \dots, T_n) \\ = \mathbf{P}(\hat{X}_{n+1} = y, \xi_n > u \mid \mathcal{F}_{T_n}). \end{aligned}$$

D'après la propriété forte de Markov (voir le théorème 7.7), on peut transformer cette quantité de la manière suivante :

$$\begin{aligned} \mathbf{P}(\hat{X}_{n+1} = y, \xi_n > u \mid \mathcal{F}_{T_n}) &= \mathbf{P}(\hat{X}_{n+1} = y, \xi_n > u \mid \hat{X}_n) \\ &= \mathbf{P}_{\hat{X}_n}(\hat{X}_1 = y, \xi_0 > u) \\ &= q(\hat{X}_n, y) \exp(-q(\hat{X}_n, \hat{X}_n)u), \end{aligned}$$

d'après (7.13). La relation (7.14) est donc vérifiée au rang n . Cette relation signifie que l'on peut construire les trajectoires de X comme dans la définition 7.1.

Par identification, (7.12) découle de (7.11) et du théorème 7.8. □

EXEMPLE (File M/M/1).– En vertu de (7.11) et des considérations de l'exemple 7.1, le générateur infinitésimal du nombre de clients dans la file M/M/1 est donné pour tout $i \in \mathbf{N}$ par :

$$\begin{cases} A(i, i + 1) &= \lambda, \\ A(i, i) &= -(\lambda + \mu \mathbf{1}_{[1, +\infty[}(i)), \\ A(i, i - 1) &= \mu \text{ si } i > 0. \end{cases}$$

A cette écriture en ligne, on préfère souvent la représentation matricielle :

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

ou la représentation graphique de la figure 7.3.

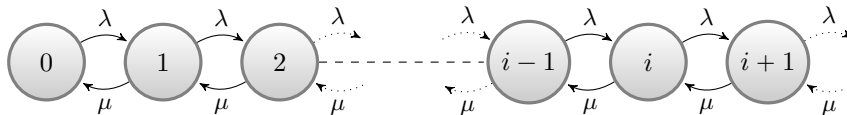


Figure 7.3 – Représentation graphique des taux de transition de la file M/M/1

THÉORÈME 7.10.– Soit X un processus de Markov régulier de paramètres (ν, Q) . On note \tilde{X} la chaîne include. Les propriétés suivantes sont équivalentes :

- \tilde{X} est irréductible;
- pour tout $x, y \in E$, il existe $t > 0$ tel que $p_t(x, y) > 0$;
- pour tout $x, y \in E$, pour tout $t > 0$, $p_t(x, y) > 0$.

Démonstration. En partant de la deuxième construction des trajectoires de X , la chaîne \tilde{X} est indépendante de N donc on a :

$$\begin{aligned} p_t(x, y) &= \mathbf{P}(X(t) = y \mid X(0) = x) \\ &= \sum_{n=0}^{\infty} \mathbf{P}(X(t) = y, N(t) = n \mid X(0) = x) \\ &= \sum_{n=0}^{\infty} \mathbf{P}(N(t) = n) \mathbf{P}(\tilde{X}_n = y \mid \tilde{X}_0 = x) \\ &= \sum_{n=0}^{\infty} e^{-\|q\|_{\infty} t} \frac{(\|q\|_{\infty} t)^n}{n!} \tilde{Q}^{(n)}(x, y). \end{aligned}$$

D'après la définition de l'irréductibilité, l'équivalence est évidente. \square

DÉFINITION 7.8.– Une mesure μ sur E est dite invariante pour le processus de Markov X lorsque $X(0) \sim \mu$ implique que $X(t) \sim \mu$ pour tout $t \geq 0$.

THÉORÈME 7.11.– Soit X un processus de Markov régulier de paramètre (ν, Q) . Une mesure μ est invariante si et seulement si elle satisfait les équations :

$$\int Af \, d\mu = 0, \text{ pour tout } f \in l^{\infty}(E). \quad (7.15)$$

En notation matricielle, cela revient à : $\mu A = 0$, où μ est le vecteur ligne $(\mu(x), x \in E)$.

Démonstration. Dire que $X(t)$ et $X(0)$ ont même loi signifie que pour tout $f \in l^{\infty}(E)$, $\mathbf{E}[f(X(t))] = \mathbf{E}[f(X(0))]$. Or :

$$\mathbf{E}[f(X(t))] = \mathbf{E}[\mathbf{E}[f(X(t)) \mid X(0)]] = \mathbf{E}[P_t f(X(0))] = \int_E P_t f(x) \, d\mu(x).$$

Donc, μ est une mesure invariante si et seulement si :

$$\int_E P_t f(x) \, d\mu(x) = \int_E f(x) \, d\mu(x).$$

Par dérivation, on en déduit (7.15). Réciproquement, si (7.15) est vraie, d'après (7.7) et le théorème de Fubini :

$$\begin{aligned} \int_E (P_t f(x) - f(x)) \, d\mu(x) &= \int_E \int_0^t P_s A f(x) \, ds \, d\mu(x) \\ &= \int_0^t \left(\int_E P_s A f(x) \, d\mu(x) \right) \, ds \\ &= 0. \end{aligned}$$

Comme E est un espace discret, l'équation (7.15) s'écrit :

$$\sum_{x \in E} \left(\sum_{z \in E} A(x, z) f(z) \right) \mu(x) = 0.$$

Pour tout $y \in E$, en prenant $f = \mathbf{1}_y$, il vient donc :

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

Dans le langage matriciel, cela signifie exactement que le produit du vecteur ligne μ par la matrice A est nul. \square

De manière analogue au cas discret, posons :

$$\tau_x^1 = \inf \{t > 0, X(t) = x\},$$

avec la convention $\tau_x^1 = \infty$ si $X(t) \neq x$ pour tout $t > 0$.

THÉORÈME 7.12.— Soit X un processus de Markov régulier irréductible récurrent. Il existe une unique mesure invariante à un facteur multiplicatif près. Cette mesure est proportionnelle à l'une des trois mesures suivantes :

(i) pour tout $y \in E$:

$$\mu(y) = \mathbf{E}_x \left[\int_0^{\tau_x^1} \mathbf{1}_{\{X(s)=y\}} \, ds \right], \quad (7.16)$$

où $x \in E$ est quelconque mais fixé ;

(ii) pour tout $y \in E$, $\mu(y) = \hat{\mu}(y)/q(y, y)$ où $\hat{\mu}$ est une mesure invariante de la chaîne incluse \tilde{X} ;

(iii) μ , solution de l'équation matricielle $\mu A = 0$.

Démonstration. Pour prouver le point (i), il faut vérifier que pour tout $t > 0$, pour tout $f \in l^\infty(E)$, l'identité suivante est satisfaite :

$$\int_E P_t f(x) \, d\mu(x) = \int_E f(x) \, d\mu(x). \quad (7.17)$$

Or, avec l'expression (7.16), on a :

$$\begin{aligned} \int_E P_t f(y) \, d\mu(y) &= \sum_{y \in E} \mathbf{E}_x \left[\int_0^{\tau_x^1} P_t f(y) \mathbf{1}_{\{X(s)=y\}} \, ds \right] \\ &= \sum_{y \in E} \mathbf{E}_x \left[\int_0^\infty P_t f(X(s)) \mathbf{1}_{\{X(s)=y\}} \mathbf{1}_{\{s < \tau_x^1\}} \, ds \right] \\ &= \mathbf{E}_x \left[\int_0^\infty \mathbf{E} [f(X(s+t)) \mid \mathcal{F}_s] \mathbf{1}_{\{s < \tau_x^1\}} \, ds \right] \\ &= \int_0^\infty \mathbf{E}_x [\mathbf{E} [f(X(s+t)) \mathbf{1}_{\{s < \tau_x^1\}} \mid \mathcal{F}_s]] \, ds \\ &= \int_0^\infty \mathbf{E}_x [f(X(s+t)) \mathbf{1}_{\{s < \tau_x^1\}}] \, ds \\ &= \mathbf{E}_x \left[\int_t^{\tau_x^1+t} f(X(s)) \, ds \right] \\ &= \mathbf{E}_x \left[\int_t^{\tau_x^1} f(X(s)) \, ds \right] + \mathbf{E}_x \left[\int_{\tau_x^1}^{\tau_x^1+t} f(X(s)) \, ds \right]. \end{aligned}$$

D'après la propriété de Markov et en remarquant que, par définition de τ_x^1 et par continuité à droite de X , on a $X(\tau_x^1) = x$, on obtient :

$$\begin{aligned} \int_E P_t f(y) \, d\mu(y) &= \mathbf{E}_x \left[\int_t^{\tau_x^1} f(X(s)) \, ds \right] + \mathbf{E}_x \left[\int_0^t f(X(s)) \, ds \right] \\ &= \mathbf{E}_x \left[\int_0^{\tau_x^1} f(X(s)) \, ds \right] \\ &= \sum_{y \in E} \mathbf{E}_x \left[\int_0^{\tau_x^1} f(X(s)) \mathbf{1}_{\{X(s)=y\}} \, ds \right] \\ &= \sum_{y \in E} \mathbf{E}_x \left[\int_0^{\tau_x^1} f(y) \mathbf{1}_{\{X(s)=y\}} \, ds \right] = \int_E f(y) \, d\mu(y). \end{aligned}$$

Pour le point (ii), rappelons qu'avec les notations du chapitre 3, on a :

$$\hat{\tau}_x^1 = \inf\{n > 0, \hat{X}_n = x\}.$$

En utilisant la première construction des trajectoires de X , on obtient :

$$\mathbf{E}_x \left[\int_0^{\hat{\tau}_x^1} \mathbf{1}_{\{X(s)=y\}} \, ds \right] = \sum_{n=1}^{\infty} \mathbf{E}_x \left[\xi_n \mathbf{1}_y(\hat{X}_{n-1}) \mathbf{1}_{\{n \leq \hat{\tau}_x^1\}} \right].$$

L'événement $\{n \leq \hat{\tau}_x^1\}$ est le complémentaire de $\{\tau_x^1 < n\} = \{\tau_x^1 \leq n-1\}$, donc il est $\hat{\mathcal{F}}_{n-1}$ mesurable (avec des notations évidentes). Par construction :

$$\mathbf{E} \left[\xi_n \mid \hat{\mathcal{F}}_{n-1} \right] = \mathbf{E} \left[\xi_n \mid \hat{X}_{n-1} \right] = \frac{1}{q(\hat{X}_{n-1}, \hat{X}_{n-1})}.$$

On en déduit que :

$$\begin{aligned} \mathbf{E}_x \left[\xi_n \mathbf{1}_y(\hat{X}_{n-1}) \mathbf{1}_{\{n \leq \hat{\tau}_x^1\}} \right] &= \mathbf{E}_x \left[\mathbf{E} \left[\xi_n \mid \hat{\mathcal{F}}_{n-1} \right] \mathbf{1}_y(\hat{X}_{n-1}) \mathbf{1}_{\{n \leq \hat{\tau}_x^1\}} \right] \\ &= \frac{1}{q(y, y)} \mathbf{P}_x(\hat{X}_{n-1} = y, \hat{\tau}_x^1 \geq n). \end{aligned}$$

Pour $\hat{\mu}$ mesure stationnaire de \hat{X} , on obtient alors :

$$\mathbf{E}_x \left[\int_0^{\hat{\tau}_x^1} \mathbf{1}_{\{X(s)=y\}} \, ds \right] = \frac{1}{q(y, y)} \sum_{n=1}^{\infty} \mathbf{P}_x(\hat{X}_{n-1} = y, \hat{\tau}_x^1 \geq n) = \frac{1}{q(y, y)} \hat{\mu}(y).$$

Le point (iii) a déjà été démontré dans le théorème 7.11. □

EXEMPLE (File M/M/1).— Pour cet exemple, la résolution du système :

$$(\pi(0), \pi(1), \dots, \pi(i), \dots) \begin{pmatrix} -\lambda & \lambda & & & \\ \mu & -(\lambda + \mu) & \lambda & (0) & \\ & & \ddots & & \\ & (0) & \mu & -(\lambda + \mu) & \lambda \\ & & & & \ddots \end{pmatrix} = 0$$

équivalent aux équations :

$$\begin{aligned} -\lambda\pi(0) + \mu\pi(1) &= 0 \\ \lambda\pi(0) - (\lambda + \mu)\pi(1) + \mu\pi(2) &= 0 \\ &\vdots \\ \lambda\pi(i-1) - (\lambda + \mu)\pi(i) + \mu\pi(i+1) &= 0 \\ &\vdots \end{aligned}$$

En additionnant ces équations deux à deux successivement, on obtient :

$$-\lambda\pi(0) + \mu\pi(1) = 0, -\lambda\pi(1) + \mu\pi(2) = 0, \dots, -\lambda\pi(i) + \mu\pi(i+1) = 0,$$

soit, pour tout entier i :

$$\pi(i+1) = \rho\pi(i) \text{ avec } \rho = \lambda/\mu.$$

Par conséquent, $\pi(i) = \rho^i\pi(0)$. On sait de l'étude de la file M/GI/1 que la file M/M/1 n'est stable (le processus comptant le nombre de clients est récurrent) que si $\rho < 1$. On retrouve bien ici sur la forme particulière des mesures invariantes, que l'on ne peut construire une probabilité invariante que si $\rho < 1$. Dans ce cas, la probabilité invariante est donnée par $\pi(i) = \rho^i(1 - \rho)$. Cela correspond à une loi géométrique de paramètre ρ , décalée de 1.

Compte tenu de la propriété de Markov forte du théorème 7.12, la preuve du résultat suivant est rigoureusement similaire à celle de son analogue dans le cas discret, le théorème 3.21.

THÉORÈME 7.13.— Soit X un processus de Markov régulier irréductible récurrent. On note π l'unique probabilité invariante. Pour tout $f \in L^1(\pi)$, on a :

$$\lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t f(X(s)) \, ds = \sum_{x \in E} f(x)\pi(x).$$

REMARQUE.— Notons que ce théorème permet d'approcher π . En effet, il suffit de mener la simulation d'une trajectoire, en suivant par exemple la construction trajectorielle du début de ce chapitre et de calculer la proportion du temps passé par le processus dans chaque état. Cela donnera pour un temps long une approximation de π qui permettra ensuite de calculer les espérances de fonctions plus générales à l'état stationnaire. Il reste à déterminer ce que « long » signifie, donc d'identifier la vitesse de cette convergence. Cela dépasse le cadre de cet ouvrage, mais n'en est pas moins un sujet particulièrement actuel de recherche.

THÉORÈME 7.14.– Soit X un processus de Markov régulier irréductible. Si X est transient alors :

$$p_t(x, x) \xrightarrow[t \rightarrow \infty]{} 0, \text{ pour tout } x \in E.$$

Si X est récurrent de probabilité invariante π :

$$p_t(x, y) \xrightarrow[t \rightarrow \infty]{} \pi(y) \text{ et } \mathbf{E}_x [\tau_x^1] = \frac{1}{\pi(x)},$$

pour tout $x \in E$.

Démonstration. Nous ne développons la preuve que dans le cas récurrent, le cas transient se traite de manière analogue. Soit $h > 0$ et $X_n^h = X(nh)$. D'après le théorème A.12 :

$$\begin{aligned} \mathbf{E} [f(X_n^h) | X_j^h, j = 0, \dots, n-1] &= \mathbf{E} [f(X_n^h) | X_{n-1}^h] \\ &= \mathbf{E} [f(X(nh)) | X((n-1)h)] = P_h f(X((n-1)h)). \end{aligned}$$

La suite X^h est donc une chaîne de Markov d'opérateur de transition P_h . En vertu du théorème 7.10, X^h est irréductible et d'après (7.17), π en est une probabilité invariante. D'après le théorème 3.25 :

$$p_{nh}(x, y) \xrightarrow[n \rightarrow \infty]{} \pi(y).$$

Comme l'on a :

$$\begin{aligned} |p_t(x, y) - p_s(x, y)| &= \left| \int_s^t A p_u(x, y) \, du \right| \\ &\leq \int_s^t |A p_u(x, y)| \, du \\ &\leq \|q\|_\infty (t - s), \end{aligned}$$

on en déduit que :

$$|p_t(x, y) - \pi(y)| \leq \|q\|_\infty |t - nh| + |p_{nh}(x, y) - \pi(y)|.$$

Soit $\epsilon > 0$, on choisit h tel que $h\|q\|_\infty < \epsilon$ et n_0 tel que :

$$n \geq n_0 \implies |p_{nh}(x, y) - \pi(y)| \leq \epsilon.$$

Si $t \geq n_0 h$, il existe n tel que $|t - nh| \leq h$ et alors $|p_t(x, y) - \pi(y)| \leq 2\epsilon$. Donc, $\lim_{t \rightarrow \infty} p_t(x, y) = \pi(y)$.

On sait que la probabilité invariante est proportionnelle à la mesure définie dans (7.16). Or :

$$\sum_{y \in E} \mathbf{E}_x \left[\int_0^{\tau_x^1} \mathbf{1}_{\{y\}}(X(s)) \, ds \right] = \mathbf{E}_x [\tau_x^1],$$

donc en appliquant (7.16) avec $y = x$, on obtient :

$$\pi(x) = \frac{1}{\mathbf{E}_x [\tau_x^1]},$$

d'où le résultat. □

7.4. Problème de martingale

DÉFINITION 7.9.— Soit ν une mesure de probabilité sur E dénombrable et A un opérateur continu de $l^\infty(E)$ dans lui-même, c'est-à-dire qu'il existe $c > 0$ tel que pour tout $f \in l^\infty(E)$, l'identité suivante soit satisfaite :

$$\|Af\|_\infty = \sup_{x \in E} |Af(x)| \leq c \|f\|_\infty.$$

On note alors :

$$\|A\|_\infty = \sup_{\|f\|_\infty=1} \|Af\|_\infty.$$

THÉORÈME 7.15.— Le processus X est un processus de Markov régulier de loi initiale ν et de générateur infinitésimal A , si et seulement si la loi de $X(0)$ est ν et pour toute fonction bornée f , le processus :

$$M^f : t \mapsto f(X(t)) - f(X(0)) - \int_0^t Af(X(s)) \, ds$$

est une martingale locale. Pour tout t , son crochet s'écrit alors :

$$\langle M^f, M^f \rangle_t = \int_0^t (Af^2(X(s)) - 2f(X(s))Af(X(s))) \, ds. \quad (7.18)$$

Démonstration. Soit $f \in l^\infty(E)$, supposons dans un premier temps que les processus :

$$t \mapsto \int_0^t Af(X(s)) \, ds \text{ et } t \mapsto \int_0^t Af^2(X(s)) \, ds$$

soient bornés. On remarque que le processus $(s \mapsto Af(X(s)))$ est adapté. Pour tout $s \geq 0$, on a :

$$\begin{aligned} & \mathbf{E} [M^f(t+s) | \mathcal{F}_t] \\ &= \mathbf{E} [f(X(t+s)) | \mathcal{F}_t] - \int_0^{t+s} \mathbf{E} [Af(X(u)) | \mathcal{F}_t] \, du \\ &= \mathbf{E} [f(X(t+s)) | X(t)] - \int_0^t Af(X(u)) \, du - \int_t^{t+s} \mathbf{E} [Af(X(u)) | \mathcal{F}_t] \, du \\ &= P_s f(X(t)) - \int_0^t Af(X(u)) \, du - \int_0^s P_u Af(X(t)) \, du. \end{aligned}$$

D'après (7.5) :

$$P_s f(X(t)) - \int_0^s P_u Af(X(t)) \, du = f(X(t)),$$

donc :

$$\mathbf{E} [M^f(t+s) | \mathcal{F}_t] = f(X(t)) - \int_0^t Af(X(u)) \, du = M^f(t),$$

ce qui signifie que M^f est une martingale.

Pour $f \in l^\infty(E)$, le processus $(t \mapsto f^2(X(t)))$ est intégrable. D'une part, d'après la première partie de la preuve on peut écrire :

$$f^2(X(t)) - f^2(X(0)) = \int_0^t Af^2(X(s)) \, ds + M^{f^2}(t). \quad (7.19)$$

D'autre part, la formule d'Itô (voir théorème A.23) pour les martingales cadlag à variation finie, donne :

$$\begin{aligned} f^2(X(t)) - f^2(X(0)) &= 2 \int_0^t f(X(s^-)) \, dM^f(s) \\ &\quad + 2 \int_0^t f(X(s^-)) Af(X(s)) \, ds + [f \circ X, f \circ X]_t. \end{aligned} \quad (7.20)$$

En comparant (7.19) et (7.20), par définition du crochet, on en déduit que :

$$[f \circ X, f \circ X]_t = \int_0^t Af^2(X(s)) \, ds - 2 \int_0^t f(X(s^-)) Af(X(s)) \, ds.$$

Comme ce processus est continu et adapté, il est prévisible, d'où (7.18). Noter que l'on peut remplacer $f(X(s^-))$ par $f(X(s))$ dans la deuxième intégrale car il n'y a pas de saut fixe en s donc ds presque sûrement $X(s^-) = X(s)$, voir la remarque 7.2.

Traisons maintenant le cas général. Pour tout entier n , posons :

$$\tau_n = \inf \left\{ t \geq 0, \int_0^t |Af(X(s))| \, ds > n \text{ ou } \int_0^t |Af^2(X(s))| \, ds > n \right\},$$

et $M_{\tau_n}^f(t) = M^f(\tau_n \wedge t)$. Notons par construction que $|M_{\tau_n}^f(t)| \leq 2\|f\|_\infty + n$ donc $M_{\tau_n}^f$ est uniformément intégrable. Par ailleurs, comme A est continu de $l^\infty(E)$ dans lui-même, τ_n tend vers l'infini donc $\{\tau_n, n \geq 1\}$ réduit M^f . On remarque enfin que pour tout temps d'arrêt τ :

$$\begin{aligned} M_\tau^f(t) &= f(X(t \wedge \tau)) - f(X(0)) - \int_0^{t \wedge \tau} Af(X(s)) \, ds \\ &= f(X^\tau(t)) - f(X^\tau(0)) - \int_0^{t \wedge \tau} Af(X^\tau(s)) \, ds, \end{aligned}$$

où $X^\tau(s) = X(\tau \wedge s)$. Par conséquent, on peut appliquer ce qui précède à $M_{\tau_n}^f$ et X^{τ_n} . Il en résulte que $M_{\tau_n}^f$ est une martingale de carré intégrable dont le crochet est donné par (7.18).

Pour la réciproque, admettons le résultat du lemme 7.16 ci-après. On va montrer que pour toute $f \in l^\infty(E)$, pour tout $t, s \geq 0$, on a :

$$\mathbf{E}[f(X(t+s)) | \mathcal{F}_t] = \lim_{n \rightarrow \infty} (\text{Id} - n^{-1}A)^{-[ns]} f(X(t)). \quad (7.21)$$

L'espérance conditionnelle sachant \mathcal{F}_t apparaît donc comme dépendant uniquement de $X(t)$ et de s d'où la propriété de Markov simple et l'homogénéité. Soit A' le générateur infinitésimal de X . D'après la première partie de la preuve :

$$t \mapsto \int_0^t (A - A')f(X(s)) \, ds$$

est une martingale nulle en 0. Qui plus est, ce processus est continu et à variation finie. Cela implique que ce processus est nul $\mathbf{P} \otimes dt$ -presque sûrement. Par ailleurs :

$$(A - A')f(x) = \lim_{t \rightarrow 0} \frac{1}{t} \int_0^t (A - A')f(X(s)) \, ds = 0,$$

donc $A = A'$. Finalement, comme A est continu de $l^\infty(E)$ dans lui-même et :

$$\sup_{x \in E} |a(x, x)| = \sup_{x \in E} |A \mathbf{1}_{\{x\}}(x)| \leq c,$$

on en déduit que X est un processus de Markov régulier.

Montrons maintenant (7.21). Quitte à localiser X , on peut toujours supposer que $E^{\lambda, f}$ est une martingale uniformément intégrable. On utilise (7.22) pour $\lambda = n$, on en déduit :

$$f(X(t)) = \mathbf{E} \left[\int_0^\infty e^{-ns} ((n \text{Id} - A)f)(X(t+s)) \, ds \mid \mathcal{F}_t \right].$$

On peut appliquer ce résultat à $(\text{Id} - \frac{1}{n}A)^{-1} f$:

$$\begin{aligned} \left(\text{Id} - \frac{1}{n}A\right)^{-1} f(X(t)) &= n\mathbf{E} \left[\int_0^\infty e^{-ns} f(X(t+s)) \, ds \mid \mathcal{F}_t \right] \\ &= \mathbf{E} \left[\int_0^\infty e^{-s} f(X(t+s/n)) \, ds \mid \mathcal{F}_t \right]. \end{aligned}$$

Par récurrence, pour tout entier k , il vient donc :

$$\begin{aligned} &\left(\text{Id} - \frac{1}{n}A\right)^{-k} f(X(t)) \\ &= \mathbf{E} \left[\int_0^\infty \dots \int_0^\infty e^{-(s_1 + \dots + s_k)} f(X(t + n^{-1}(s_1 + \dots + s_k))) \, ds_1 \dots ds_k \mid \mathcal{F}_t \right]. \end{aligned}$$

Soit $\{Z_k, k \geq 1\}$ une suite de variables aléatoires indépendantes, indépendantes de X , de loi exponentielle de paramètre 1. Pour tout réel u non nul, la loi forte des grands nombres indique que :

$$u \frac{1}{[nu]} \sum_{j=1}^{[nu]} Z_j \xrightarrow{n \rightarrow \infty} u\mathbf{E}[Z_1] = u.$$

On a alors :

$$\begin{aligned} &\int_0^\infty \dots \int_0^\infty e^{-(s_1 + \dots + s_{[nu]})} f(X(t + n^{-1}(s_1 + \dots + s_k))) \, ds_1 \dots ds_k \\ &= \mathbf{E} \left[f \left(X \left(t + u(nu)^{-1} \sum_{j=1}^{[nu]} Z_j \right) \right) \right] \xrightarrow{n \rightarrow \infty} f(X(t+u)). \end{aligned}$$

Par convergence dominée, on obtient (7.21). \square

REMARQUE.— Le calcul du crochet est intéressant en lui-même car il est à la base de nombreuses méthodes d'approximations de processus, en particulier les limites fluides (ou approximations champ-moyen) et les approximations diffusions.

LEMME 7.16.— Soit X un processus cadlag. Pour toute $f \in l^\infty(E)$, le processus :

$$M^f : t \mapsto f(X(t)) - f(X(0)) - \int_0^t Af(X(s)) \, ds$$

est une martingale locale si et seulement si pour tout $\lambda \in \mathbf{R}$, le processus :

$$E^{\lambda, f} : t \mapsto e^{-\lambda t} f(X(t)) + \int_0^t e^{-\lambda s} (\lambda f(X(s)) - Af(X(s))) \, ds$$

est une martingale locale. En particulier, si $\{\tau_n, n \geq 1\}$ réduit $E^{\lambda, f}$, on obtient l'identité suivante :

$$f(X^{\tau_n}(t)) = \mathbf{E} \left[\int_0^\infty e^{-\lambda s} (\lambda f(X^{\tau_n}(t+s)) - Af(X^{\tau_n}(t+s))) ds \mid \mathcal{F}_t \right]. \quad (7.22)$$

Démonstration. Quitte à localiser, on peut toujours supposer que les processus $E^{\lambda, f}$ et M^f sont uniformément intégrables. On pose pour tout $t \geq 0$, $U(t) = \exp(-\lambda t)$. Ce processus est continu, à variation bornée, d'après la formule d'intégration par parties, pour tout t , on a :

$$U(t)M^f(t) = \int_0^t U(s) dM^f(s) - \int_0^t M^f(s)\lambda e^{-\lambda s} ds. \quad (7.23)$$

Par hypothèse, M^f est une martingale donc par construction de l'intégrale stochastique, $\int U dM^f$ aussi. Notons maintenant que :

$$\begin{aligned} \int_0^t \int_0^s Af(X(u)) du \lambda e^{-\lambda s} ds &= \int_0^t Af(X(u)) \int_u^t \lambda e^{-\lambda s} ds du \\ &= \int_0^t Af(X(u)) (e^{-\lambda u} - e^{-\lambda t}) du. \end{aligned} \quad (7.24)$$

En reportant (7.24) dans (7.23), on voit que le processus :

$$\begin{aligned} t \mapsto U(t)M^f(t) + \int_0^t M^f(s)\lambda e^{-\lambda s} ds \\ = e^{-\lambda t} f(X(t)) + \int_0^t f(X(s))\lambda e^{-\lambda s} ds - \int_0^t Af(X(s))e^{-\lambda s} ds \end{aligned}$$

est une martingale, d'où le résultat. \square

7.5. Réversibilité et applications

Dans cette section, $(X(t), t \geq 0)$ désigne un processus de Markov homogène, régulier, irréductible, récurrent à trajectoires cadlag sur l'espace dénombrable E , d'opérateur de transition Q et de générateur infinitésimal A . Dans la suite, on dira que $(X(t), t \geq 0)$ est stationnaire s'il admet une probabilité invariante π et si la loi de $X(0)$ est π . On peut toujours supposer que $\pi(x) > 0$ pour tout $x \in E$.

DÉFINITION 7.10.— On suppose que $(X(t), t \geq 0)$ est stationnaire. Pour tout $T > 0$, le processus retourné de $(X(t), t \geq 0)$ à partir de T est le processus $(\bar{X}^T(t), t \in [0, T])$ défini pour tout $t \in [0, T]$ par :

$$\bar{X}^T(t) = X((T-t)^-) = \lim_{s \searrow t} X((T-s)).$$

Rappelons (voir la définition A.5 pour plus de détails) que $l^2(\mathbf{N}, \pi)$ est l'espace de Hilbert des suites de carré intégrable pour la mesure π .

LEMME 7.17.– Si A est le générateur infinitésimal d'un processus de Markov régulier X alors A est continu de $l^2(\mathbf{N}, \pi)$ dans lui-même. Il admet donc un adjoint \bar{A} dans $l^2(\mathbf{N}, \pi)$ défini par :

$$\bar{A}(x, y) = A(y, x) \frac{\pi(y)}{\pi(x)}.$$

Démonstration. Rappelons que $A(x, x) = -\sum_{y \neq x} A(x, y) < 0$ donc on a :

$$\sum_{y \in E} |A(x, y)| = 2|A(x, x)|.$$

Par conséquent, la mesure ν définie par :

$$\nu(y) = \frac{1}{2|A(x, x)|} |A(x, y)|, \text{ pour tout } y \in E,$$

est une mesure de probabilité sur E . Par conséquent :

$$\begin{aligned} \left(\sum_{y \in E} A(x, y) f(y) \right)^2 &\leq 4A(x, x)^2 \left(\sum_{y \in E} \frac{|A(x, y)|}{2|A(x, x)|} f(y) \right)^2 \\ &\leq 2|A(x, x)| \sum_{y \in E} |A(x, y)| |f(y)|^2 \\ &\leq 2\|A\|_\infty \sum_{y \in E} |A(x, y)| |f(y)|^2, \end{aligned}$$

d'après l'inégalité de Jensen et la régularité de A . Par conséquent :

$$\begin{aligned} \sum_{x \in E} (AF)(x)^2 \pi(x) &\leq 2\|A\|_\infty \sum_{x \in E} \sum_{y \in E} |A(x, y)| |f(y)|^2 \pi(x) \\ &= 2\|A\|_\infty \sum_{y \in E} (1 + |A(y, y)|) \pi(y) f(y)^2 \\ &\leq 2\|A\|_\infty (1 + \|A\|_\infty) \|f\|_{l^2(\pi)}^2. \end{aligned}$$

Donc, A est continu de $l^2(\mathbf{N}, \pi)$ dans lui-même. Par conséquent, l'adjoint \bar{A} de A , défini pour tous $u, v \in l^2(\mathbf{N}, \pi)$ par :

$$\langle Au, v \rangle_{l^2(\mathbf{N}, \pi)} = \langle u, Av \rangle_{l^2(\mathbf{N}, \pi)} \quad (7.25)$$

est aussi continu de $l^2(\mathbf{N}, \pi)$ dans lui-même. De plus :

$$\bar{A}(x, y) = A(y, x) \frac{\pi(y)}{\pi(x)}.$$

En effet : $\bar{A}(x, y) = (\bar{A} \mathbf{1}_{\{x\}})(y)$ donc pour $u = \mathbf{1}_{\{y\}}$, (7.25) donne :

$$(\bar{A} \mathbf{1}_{\{y\}})(x)\pi(x) = (\bar{A} \mathbf{1}_{\{x\}})(y)\pi(y), \text{ soit } A(x, y)\pi(x) = \bar{A}(y, x)\pi(y),$$

d'où le résultat. \square

THÉORÈME 7.18.— Sous les hypothèses de cette section, pour tout $T > 0$ le processus $(\bar{X}^T(t), t \in [0, T])$ est markovien et à trajectoires cadlag, de générateur infinitésimal \bar{A} .

Démonstration. Les trajectoires du processus $(\bar{X}^T(t), t \geq 0)$ sont presque sûrement continues à droite, car par existence presque sûre d'une limite à gauche en tout point pour X , on a presque sûrement :

$$\begin{aligned} \lim_{h \searrow 0} \bar{X}^T(t+h) &= \lim_{h \searrow 0} \lim_{h' \searrow 0} X(T - (t+h) - h') \\ &= \lim_{\epsilon \searrow 0} X(T - (t+\epsilon)) = X(T-t)^- = \bar{X}^T(t), \end{aligned}$$

et l'existence d'une limite à gauche se démontre de manière similaire. On prend donc la limite à gauche en $T-t$ pour rendre cadlag le processus retourné.

Compte tenu du théorème 7.15, pour tout $f \in l^\infty(E)$, nous devons prouver que :

$$t \mapsto f(\bar{X}^T(t)) - f(\bar{X}^T(0)) - \int_0^t \bar{A}f(\bar{X}^T(r)) \, dr$$

est une martingale locale. La filtration est bien évidemment celle engendrée par les trajectoires de \bar{X}^T et non celles de X , c'est-à-dire que l'on a $\bar{\mathcal{F}}_t = \sigma\{\bar{X}^T(s), s \leq t\}$. En utilisant un raisonnement de classe monotone, il faut et il suffit que l'on prouve que pour $0 \leq s_1 < \dots < s_n \leq s < t$, pour toute fonction $\varphi : E^n \rightarrow \mathbf{R}$ bornée, on a :

$$\begin{aligned} \mathbf{E} \left[\left(f(\bar{X}^T(t)) - f(\bar{X}^T(s)) - \int_s^t \bar{A}f(\bar{X}^T(r)) \, dr \right) \varphi(\bar{X}^T(s_1), \dots, \bar{X}^T(s_n)) \right] \\ = 0. \end{aligned}$$

Comme X est markovien, on peut réécrire cette identité sous la forme :

$$\begin{aligned} 0 = \mathbf{E} \left[\left(f(X(T-t)) - f(X(T-s)) - \int_s^t (\bar{A}f)(X(T-r)) \, dr \right) \right. \\ \left. \times \mathbf{E} [\varphi(\bar{X}^T(s_1), \dots, \bar{X}^T(s_n)) | X(T-s)] \right]. \end{aligned}$$

D'après le théorème A.7, il suffit que pour ψ bornée, l'on démontre :

$$\mathbf{E} \left[\left(f(X(T-t)) - f(X(T-s)) - \int_{T-t}^{T-s} (\bar{A}f)(X(r)) \, dr \right) \psi(X(T-s)) \right] = 0. \quad (7.26)$$

Or, en introduisant le semi-groupe associé à X et en se rappelant que par stationnarité, la loi de $X(T-s)$ est celle de $X(0)$, c'est-à-dire π , on obtient d'une part :

$$\begin{aligned} \mathbf{E} [(f(X(T-t)) - f(X(T-s))) \psi(X(T-s))] \\ = \int_E \psi P_{t-s} f \, d\pi - \int_E f \psi \, d\pi \end{aligned} \quad (7.27)$$

et d'autre part, d'après (7.25) et (7.5), on a :

$$\begin{aligned} \mathbf{E} \left[\int_s^t \bar{A}f(X(T-r)) \, dr \, \psi(X(T-s)) \right] &= \int_E \int_s^t P_{r-s} \psi \bar{A}f \, dr \, d\pi \\ &= \int_0^{t-s} \int_E P_r \psi \bar{A}f \, d\pi \, dr \\ &= \int_0^{t-s} \int_E A P_r \psi f \, d\pi \, dr \\ &= \int_E \int_0^{t-s} A P_r \psi \, dr \, f \, d\pi \\ &= \int_E (P_{t-s} \psi - \psi) f \, d\pi. \end{aligned} \quad (7.28)$$

En faisant la différence de (7.27) et (7.28), on obtient (7.26). \square

Remarquons en particulier que le générateur de $(\bar{X}^T(t), t \geq 0)$ ne dépend pas de T , ce qui va s'avérer très précieux dans la construction suivante.

DÉFINITION 7.11.— On dit que le processus $(X(t), t \geq 0)$ est réversible, s'il est stationnaire et pour tout $T > 0$, de même loi que son processus retourné sur $[0, T]$.

Pour tout $T > 0$, $X(0)$ et $\bar{X}^T(0)$ ont même loi π . Comme la loi d'un processus de Markov est entièrement caractérisée par son générateur et sa loi initiale, on déduit du théorème 7.18 que $(X(t), t \geq 0)$ est réversible si et seulement si pour tout $x, y \in E$, on a :

$$\pi(x)A(x, y) = \pi(y)A(y, x). \quad (7.29)$$

Cette équation, appelée de balance locale, est une propriété « miroir » : le flux de passage de x à y est le même que celui de y à x .

THÉORÈME 7.19.– Réciproquement, pour tout processus markovien X , s'il existe une probabilité π vérifiant (7.29) alors X admet π pour probabilité stationnaire.

Démonstration. Il suffit pour cela de vérifier que $\pi A = 0$. Or, pour tout $x \in E$, on a :

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

Qui plus est, X est un processus de Markov réversible. □

Une classe importante de processus markoviens, dans laquelle on pourra ranger pratiquement tous les processus rencontrés dans la suite, a automatiquement la propriété de réversibilité dès qu'il existe une probabilité invariante : ce sont les processus de naissance et de mort.

DÉFINITION 7.12.– Un processus markovien $(X(t), t \geq 0)$ homogène à valeurs dans E , où $E = \mathbf{N}$ ou $\llbracket 0, n \rrbracket$, est dit de naissance et de mort si son générateur infinitésimal est tridiagonal sur E , c'est-à-dire que pour tout $i \in E$, on a :

$$A(i, j) = 0 \text{ pour tout } j \text{ tel que } |j - i| \geq 2.$$

Cette terminologie est héritée de la dynamique des populations. Si $(X(t), t \geq 0)$ compte la taille d'une population donnée à chaque instant, les sauts de $(X(t), t \geq 0)$ ne se produisent qu'au moment d'une naissance (de i à $i + 1$) ou d'une mort (de i à $i - 1$).

THÉORÈME 7.20.– Tout processus de naissance et mort stationnaire de loi π est réversible.

Démonstration. Comme A est tridiagonale, il suffit de vérifier que :

$$\pi(i - 1)A(i - 1, i) = \pi(i)A(i, i - 1)$$

pour tout $i \in E$, ce que nous faisons par récurrence. Tout d'abord, on a :

$$\pi(0)A(0, 1) = -\pi(0)A(0, 0) = \pi(1)A(1, 0),$$

puisque $\pi A(0) = 0$. Ensuite, si $\pi(i-1)A(i-1, i) = \pi(i)A(i, i-1)$ pour un indice $i \geq 1$ tel que $i+1 \in E$, on a :

$$\begin{aligned}\pi(i)A(i, i+1) &= \pi(i)(-A(i, i) - A(i, i-1)) \\ &= -\pi(i)A(i, i) - \pi(i-1)A(i-1, i) \\ &= \pi(i+1)A(i+1, i),\end{aligned}$$

puisque $\pi A(i) = 0$. Le résultat est démontré. \square

Le théorème suivant permet de calculer très facilement de nombreuses lois stationnaires dans des cas pratiques.

THÉORÈME 7.21 (Théorème de Kelly).— Soit $(X(t), t \geq 0)$ un processus de Markov réversible sur E , de générateur infinitésimal A et de probabilité invariante π . Soit $F \subset E$. On définit, pour un certain $\alpha \geq 0$, la matrice \tilde{A} suivante sur $E \times E$:

$$\begin{aligned}\tilde{A}(x, y) &= \begin{cases} \alpha A(x, y) & \text{si } x \in F, y \in E \setminus F; \\ A(x, y) & \text{sinon pour } x \neq y; \end{cases} \\ \tilde{A}(x, x) &= - \sum_{y \neq x} \tilde{A}(x, y) \text{ pour tout } x \in E.\end{aligned}$$

Alors, le processus de Markov $(\tilde{X}(t), t \geq 0)$ de générateur \tilde{A} est réversible, de probabilité invariante $\tilde{\pi}$ donnée pour tout $i \in E$ par :

$$\pi(x) = \begin{cases} C\pi(x) & \text{si } x \in F; \\ C\alpha\pi(x) & \text{si } x \in E \setminus F, \end{cases}$$

où $C = \left(\sum_{k \in F} \pi(k) + \alpha \sum_{k \in E \setminus F} \pi(k) \right)^{-1}$ est la constante de normalisation.

Démonstration. On montre très facilement que $\tilde{\pi}$ définit bien une probabilité sur E . Il suffit de vérifier qu'elle est réversible, ce qui est immédiat en remarquant que :

$$\tilde{\pi}(x)\tilde{A}(x, y) = \alpha C\pi(x)A(x, y) = \tilde{\pi}(y)\tilde{A}(y, x)$$

pour tout $x \in F$ et $y \in E \setminus F$. \square

Si l'on pose $\alpha = 0$, on interdit donc au processus de quitter le sous-ensemble F , tout en gardant la propriété de réversibilité. Nous verrons une application de ce résultat au chapitre 9.

7.6. MMPP

Les processus de Poisson modulé, *Markov Modulated Poisson Process* (MMPP), correspondent à une classe de processus markoviens qui généralisent les processus de Poisson tout en gardant des caractéristiques calculables. Ils apparaissent naturellement dans la modélisation des processus de débordement, voir chapitre 9. On a pendant un temps pensé qu'ils pouvaient servir de modèles pour les flux de données. Même si cette approche semble tombée en désuétude, elle n'en demeure pas moins intéressante, voir le chapitre 1.

DÉFINITION 7.13.— Soit $(J(t), t \geq 0)$ un processus de Markov stationnaire à valeurs dans $E = \llbracket 1, m \rrbracket$. Soit Q son générateur infinitésimal et ν sa probabilité stationnaire. Soit λ une fonction de E dans \mathbf{R}^+ . Le processus ponctuel N est un MMPP de paramètres Q et λ si et seulement si pour toute fonction f à support compact, on a :

$$\mathbf{E} \left[\exp \left(- \int_0^t f(s) dN(s) \right) \right] = \mathbf{E} \left[\exp \left(- \int_0^t (1 - e^{-f(s)}) \lambda(J(s)) ds \right) \right].$$

☞ Un MMPP est donc un processus de Cox dont l'intensité, variable dans le temps, dépend de l'évolution d'un processus de Markov à espace d'états fini.

Cela revient à dire que lorsque le processus des phases J est dans la phase j , les arrivées du processus ponctuel N se font selon un processus de Poisson d'intensité $\lambda(j)$. Quand le processus des phases change d'état, selon la dynamique induite par son générateur infinitésimal, l'intensité du processus de Poisson des arrivées change. La figure 7.4 représente une réalisation d'un MMPP à deux phases.

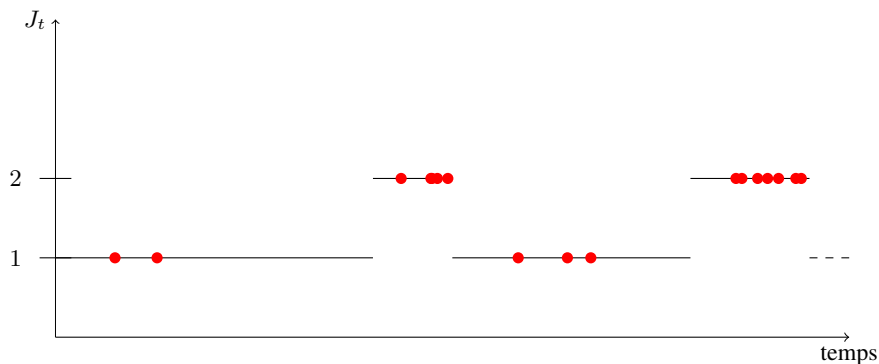


Figure 7.4 – Une réalisation d'un MMPP à deux phases

Le processus N n'est pas markovien seul car à l'instant t , le temps avant la prochaine arrivée suit, certes, une loi exponentielle, mais dont le paramètre dépend de la phase. Le processus (N, J) est markovien et son générateur s'écrit sous forme de matrices par blocs. On note Λ , la matrice diagonale dont le coefficient en (i, i) est $\lambda(i)$. Avec ces notations, le générateur infinitésimal de (N, J) est déterminé par :

$$\begin{pmatrix} Q - \Lambda & \Lambda & & & \\ 0 & Q - \Lambda & \Lambda & & \\ & & \ddots & \ddots & \ddots \\ & & & \ddots & \ddots \\ & & & & \ddots \end{pmatrix}.$$

En effet, les événements sont de deux types : arrivée ou changement de phase. Une arrivée engendre une transition d'un état (i, ϕ) vers $(i + 1, \phi)$. Un changement de phase laisse la première composante inchangée et modifie la deuxième. On en déduit un algorithme (voir 7.1) de simulation de trajectoires de MMPP. On suppose toujours que l'espace d'états est $E = \llbracket 1, m \rrbracket$.

EXEMPLE 7.2.– Un autre exemple de MMPP est donné par le processus de débordement d'une file $M_\lambda/M_\mu/S/S$. Dans un tel système, tant que les serveurs ne sont pas tous occupés, rien ne déborde. Lorsque tous les serveurs travaillent, les nouvelles arrivées sont refusées et forment le processus de débordement. Pour représenter cela sous forme d'un MMPP, il suffit de considérer comme processus des phases le processus J qui compte le nombre de serveurs occupés et la fonction λ définie par :

$$\lambda(i) = \begin{cases} 0 & \text{si } i < S, \\ \lambda & \text{si } i = S. \end{cases}$$

Pour qu'elle soit utilisable en pratique, une classe de processus doit être stable par superposition : si l'on superpose deux processus de la même classe (par exemple Poisson, Cox ou MMPP), il est souhaitable que le processus « somme » soit de la même classe. Nous savons déjà que les processus de Poisson satisfont cette propriété. Nous allons l'étendre au cas des MMPP. Cela nécessite d'introduire de nouvelles notations.

DÉFINITION 7.14.– Le produit de Kronecker de deux matrices A et B , noté $A \otimes B$, est la matrice :

$$\begin{pmatrix} A(1, 1)B & A(1, 2)B & \dots & A(1, n)B \\ \vdots & \vdots & & \vdots \\ A(n, 1)B & \dots & \dots & A(n, n)B \end{pmatrix}.$$

La somme de Kronecker est alors définie par $A \oplus B = (A \otimes \text{Id}_B) + (\text{Id}_A \otimes B)$ où Id_A est la matrice identité de même taille que A .

Algorithme 7.1 . Réalisation d'une trajectoire d'un MMPP

Données : $m, Q, \lambda, \{i_0\}, T$
Résultat : une trajectoire $(t_n, n \geq 1)$ sur $[0, T]$ d'un MMPP (m, Q, λ) d'état initial $\{i_0\}$.

pour $i = 1, \dots, m$ **faire**
 $r_{i0} = \lambda_i$;
 pour $j = 1, \dots, m$ **faire**
 $r_{i,j} = r_{i,j-1} + q(i, j)$;
 fin
fin
phase $\leftarrow i_0$;
 $t \leftarrow 0$;
 $n \leftarrow 0$;
tant que $t \leq T$ **faire**
 $x \leftarrow$ réalisation d'une ε $\left(\lambda(\text{phase}) + \sum_j q(\text{phase}, j) \right)$;
 $t \leftarrow t + x$;
 $u \leftarrow$ réalisation d'une variable aléatoire de loi uniforme sur $[0, 1]$;
 si $u \leq r_{i_0}$ **alors**
 $t_n \leftarrow t$
 sinon
 $j \leftarrow 1$;
 tant que $u > r_{i,j}$ **faire**
 $j \leftarrow j + 1$;
 fin
 phase $\leftarrow j$;
 fin
 $n \leftarrow n + 1$
fin
retourner t_1, t_2, \dots, t_n

Si A est diagonale de taille n , on peut lui associer la fonction :

$$a : \begin{cases} \llbracket 1, n \rrbracket & \longrightarrow \mathbf{R} \\ i & \longmapsto A(i, i). \end{cases}$$

Pour deux matrices diagonales A_1 et A_2 donnant lieu à deux fonctions a_1 et a_2 définies respectivement sur $\llbracket 1, n_1 \rrbracket$ et $\llbracket 1, n_2 \rrbracket$, la matrice $A_1 \oplus A_2$ est la matrice qui correspond à la fonction :

$$a_1 \oplus a_2 : \begin{cases} \llbracket 1, n_1 \rrbracket \times \llbracket 1, n_2 \rrbracket & \longrightarrow \mathbf{R} \\ (i, j) & \longmapsto A_1(i, i) + A_2(j, j) = a_1(i) + a_2(j). \end{cases}$$

THÉORÈME 7.22.— Soit $J_1 = (E_1, A_1, \nu_1)$ et $J_2 = (E_2, A_2, \nu_2)$ deux processus de Markov indépendants. Le processus « produit » $J = (J_1, J_2)$ est un processus de Markov de caractéristiques $(E_1 \times E_2, A_1 \oplus A_2, \nu_1 \otimes \nu_2)$.

Démonstration. Il est évident que l'espace d'états est $E = E_1 \times E_2$. Compte tenu du théorème 7.15, il suffit de démontrer que pour toute fonction $f \in l^\infty(E)$, le processus :

$$t \rightarrow \Theta^f(t) = f(J_1(t), J_2(t)) - \int_0^t (A_1 \oplus A_2)f(J_1(s), J_2(s)) \, ds$$

est une martingale. Supposons d'abord que $f = f_1 \otimes f_2$ avec f_1 et f_2 bornées. Puisque J_1 et J_2 sont des processus de Markov, pour $i = 1, 2$, pour tout $t \geq 0$, on peut écrire :

$$f_i(J_i(t)) = f_i(J_i(0)) + \int_0^t A_i f_i(J_i(s)) \, ds + M_i(t),$$

où M_1 et M_2 sont des martingales. La formule d'Itô donne alors :

$$\begin{aligned} & f_1(J_1(t))f_2(J_2(t)) - f_1(J_1(0))f_2(J_2(0)) \\ &= \int_0^t f_1(J_1(s))A_2f_2(J_2(s)) \, ds + \int_0^t f_2(J_2(s))A_1f_1(J_1(s)) \, ds \\ &\quad + \sum_{s \leq t} \Delta f_1(J_1(s))\Delta f_2(J_2(s)) + M(t), \end{aligned}$$

où M est une martingale. Comme J_1 et J_2 sont indépendants, presque sûrement ils n'ont aucun instant de sauts commun, donc l'avant-dernier terme est nul. On peut donc réécrire :

$$\begin{aligned} & f_1(J_1(t))f_2(J_2(t)) - f_1(J_1(0))f_2(J_2(0)) \\ &= \int_0^t (A_1 \oplus A_2)(f_1 \otimes f_2)(J_1(s), J_2(s)) \, ds + M(t). \end{aligned}$$

Le résultat est donc prouvé pour $f = f_1 \otimes f_2$. Par linéarité, pour $f = \sum_{k=1}^n f_{1k} \otimes f_{2k}$ et $\psi \in \mathcal{F}_t$ bornée, on a donc :

$$\mathbf{E} [\Theta^f(t+s)\psi] = \mathbf{E} [\Theta^f(t)\psi], \text{ pour tout } s \geq 0. \quad (7.30)$$

Pour toute fonction f bornée, il existe une suite de fonctions $(f^l, l \geq 1)$ de la forme $f^l = \sum_{k=1}^{n_l} f_{1k}^l \otimes f_{2k}^l$, qui converge uniformément vers f . Comme A_1 et A_2 sont des noyaux markoviens, $\|A_i f\|_\infty \leq \|f\|_\infty$. Par conséquent, d'après le théorème de convergence dominée, Θ^{f^l} converge dans $l^\infty(E)$ vers Θ^f . Comme ψ est bornée, (7.30)

est vraie aussi pour toute fonction f bornée, par conséquent Θ^f est une martingale. Finalement, il vient :

$$\begin{aligned}\nu_1 \otimes \nu_2(A_1 \oplus A_2) &= \nu_1 \otimes \nu_2(A_1 \otimes \text{Id}_2 + \text{Id}_1 \otimes A_2) \\ &= \nu_1 A_1 \otimes \nu_2 + \nu_1 \otimes \nu_2 A_2 \\ &= 0\end{aligned}$$

et donc $\nu_1 \otimes \nu_2$ est une mesure invariante du processus (J_1, J_2) . □

Il est intéressant d'écrire « à la main » le générateur du processus produit et de retrouver la forme $A_1 \oplus A_2$ (voir l'exercice 15).

THÉORÈME 7.23.— Soit N_i , $i = 1, \dots, K$ des MMPP indépendants tels que pour tout i , J_i est à valeurs dans $E_i = \llbracket 1, m_i \rrbracket$, de générateur infinitésimal A_i et de probabilité invariante ν_i . On note Λ_i la matrice diagonale des taux pour le MMPP N_i .

Le processus superposition N est un MMPP de processus des phases J , dont le générateur infinitésimal est donné par :

$$A = A_1 \oplus A_2 \oplus \dots \oplus A_K$$

et de fonction de taux $\lambda = \lambda_1 \otimes \lambda_2 \otimes \dots \otimes \lambda_K$. La probabilité invariante du processus des phases est $\nu_1 \otimes \dots \otimes \nu_K$.

☞ Donnons une idée intuitive de ce qui se passe dans le cas de deux MMPP. Tant qu'aucun processus de phases ne change d'état, les clients arrivent selon la superposition de deux processus de Poisson indépendants, donc selon un processus de Poisson d'intensité somme. On a donc bien sur des intervalles aléatoires, un processus de Poisson dont l'intensité est modulée par les couples de phases sous-jacents. Toutes les combinaisons de phases sont *a priori* possibles donc il y a $m_1 \cdot m_2$ phases et intensités possibles, ce que traduit le fait que λ est définie sur l'espace produit $E_1 \times E_2$. Il faudrait ensuite s'assurer que le processus des phases est bien markovien. La démonstration par les transformées de Laplace est plus abstraite mais largement plus facile que ce début de raisonnement.

Démonstration. Nous pouvons nous contenter de faire la preuve pour $K = 2$, le cas général s'en déduit par récurrence. D'après la définition d'un MMPP, pour $i = 1, 2$ et f_1, f_2 des fonctions bornées, pour tout $t \geq 0$, on a :

$$f_i(N_i(t)) = f_i(N_i(0)) + \int_0^t f_i(N_i(s)) \lambda_i(J_i(s)) \, ds + M^i(t),$$

où M^i est une martingale. Comme les processus sont indépendants, les martingales M^i sont indépendantes et donc leur crochet est nul. Par conséquent, d'après la formule d'intégration par parties (A.12), on a :

$$\begin{aligned} & f_1(N^1(t))f_2(N^2(t)) - f_1(N^1(0))f_2(N^2(0)) \\ &= \int_0^t f_2(N_2(s)) dM^1(s) + \int_0^t f_1(N_1(s)) dM^2(s) \\ &\quad + \int_0^t f_1(N_1(s))f_2(N_2(s))(\lambda_1(J_1(s)) + \lambda_2(J_2(s))) ds. \end{aligned}$$

En vertu du théorème A.32, on a prouvé que :

$$\begin{aligned} t \mapsto & (f_1 \otimes f_2)(N_1(t), N_2(t)) - (f_1 \otimes f_2)(N_1(0), N_2(0)) \\ & - \int_0^t (f_1 \otimes f_2)(N_1(s), N_2(s))(\lambda_1(J_1(s)) + \lambda_2(J_2(s))) ds \end{aligned}$$

est une martingale. Par densité, ce résultat reste vrai pour toutes les fonctions bornées sur $\mathbf{N} \times \mathbf{N}$, en particulier pour les fonctions de la forme $(n_1, n_2) \mapsto f(n_1 + n_2)$ avec f bornée de \mathbf{N} dans \mathbf{R} . Il s'ensuit que le processus $N_1 + N_2$ est un processus de Cox de compensateur $\lambda_1(J_1) + \lambda_2(J_2)$. En vertu du théorème 7.22, cela correspond bien au processus intensité associé au processus produit (J_1, J_2) . \square

THÉORÈME 7.24 (PASTA modifiée).— Soit $N = (E, J, \nu)$ un MMPP et $(\psi(t), t \geq 0)$ un processus \mathcal{F}^N prévisible borné. On a l'identité :

$$\lim_{t \rightarrow \infty} \frac{1}{N_t} \int_0^t \psi(s) dN_s = \frac{1}{\sum_{j \in E} \lambda(j)\nu(j)} \lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t \psi(s)\lambda(J(s)) ds, \quad (7.31)$$

dès lors que chacune des limites existe.

Démonstration. D'après le théorème A.37, comme ψ est bornée, on a :

$$\lim_{t \rightarrow \infty} \frac{1}{N(t)} \int_0^t \psi(s) dN(s) = \lim_{t \rightarrow \infty} \frac{t}{\int_0^t \lambda(J(s)) ds} \frac{1}{t} \int_0^t \psi(s)\lambda(J(s)) ds.$$

Comme J est un processus de Markov ergodique, on a aussi :

$$\lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t \lambda(J(s)) ds = \sum_{j \in E} \lambda(j)\nu(j),$$

d'où le résultat. \square

7.7. Problèmes

EXERCICE 13.– Soit $(J(t), t \geq 0)$ le processus de Markov à valeurs dans $\{A, B\}$ dont le générateur infinitésimal est donné par :

$$\begin{pmatrix} -\sigma_A & \sigma_A \\ \sigma_B & -\sigma_B \end{pmatrix}$$

1) Déterminer la probabilité stationnaire π de J .

2) Quelle est la loi du temps de séjour à l'état A ? On construit maintenant un processus d'arrivées de la manière suivante : pendant les périodes où $J(t) = A$ il n'y a pas d'arrivées, pendant les périodes où $J(t) = B$, les arrivées se font un processus de Poisson d'intensité λ . Notons comme d'habitude par N_t le nombre de clients qui arrivent jusqu'à l'instant t . On admet qu'il ne peut y avoir simultanément changement de phase (c'est-à-dire changement d'état du processus J et une arrivée).

3) Ecrire les coefficients non nuls du générateur infinitésimal du processus de Markov (J, N) .

4) Ce processus admet-il une probabilité stationnaire? On appelle ce processus, un processus IPP (*Interrupted Poisson Process*), c'est un modèle possible de la voix humaine (la phase A représentant les périodes de silence et la phase B les périodes de parole). On veut maintenant étudier la file $IPP/M/1/1$. On appelle $1/\mu$ le temps moyen de service. Pour ce faire, on étudie le processus (X, J) où X représente le nombre de clients dans la file. On énumère les états dans l'ordre lexicographique :

$$(0, A), (0, B), (1, A), (1, B).$$

5) Ecrire le générateur infinitésimal de (X, J) .

6) Ce processus admet-il une probabilité stationnaire? Expliciter les valeurs de trois de ses composantes en fonction de la quatrième.

7) Exprimer la probabilité stationnaire de X en fonction de celle de (X, J) . On suppose maintenant que :

$$\lambda = 3, \mu = 1, \sigma_A = 1, \sigma_B = 2.$$

8) Calculer complètement les probabilités stationnaires de (X, J) puis de X .

9) Quelle est la probabilité de blocage à l'état stationnaire?

10) Quelle est la probabilité de perte à l'état stationnaire?

EXERCICE 14.– Pour tout entier n et tout réel positif β , on rappelle que :

$$\int_0^{+\infty} x^n \beta e^{-\beta x} dx = \frac{n!}{\beta^n}.$$

On considère une file d'attente dans laquelle les temps de service sont indépendants et de même loi. Une requête a une probabilité p que son temps de service suive une loi exponentielle de paramètre μ_α et une probabilité $q = 1 - p$ que son temps de service suive une loi exponentielle de paramètre μ_β . Formellement, si X_1, X_2 sont deux variables aléatoires indépendantes de loi exponentielle de paramètres respectifs μ_α et μ_β , si Y est une variable aléatoire indépendante de X_1 et X_2 telle que $P(Y = 1) = p = 1 - P(Y = 0)$, alors la variable aléatoire $X = X_Y = X_1 \mathbf{1}_{\{Y=1\}} + X_2 \mathbf{1}_{\{Y=2\}}$ a la loi d'une durée de service.

Les arrivées forment un processus de Poisson d'intensité λ . Il y a un seul serveur et un buffer de taille K .

- 1) Montrer que le temps moyen de service est donné par :

$$p \frac{1}{\mu_\alpha} + q \frac{1}{\mu_\beta}.$$

On notera cette quantité par $1/\mu$.

- 2) Montrer que la variance du temps moyen de service est donnée par :

$$p \frac{2}{\mu_\alpha^2} + q \frac{2}{\mu_\beta^2} - \frac{1}{\mu^2}.$$

Afin d'étudier les performances de cette file d'attente, on considère le processus de Markov X qui représente le nombre de clients dans le système et la phase du client en service. L'espace d'état est donc :

$$E = \{0\} \cup \{(i, \gamma), 1 \leq i \leq K + 1, \gamma \in \{\alpha, \beta\}\}.$$

Par exemple, lorsque X est dans l'état $(5, \alpha)$, cela signifie qu'il y a quatre clients dans le buffer et que le client en service est en phase α donc que son temps de service est distribué comme une exponentielle de paramètre μ_α .

- 3) Pour $K = 0$, écrire le générateur infinitésimal de X .
- 4) Pour $K = 0$, calculer la probabilité de perte à l'état stationnaire.
- 5) Comparer au résultat obtenu pour la file M/M/1/1 pour la même charge.

On suppose dorénavant que le buffer est de taille infinie.

- 1) Quelle est la condition d'existence d'une probabilité stationnaire ? (Aucun calcul à faire !)
- 2) Quel est le nombre moyen de clients dans le système à l'état stationnaire ?
- 3) Si $\mu_\alpha = 1$ et μ est fixé, quelle est la relation entre p et μ_β ?
- 4) Représenter les variations du nombre moyen de clients quand p varie de 0 à $1/\mu$.
- 5) Ecrire les coefficients non nuls du générateur infinitésimal de X .

6) Soit π le vecteur représentant la probabilité stationnaire, on pose $x_0 = \pi(0)$, $x_i = (\pi(i, \alpha), \pi(i, \beta))$. Ecrire les équations satisfaites par les x_i en utilisant des produits matriciels par blocs. Il n'est pas demandé de résoudre le système obtenu.

EXERCICE 15.— Soit $J_1 = (E_1, A_1, \nu_1)$ et $J_2 = (E_2, A_2, \nu_2)$ deux processus de Markov indépendants. Ecrire « à la main » le générateur du processus produit $J = (J_1, J_2)$, pour retrouver la conclusion du théorème 7.22.

7.8. Notes et commentaires

Il existe de nombreux livres sur les processus de Markov à espaces d'états discrets et leurs applications notamment aux files d'attente, par ordre plus ou moins chronologique, citons de façon non exhaustive [ASM 03, CIN 75, KEL 79, KLE 76]. Ces livres n'abordent pas les problèmes de martingales à cause de leur complexité. Cet outil est néanmoins à la base de nombreux travaux notamment sur les limites fluides comme l'on en trouvera de magnifiques exemples dans [ROB 03]. Les livres de référence plus mathématiques sur ce sujet sont [BRE 81, DEL 76, ETH 86, JAC 79]. Nous avons délibérément choisi cette approche un peu plus formelle que celle utilisée dans le chapitre sur les chaînes de Markov à la fois pour son élégance mais aussi pour préparer le terrain pour les lecteurs qui voudraient se lancer dans la théorie, autrement plus ardue, des processus de Markov à espaces d'états continus.

Pour en savoir plus sur les MMPP, on pourra se référer à [FIS 93, NEU 94] et à leurs références. Noter toutefois que l'approche par les martingales développée ici est originale.

Résumé

– Un processus de Markov à valeurs dans un espace d'états discret est un processus dont les trajectoires sont faites de sauts régulés par une chaîne de Markov et dont les durées de séjour dans chaque état sont de loi exponentielle. Le paramètre de la loi exponentielle dépend de l'état. Entre deux sauts, le processus n'évolue pas.

– Son générateur infinitésimal A est une matrice éventuellement infinie dont les coefficients s'interprètent par :

- $A(x, x)$ est l'opposé du paramètre de la loi exponentielle qui régit le temps de séjour dans l'état x . Attention au signe moins !

- $A(x, y)/|A(x, x)|$ est la probabilité que le processus aille en y lorsqu'il quitte x .

– La nature des états (récurrents, transients) du processus de Markov est la même que la nature des états de la chaîne de Markov incluse.

– La probabilité invariante π s'obtient en résolvant les équations $\pi A = 0$ et $\sum_{x \in E} \pi(x) = 1$ où π est écrite sous forme de vecteur ligne.

– La probabilité stationnaire par une multiplication de tous les coefficients du générateur infinitésimal par un même réel positif. La dynamique change : les temps dans chaque état sont plus courts, les transitions restent les mêmes. En modélisation de files d'attente, on tire profit de cela en choisissant la durée moyenne de service comme l'unité de temps, c'est-à-dire $\mu = 1$.

– Si le processus est irréductible, récurrent, de probabilité invariante π , on a :

$$\frac{1}{t} \int_0^t f(X(s)) \, ds \xrightarrow{t \rightarrow \infty} \sum_{x \in E} f(x) \pi(x)$$

$$\mathbf{P}(X(t) = y \mid X(0) = x) \xrightarrow{t \rightarrow \infty} \pi(y), \forall x, y \in E.$$

– Le temps de retour à un point se déduit de la probabilité invariante :

$$\mathbf{E}_x [\tau_x^1] = \frac{1}{\pi(x)}.$$

– Les trajectoires d'un processus de Markov se simulent en utilisant la construction de la définition 7.1.

Chapitre 8

Systemes à attente

Tous les systemes de telecommunications reels sont des processus à pertes puisque toutes les memoires ont des capacites finies, donc limitees. On entend par systeme à attente, un systeme dans lequel le dimensionnement est tel que les pertes par debordement sont negligeables et pour lesquels le critere pertinent d'evaluation des performances est le temps d'attente.

Avant de nous lancer dans l'etude detaillee de ces systemes, il est important de mentionner une formule bien connue sous le nom de formule de Little.

8.1. Formule de Little

On considere un systeme à attente dans lequel les clients arrivent à des instants $(T_n, n \geq 1)$, restent un temps $(W_n, n \geq 1)$ et en sortent aux dates $(D_n = T_n + W_n, n \geq 1)$. On note N , le processus ponctuel des entrees et X le processus qui compte le nombre de clients presents dans le systeme à t . A l'instant 0, le systeme est suppose vide, donc $X(0) = 0$. L'important est que le systeme soit conservatif : autant de travail en sortie qu'en entree.

THEOREME 8.1 (Formule de Little).— On suppose que N est asymptotiquement lineaire, c'est-à-dire qu'il existe $\lambda > 0$ tel que :

$$\frac{N(t)}{t} \xrightarrow{t \rightarrow \infty} \lambda, \text{ p.s.}$$

On suppose que la suite W est ergodique, c'est-à-dire que :

$$\frac{1}{n} \sum_{j=1}^n W_j \xrightarrow{n \rightarrow \infty} W, \text{ p.s.}$$

Sous ces hypothèses, on a :

$$X = \lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t X(s) \, ds = \lambda W.$$

L'existence de cette dernière limite est montrée dans la preuve.

Démonstration. Fixons $t \geq 0$ et appliquons la formule d'intégration par parties (A.12) aux processus X et $t \mapsto t$. Comme le deuxième est continu, il n'y a pas de parties « crochet » et l'on a :

$$tX(t) = \int_0^t X(s) \, ds + \int_0^t s \, dX(s).$$

Or les sauts de X sont ceux de N et D , le processus des départs. Par conséquent :

$$\begin{aligned} \int_0^t X(s) \, ds &= \int_0^t (t-s) \, dN(s) - \int_0^t (t-s) \, dD(s) \\ &= \sum_{T_n \leq t} (t-T_n) - \sum_{T_n+W_n \leq t} (t-T_n-W_n). \end{aligned}$$

Il faut maintenant distinguer les clients qui sont entrés puis sortis avant t , de ceux qui sont entrés avant t mais pas encore sortis à t . D'où :

$$\begin{aligned} \int_0^t X(s) \, ds &= \sum_{T_n+W_n \leq t} (t-T_n - (t-T_n-W_n)) + \sum_{T_n+W_n > t, T_n \leq t} (t-T_n) \\ &= \sum_{T_n+W_n \leq t} W_n + \sum_{T_n+W_n > t, T_n \leq t} (t-T_n). \end{aligned}$$

Pour les instants T_n tels que $(T_n + W_n > t, T_n \leq t)$, on a évidemment $0 \leq t - T_n \leq W_n$ donc :

$$\sum_{T_n+W_n \leq t} W_n \leq \int_0^t X(s) \, ds \leq \sum_{T_n+W_n \leq t} W_n + \sum_{T_n+W_n > t, T_n \leq t} W_n = \sum_{T_n \leq t} W_n.$$

Par définition de $N(t)$, on a :

$$\sum_{T_n \leq t} W_n = \sum_{n=1}^{N(t)} W_n,$$

donc :

$$\frac{1}{t} \sum_{T_n \leq t} W_n = \frac{N(t)}{t} \frac{1}{N(t)} \sum_{n=1}^{N(t)} W_n \xrightarrow{t \rightarrow \infty} \lambda W,$$

d'après les deux hypothèses. Il reste à prouver que l'on a la même limite pour la borne inférieure. Pour cela, il faut contrôler le nombre de clients entrés avant t et pas encore sortis à cet instant. Remarquons que :

$$\begin{aligned} \frac{W_n}{n} &= \frac{1}{n} \left(\sum_{j=1}^n W_j - \sum_{j=1}^{n-1} W_j \right) \\ &= \frac{1}{n} \sum_{j=1}^n W_j - \frac{n-1}{n} \frac{1}{n-1} \sum_{j=1}^{n-1} W_j \\ &\xrightarrow{n \rightarrow \infty} 0, \text{ p.s.} \end{aligned}$$

Par conséquent :

$$\frac{W_n}{T_n} = \frac{W_n}{n} \frac{N(T_n)}{T_n} \xrightarrow{n \rightarrow \infty} 0 \cdot \lambda = 0, \text{ p.s.}$$

Raisonnons à trajectoire fixée. Pour tout $\varepsilon > 0$, il existe m (dépendant de la trajectoire) tel que $W_n \leq \varepsilon T_n$ pour $n \geq m$. Pour $n \geq m$, $T_n + W_n \leq (1 + \varepsilon)T_n$. Si $n \geq m$ et $T_n \leq t(1 + \varepsilon)^{-1}$ alors le client n est sorti à l'instant t donc :

$$\sum_{T_n + W_n \leq t} W_n \geq \sum_{j=m}^{N(t(1+\varepsilon)^{-1})} W_j = \sum_{j=1}^{N(t(1+\varepsilon)^{-1})} W_j - \sum_{j=1}^{m-1} W_j.$$

Le même raisonnement que pour la borne supérieure donne :

$$\liminf_{t \rightarrow \infty} \sum_{T_n + W_n \leq t} W_n \geq \lambda W (1 + \varepsilon)^{-1}.$$

Comme ce résultat est vrai pour tout $\varepsilon > 0$, on en déduit qu'il est encore vrai pour $\varepsilon = 0$. D'après le célèbre théorème des « gendarmes », on en déduit le résultat. \square

EXEMPLE.— Prenons comme système le serveur d'une file à un seul serveur. Le temps de séjour y est égal au temps de service. Comme il y a 0 ou 1 client dans le serveur, le nombre moyen de clients correspond au taux d'occupation du serveur noté τ . De la formule de Little, on déduit $\tau = \lambda \times 1/\mu$. En d'autres termes, la charge est bien le pourcentage du temps où le serveur est occupé.

EXEMPLE.— Revenons sur l'équation (1.2) associée à la figure 1.3. On peut considérer que l'on a un système où les clients rentrent aux instants T_n et sortent aux instants $T_n + Y_n$. Le temps de traversée du client n est donc Y_n . D'après la loi forte des grands nombres, on a :

$$\frac{T_n}{n} = \frac{1}{n} \sum_{j=1}^n (X_j + Y_j) \xrightarrow{n \rightarrow \infty} \frac{1}{\mu} + \frac{1}{\tau}, \text{ presque sûrement.} \quad (8.1)$$

Si l'on note N le nombre d'arrivées entre 0 et t , on a évidemment :

$$T_n \leq t < T_{n+1} \iff N(t) = n,$$

donc :

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

Comme $N(t)$ tend vers l'infini presque sûrement, le théorème « dit des gendarmes » et (8.1) impliquent que $N(t)/t$ tend vers $1/(1/\mu + 1/\tau) = \lambda$. Par construction, Q représente le pourcentage du temps où le serveur est actif, d'après la formule de Little on obtient : $Q = \lambda/\mu$. On a donc bien montré que dans ce cas, la charge est bien égale au produit du nombre moyen d'arrivées par le temps moyen de traitement.

REMARQUE.— On remarque que dans la démonstration de la formule de Little, il n'est nullement fait mention de la discipline de service. Il s'ensuit que le temps moyen d'attente est le même dans la discipline FIFO que dans la discipline LIFO. Ce phénomène pour le moins paradoxal signifie seulement que le temps moyen d'attente n'est qu'une mesure bien pauvre en renseignements sur le comportement du système. En revanche, elle est souvent très facile à calculer. En fait, la différence entre les disciplines de service se verra sur la variance et plus généralement sur la loi du temps d'attente.

8.2. File à un serveur

THÉORÈME 8.2.— Pour la file d'attente $M_\lambda/M_\mu/1/\infty$ -FIFO, on pose $\rho = \lambda/\mu$. La condition de stabilité au sens du théorème 4.2 est donnée par $\rho < 1$. Dans ce cas, la probabilité invariante π est définie par :

$$\pi(0) = 1 - \rho, \pi(n) = \rho^n(1 - \rho), n \geq 1.$$

Le nombre moyen de clients dans le système à l'état stationnaire est $\rho(1 - \rho)^{-1}$.

Démonstration. On a vu dans l'exemple 7.1 que le processus $X = (X(t), t \geq 0)$ désignant le nombre de clients dans le système (aussi appelé processus de congestion du système) est markovien de générateur infinitésimal :

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

L'unique probabilité stationnaire π du processus X sur \mathbf{N} (si elle existe) vérifie le système matriciel :

$$\pi A = 0 \text{ et } \pi \cdot \mathbf{e} = 1, \quad (\text{S})$$

où $\mathbf{e} = (1, 1, \dots, 1)$ et $x.y$ le produit scalaire euclidien de deux vecteurs lignes x et y . La première égalité du système (S) équivaut aux équations :

$$\left\{ \begin{array}{l} -\lambda\pi(0) + \mu\pi(1) = 0 \\ \lambda\pi(0) - (\lambda + \mu)\pi(1) + \mu\pi(2) = 0 \\ \vdots \\ \lambda\pi(i-1) - (\lambda + \mu)\pi(i) + \mu\pi(i+1) = 0 \\ \vdots \end{array} \right.$$

En additionnant ces équations deux à deux successivement, on obtient :

$$-\lambda\pi(0) + \mu\pi(1) = 0, -\lambda\pi(1) + \mu\pi(2) = 0, \dots -\lambda\pi(i) + \mu\pi(i+1) = 0,$$

soit, pour tout entier i ,

$$\pi(i+1) = \rho\pi(i) \text{ avec } \rho = \lambda/\mu.$$

Par conséquent, $\pi(i) = \rho^i \pi(0)$. Il s'ensuit alors de la deuxième égalité de (S) que sous la condition :

$$\rho < 1, \quad (8.2)$$

on a :

$$\pi(0) = \frac{1}{\sum_{i \in \mathbf{N}} \rho^i} = 1 - \rho \text{ et donc } \pi(i) = \rho^i (1 - \rho), \text{ pour tout } i \geq 1.$$

Cela signifie, en d'autres termes, qu'à la condition (8.2) qui est, d'après le théorème 4.2, la condition de stabilité du système et donc de récurrence de 0 pour X , l'unique probabilité stationnaire de X est la loi d'une variable aléatoire $X_\infty = Z - 1$, où Z suit la loi géométrique de paramètre $1 - \rho$.

En particulier, le nombre de client moyen dans le système à l'équilibre est donné par :

$$\mathbf{E}[X_\infty] = \mathbf{E}[Z - 1] = \frac{1}{1 - \rho} - 1 = \frac{\rho}{1 - \rho}. \quad (8.3)$$

□

On suppose dorénavant que la condition de stabilité (8.2) est satisfaite.

Pour ce système, une métrique de performance intéressante est la proportion de client entrant dans un système vide ou, ce qui revient au même comme nous allons le

voir, la proportion du temps pendant laquelle le serveur est au repos. En rappelant que l'on note $T_1 < T_2 < \dots$, les instants d'entrée des clients dans la file, la limite en temps long de la proportion de clients entrant dans un système vide est donnée par :

$$\begin{aligned}
 \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \mathbf{1}_{\{0\}}(X(T_n^-)) &= \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \mathbf{1}_{\{0\}}(X(t^-)) dt \\
 &= \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \mathbf{1}_{\{0\}}(X(t)) dt \\
 &= \mathbf{E}[\mathbf{1}_{\{0\}}(X_\infty)] \\
 &= \pi(0) = 1 - \rho,
 \end{aligned} \tag{8.4}$$

où la première égalité presque sûre découle de la propriété PASTA pour la fonction mesurable $\mathbf{1}_{\{0\}}$ (voir le théorème A.38), la deuxième du fait que X est presque sûrement à trajectoires cadlag et la troisième de l'ergodicité du processus X (théorème 3.21). Par exemple, un système de charge 1/2 fonctionne donc la moitié du temps et sur un long intervalle de temps, la moitié des clients trouvent à leur entrée un système vide.

8.2.1. Temps d'attente, temps de séjour

Intéressons-nous au temps d'attente (temps passé dans la salle d'attente) et au temps de séjour (temps total passé dans le système) des clients dans le système. Nous reprenons les notations du chapitre 5 et notons pour tout $n \geq 1$, TA_n le temps d'attente du client C_n (qui coïncide, en FIFO, avec la quantité de travail du serveur à l'entrée de C_n étudiée dans le chapitre 4). D'après le théorème 4.2, à la condition (8.2) il existe, pour toute discipline de service admissible, un temps d'attente stationnaire TA et par voie de conséquence, un temps de séjour stationnaire TS dans le système (attente + service), donné par $\text{TS} = \text{TA} + \sigma$. La formule de Little nous apprend alors que :

$$\mathbf{E}[\text{TS}] = \frac{\mathbf{E}[X_\infty]}{\lambda} = \frac{1}{\mu - \lambda}, \tag{8.5}$$

avec (8.3). On en déduit alors que :

$$\mathbf{E}[\text{TA}] = \mathbf{E}[\text{TS}] - \mathbf{E}[\sigma] = \frac{1}{\mu - \lambda} = \frac{\rho}{\mu - \lambda}. \tag{8.6}$$

Alors même que la congestion moyenne (8.3) ne dépend que de la charge, comparons les temps d'attente et de séjour des deux systèmes de même charge $M_\lambda/M_\mu/1$ et $M_{\alpha\lambda}/M_{\alpha\mu}/1$, où α est un paramètre positif. En ajoutant l'exposant α aux paramètres du deuxième système, on déduit de (8.5) et (8.6) que :

$$\mathbf{E}[\text{TS}^\alpha] = \frac{1}{\alpha} \mathbf{E}[\text{TS}] \quad \text{et} \quad \mathbf{E}[\text{TA}^\alpha] = \frac{1}{\alpha} \mathbf{E}[\text{TA}].$$

Ainsi, les temps d'attente et de séjour moyens sont par exemple divisés par deux dès que les temps de service et d'interarrivée sont divisés par deux. Les paramètres de sortie conservent ainsi le facteur d'échelle des paramètres d'entrée.

Rappelons que l'on peut en fait obtenir explicitement la loi du temps d'attente et du temps de séjour stationnaires à travers leur transformée de Laplace. Il suffit pour cela de spécialiser les résultats du théorème 5.6 au cas où les temps de service suivent la loi $\varepsilon(\mu)$. On a donc le théorème suivant.

THÉORÈME 8.3.— Le temps d'attente stationnaire T_A et le temps de séjour stationnaire T_S dans une file M/M/1 admettent respectivement pour transformée de Laplace les fonctions définies pour tout s par :

$$\mathbf{E}[e^{-sT_A}] = (1 - \rho) \left(\frac{\mu + s}{\mu(1 - \rho) + s} \right), \quad (8.7)$$

$$\mathbf{E}[e^{-sT_S}] = (1 - \rho) \left(\frac{\mu}{\mu(1 - \rho) + s} \right). \quad (8.8)$$

La propriété PASTA permet aussi de montrer que la loi stationnaire de la chaîne incluse coïncide avec celle du processus.

LEMME 8.4.— Sous la condition $\rho < 1$, la chaîne de Markov $\hat{X} = (X(T_n^-), n \geq 0)$ est irréductible, récurrente positive de loi stationnaire π .

Démonstration. Il est évident que la chaîne est irréductible et récurrente, soit $\hat{\pi}$ sa loi invariante. D'après le théorème 7.13, pour $f \in l^\infty(\mathbf{N})$, on a :

$$\frac{1}{N} \sum_{n=1}^N f(X(T_n^-)) \xrightarrow{N \rightarrow \infty} \int_{\mathbf{N}} f \, d\hat{\pi}.$$

Comme $N(t)$ est à valeurs dans \mathbf{N} et tend vers l'infini quand t tend vers l'infini ; par extraction, on a :

$$\frac{1}{N(t)} \sum_{n=1}^{N(t)} f(X(T_n^-)) \xrightarrow{t \rightarrow \infty} \int_{\mathbf{N}} f \, d\hat{\pi}.$$

Or d'après le théorème A.38, il vient :

$$\begin{aligned} \lim_{t \rightarrow \infty} \frac{1}{N(t)} \sum_{n=1}^{N(t)} f(X(T_n^-)) &= \lim_{t \rightarrow \infty} \frac{1}{N(t)} \int_0^t f(X(s^-)) \, dN(s) \\ &= \lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t f(X(s^-)) \, ds = \int_{\mathbf{N}} f \, d\pi. \end{aligned}$$

On en déduit que pour toute fonction bornée f , on a :

$$\int_{\mathbf{N}} f d\pi = \int_{\mathbf{N}} f d\hat{\pi}.$$

Par identification, il s'ensuit que $\hat{\pi} = \pi$. □

REMARQUE.— L'installation d'un lien est un investissement qu'il faut amortir. Par conséquent, le lien doit être utilisé le plus possible, c'est-à-dire supporter une charge la plus proche possible de 1. Malheureusement, d'après (8.6), une telle charge induit des délais importants et donc un service de qualité limitée. Heureusement, dans la vraie vie, les liens sont chargés à 10 % de leur capacité...

REMARQUE.— Un même volume d'informations peut être transmis par des paquets de taille moyenne m émis à un taux λ ou par des paquets de taille moyenne cm émis à un taux λ/c . La charge, ρ , sera la même et donc le nombre moyen de clients en attente dans la file sera le même. Par conséquent, la taille des mémoires nécessaires (les mémoires physiques n'étant pas infinies) sera d'autant plus grande que la taille moyenne des paquets est grande. Cela n'est pas une évidence *a priori*, en effet le volume total d'octets transmis est le même ! A charge constante, la taille moyenne des paquets a aussi une influence sur le délai de séjour. En effet, $E[\text{TS}] = 1/\mu(1 - \rho)$. Transmettre des informations dans des grands paquets impose donc des délais de transmission grands. Il semblerait alors que la solution soit de choisir de petits paquets. Malheureusement, les petits paquets ont aussi leur côté négatif. Un paquet est constitué d'informations utiles et d'informations de contrôle. Comme les informations de contrôle sont, en général, de taille constante pour un protocole donné, limiter la taille des paquets revient alors à limiter le champ d'informations utiles. Par conséquent, plus les paquets sont petits, plus l'efficacité (définie comme le rapport entre les informations utiles transmises et la taille totale du paquet) diminue.

8.3. File à plusieurs serveurs

Nous abordons maintenant la file markovienne à S serveurs ($S \geq 1$) et à salle d'attente de capacité illimitée, noté $M_\lambda/M_\mu/S/\infty$ -FIFO avec la nomenclature habituelle. Ce modèle est représenté par le processus $(X^{(S)}(t), t \geq 0)$ comptant le nombre total de clients dans le système, à valeurs dans \mathbf{N} . Ce processus est naturellement markovien, nous déterminons ci-après sa matrice de transition :

– pour tout $i \in \llbracket 1, S - 1 \rrbracket$, si le processus est dans l'état i il y a i clients dans le système qui sont tous en service. Donc le temps de séjour en i (que nous noterons Y_i pour tout i) est le minimum d'une variable aléatoire $U \sim \varepsilon(\lambda)$ comptant le temps d'interarrivée en cours et de i variables aléatoires indépendantes et identiquement distribuées V_1, V_2, \dots, V_i de loi $\varepsilon(\mu)$, comptant les temps de service résiduels des i clients en service. Ensuite, le processus saute en $i - 1$ si $Y_i = V_k$ pour un certain k et en $i + 1$ si $Y_i = U$;

Sous la condition de stabilité $\rho < S$, on peut déduire du théorème 8.5, comme en (8.4), la limite du taux de clients qui doivent attendre pour être servis. Celle-ci est donnée par la formule d'Erlang-C :

$$\begin{aligned} \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \mathbf{1}_{\{[S, +\infty)\}} \left(X^{(S)}(T_n -) \right) &= \mathbf{E} \left[\mathbf{1}_{\{[S, +\infty)\}} \left(X_{\infty}^{(S)} \right) \right] \\ &= \sum_{i=S}^{+\infty} \pi^{(S)}(i) = \frac{\pi^{(S)}(0)}{S!S^{-S}} \sum_{i=1}^{+\infty} \left(\frac{\rho}{S} \right)^i \\ &= \left(\sum_{k=0}^{S-1} \frac{\rho^k}{k!} + \frac{S\rho^S}{S!(S-\rho)} \right)^{-1} \frac{S\rho^S}{S!(S-\rho)} \end{aligned}$$

Cette dernière formule est appelée Formule d'Erlang-C. On note :

$$C(S, \rho) = \left(\sum_{k=0}^{S-1} \frac{\rho^k}{k!} + \frac{S\rho^S}{S!(S-\rho)} \right)^{-1} \frac{S\rho^S}{S!(S-\rho)}.$$

On peut donc implémenter un algorithme simple de dimensionnement d'une file à plusieurs serveurs étant fixée une contrainte de qualité de service en termes de proportion de clients mis en attente.

Algorithme 8.1 . Dimensionnement d'un système multiserveur avec taux de mise en attente garanti

Données : ρ, p

Résultat : le nombre de serveurs optimal S , étant donné une charge ρ et la garantie d'une proportion p d'utilisateurs mis en attente.

$S \leftarrow 1$;

tant que $p < C(S, \rho)$ **faire**

 | $S \leftarrow S + 1$;

fin

retourner S

Le nombre moyen de clients dans le système à l'équilibre est donné par :

$$\begin{aligned}
 \mathbf{E} \left[X_{\infty}^{(S)} \right] &= \sum_{i=0}^{\infty} i \pi^{(S)}(i) \\
 &= \pi^{(S)}(0) \left(\sum_{i=0}^S \frac{i \rho^i}{i!} + \frac{S^{S-1} \rho}{S!} \sum_{i=S+1}^{\infty} i \left(\frac{\rho}{S} \right)^{i-1} \right) \\
 &= \pi^{(S)}(0) \rho \left(\sum_{i=0}^{S-1} \frac{\rho^i}{i!} + \frac{S^{S-1}}{S!} \frac{d}{dx} \left(\sum_{i=S+1}^{\infty} x^i \right) \left(\frac{\rho}{S} \right) \right) \\
 &= \pi^{(S)}(0) \rho \left(\sum_{i=0}^{S-1} \frac{\rho^i}{i!} + \frac{\rho^S S^2 - S\rho + S}{S! (S - \rho)^2} \right).
 \end{aligned}$$

D'après la formule de Little, à l'état stationnaire le temps de séjour $T_s^{(S)}$ et le temps d'attente $T_A^{(S)}$ ont donc pour moyennes respectives :

$$\mathbf{E} \left[T_s^{(S)} \right] = \frac{\pi^{(S)}(0)}{\mu} \left(\sum_{i=0}^{S-1} \frac{\rho^i}{i!} + \frac{\rho^S S^2 - S\rho + S}{S! (S - \rho)^2} \right), \quad (8.10)$$

$$\mathbf{E} \left[T_A^{(S)} \right] = \frac{1}{\mu} \left[\pi^{(S)}(0) \left(\sum_{i=0}^{S-1} \frac{\rho^i}{i!} + \frac{\rho^S S^2 - S\rho + S}{S! (S - \rho)^2} \right) - 1 \right]. \quad (8.11)$$

La formule (8.11) permet de dimensionner un système, étant garanti un temps d'attente moyen.

Algorithme 8.2 . Dimensionnement d'un système multiserveur avec temps d'attente garanti

Données : ρ, μ, T_s

Résultat : le nombre de serveurs optimal S , étant donné une charge ρ et la garantie d'un temps d'attente moyen T_s .

$S \leftarrow 1$;

tant que

$$T_s < \frac{1}{\mu} \left[\left(\sum_{i=0}^{S-1} \frac{\rho^i}{i!} + \frac{S\rho^S}{S!(S - \rho)} \right)^{-1} \left(\sum_{i=0}^{S-1} \frac{\rho^i}{i!} + \frac{\rho^S S^2 - S\rho + S}{S! (S - \rho)^2} \right) - 1 \right]$$

faire

| $S \leftarrow S + 1$;

fin

retourner S

Par ailleurs, comme dans le cas de la file à un seul serveur, on peut calculer la loi du temps d'attente à travers sa transformée de Laplace.

THÉORÈME 8.6.– Si $\rho < S$, le temps d'attente W_n converge en loi vers $TA^{(S)}$ dont la transformée de Laplace est donnée par :

$$\mathbf{E} \left[e^{-sTA^{(S)}} \right] = 1 - C(S, \rho) + \pi^{(S)}(0) \frac{\rho^S}{S!} \frac{S\mu}{s + (S - \rho)\mu}. \quad (8.12)$$

Ce qui équivaut à dire que :

$$d\mathbf{P}_{TA^{(S)}}(x) = (1 - C(S, \rho))\delta_0(x) + \pi_0 \frac{\rho^S}{S!} S\mu \exp(-(S - \rho)\mu x) \mathbf{1}_{\mathbf{R}^+}(x) dx.$$

↳ Cette dernière équation signifie que le temps d'attente est nul avec probabilité $1 - C(S, \rho)$ (ce qui est évident) et que conditionnellement à ce qu'il soit positif, $TA^{(S)}$ est exponentiellement distribué de paramètre $(S - \rho)\mu$.

Démonstration. Comme les durées de service suivent des lois exponentielles, conditionnellement à ce que tous les serveurs soient pris, les interdéparts sont non seulement tous exponentiellement distribués de paramètres $S\mu$ mais de plus les interdéparts sont indépendants les uns des autres. On a donc :

$$W_n \stackrel{\text{loi}}{=} \sum_{j=1}^{\hat{X}_n - S + 1} \eta_j \mathbf{1}_{\{\hat{X}_n \geq S\}},$$

où $(\eta_k, k \geq 0)$ est une suite de variables aléatoires indépendantes et identiquement distribuées de loi exponentielle de paramètre $S\mu$. Par conséquent :

$$\mathbf{E} \left[e^{-sW_n} \mid \hat{X}_n \right] = \left(\frac{S\mu}{s + S\mu} \right)^{\hat{X}_n - S + 1} \mathbf{1}_{\{\hat{X}_n \geq S\}} + \mathbf{1}_{\{\hat{X}_n < S\}}.$$

D'après le théorème 8.4, \hat{X}_n converge en loi vers X . Comme cette espérance conditionnelle est une fonction bornée de \hat{X}_n , si $\rho < S$, $\mathbf{E} \left[e^{-sW_n} \right]$ converge simplement pour tout s vers :

$$\mathbf{E} \left[\left(\frac{S\mu}{s + S\mu} \right)^{\hat{X}_n - S + 1} \mathbf{1}_{\{\hat{X}_n \geq S\}} + \mathbf{1}_{\{\hat{X}_n < S\}} \right] = \mathbf{E} \left[e^{-sTA^{(S)}} \right].$$

On obtient donc :

$$\begin{aligned}
 \mathbf{E} \left[e^{-sT_A^{(S)}} \right] &= 1 - C(S, \rho) + \sum_{j=S}^{\infty} \left(\frac{\rho}{S} \frac{S\mu}{s + S\mu} \right)^{j-S+1} \\
 &= 1 - C(S, \rho) + \pi^{(S)}(0) \frac{\rho^S}{S!} \frac{S\mu}{s + S\mu} \sum_{j=0}^{\infty} \left(\frac{\rho\mu}{s + S\mu} \right)^j \\
 &= 1 - C(S, \rho) + \pi_0 \frac{\rho^S}{S!} \frac{S\mu}{s + (S - \rho)\mu}.
 \end{aligned}$$

□

8.3.1. Comparaison de systèmes

A partir d'une file d'attente simple, nous comparons ici qualitativement trois types d'opérations visant à en améliorer les performances : le multiplexage des ressources, le parallélisme et le débit d'exécution. On considère les quatre systèmes suivants, tous soumis à des arrivées poissonniennes d'intensité λ , de clients demandant des temps de service de loi $\varepsilon(\mu)$. La charge égale donc toujours $\rho = \frac{\lambda}{\mu}$ Erlang et l'on supposera que $\rho < 2$. On suppose de plus que tous les serveurs considérés travaillent en FCFS et que les salles d'attente sont toujours de capacité illimitée :

- le système A a un serveur travaillant à vitesse unité et correspond donc à une file d'attente simple $M_\lambda/M_\mu/1$;
- dans le système B, on dirige les clients avec probabilité 1/2 (de manière indépendantes entre les clients et indépendantes de toutes les autres variables aléatoires) vers l'un ou l'autre de deux systèmes indépendants fonctionnant en parallèle, chacun à un serveur travaillant à vitesse unité. D'après le théorème 6.6, chacune des deux files en parallèle est donc une file $M_{\frac{\lambda}{2}}/M_\mu/1$;
- le système C est une file d'attente unique à deux serveurs, chacun travaillant à vitesse unité. C est donc exactement une file $M_\lambda/M_\mu/2$;
- le système D a un serveur travaillant à vitesse double : le client C_n qui demande un temps de service de durée σ_n met effectivement un temps $\frac{\sigma_n}{2}$ pour être servi. Il est alors facile de voir que les temps de service des clients sont indépendantes et identiquement distribués de loi $\varepsilon(2\mu)$ et D correspond donc à une file d'attente $M_\lambda/M_{2\mu}/1$.

Dans la suite, nous ajoutons respectivement les exposants A, B, C et D aux caractéristiques des différents systèmes.

D'après (8.5) le temps de séjour moyen d'un client à l'équilibre dans le système D est donné par :

$$\mathbf{E} [Ts^D] = \frac{1}{2\mu - \lambda}, \quad (8.13)$$

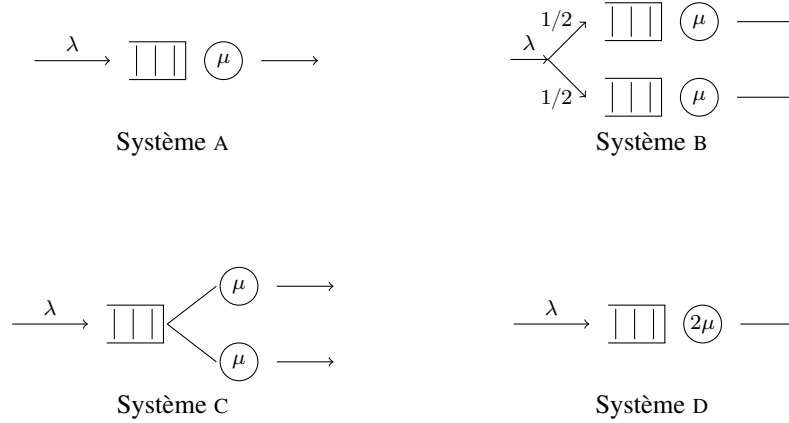


Figure 8.1 – Les quatre systèmes comparés ; la charge est ρ dans A et $\rho/2$ dans les trois autres

et dans le système A, à condition que $\rho < 1$, on a :

$$\mathbf{E}[\text{Ts}^A] = \frac{1}{\mu - \lambda}. \quad (8.14)$$

Considérons maintenant le système B. Soit X^1 et X^2 , les processus comptant respectivement le nombre de clients dans les deux files en parallèle et pour tout t , $X_t^B = X_t^1 + X_t^2$ le nombre de clients dans l'ensemble du système à t . Comme $\rho < 2$, X^1 et X^2 sont ergodiques et en notant X_∞^1 et X_∞^2 leurs limites en loi, d'après (8.3), on a :

$$\begin{aligned} \lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t X_s^B ds &= \lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t X_s^1 ds + \lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t X_s^2 ds \\ &= \mathbf{E}[X_\infty^1] + \mathbf{E}[X_\infty^2] \\ &= \frac{\lambda/2}{\mu - \lambda/2} + \frac{\lambda/2}{\mu - \lambda/2} = \frac{2\lambda}{2\mu - \lambda}. \end{aligned} \quad (8.15)$$

D'après le théorème 4.20, il existe à la condition $\rho < 2$ un temps d'attente et donc un temps de séjour stationnaire Ts^B pour ce système, vérifiant la convergence presque sûre :

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \text{Ts}_n^B = \mathbf{E}[\text{Ts}^B],$$

où pour tout n , Ts_n^B désigne le temps de séjour proposé au n^{e} client. La formule de Little s'applique donc pour ce système. Avec (8.15), il vient :

$$\mathbf{E}[\text{Ts}^B] = \frac{2}{2\mu - \lambda}. \quad (8.16)$$

Finalement, d'après (8.10) et (8.9), on a :

$$\begin{aligned}
 \mathbf{E}[W^c] &= \frac{1}{\mu} \pi(0) \left(1 + \rho + \frac{\rho^2}{2} \frac{4 - 2\rho + 2}{(2 - \rho)^2} \right) \\
 &= \frac{1}{\mu} \frac{2 - \rho}{2 + \rho} \left(\frac{4}{(2 + \rho)(2 - \rho)} \right) \\
 &= \frac{4\mu}{(2\mu - \lambda)(2\mu + \lambda)}. \tag{8.17}
 \end{aligned}$$

Regroupons (8.13), (8.14), (8.16) et (8.17). On retrouve que, dès que $\rho < 2$:

$$\begin{aligned}
 \mathbf{E}[Ts^B] - \mathbf{E}[Ts^c] &= \frac{2\lambda}{(2\mu - \lambda)(2\mu + \lambda)}; \\
 \mathbf{E}[Ts^c] - \mathbf{E}[Ts^D] &= \frac{1}{2\mu + \lambda}.
 \end{aligned}$$

Ainsi :

$$\mathbf{E}[Ts^D] < \mathbf{E}[Ts^c] < \mathbf{E}[Ts^B] < \mathbf{E}[Ts^A]; \tag{8.18}$$

$$\mathbf{E}[Ts^D] < \frac{1}{2} \mathbf{E}[Ts^A], \tag{8.19}$$

où (8.19) et la dernière inégalité de (8.18) n'ont de sens que si de plus $\rho < 1$. Du point de vue du temps passé dans le système, il est donc préférable de doubler la vitesse d'exécution que de doubler le nombre de ressources, ce qui est à son tour plus performant qu'une mise en parallèle de deux systèmes. En outre, d'après (8.19) on fait mieux que doubler les performances d'un système à un serveur en doublant sa vitesse d'exécution.

8.4. La file *processor sharing*

Nous abordons succinctement le cas de la file markovienne $M_\lambda/M_\mu/PS$, étudiée dans le cas général en section 4.2. Les clients entrent suivant un processus de Poisson d'intensité λ , demandent des temps de service de loi $\varepsilon(\mu)$ et sont immédiatement servis par un serveur selon la discipline *processor sharing* : tous sont servis simultanément, à une vitesse inversement proportionnelle au nombre de clients présents.

Notons X^{PS} , le processus comptant le nombre de clients dans le système. Il est facile de se convaincre que ce processus est Markovien et de calculer son générateur. Pour tout $t \geq 0$, notons comme précédemment :

- $W(t)$ le temps résiduel avant la prochaine arrivée après t ;
- $R(t)$, le temps résiduel avant le prochain départ après t .

Soit $i \geq 1$. Sur $\{X^{\text{PS}}(t) = i\}$, notons $R_1(t), R_2(t), \dots, R_i(t)$, les temps de service résiduels des i clients à t , comptés en unités de temps. D'après le paradoxe de l'autobus, ces temps de services sont indépendants et de loi $\varepsilon(\mu)$. Tant qu'il n'y a pas de nouvelle arrivée après t (c'est-à-dire jusqu'à $t + W(t)$), le serveur travaille à vitesse $1/i$ (ce qui multiplie l'échelle de temps par ce facteur) et donc du point de vue du serveur les temps de service résiduels des clients suivent la loi $\varepsilon(\mu/i)$ et alors :

– si $\min_{j=1, \dots, i} R_j(t) \leq W(t)$, le serveur travaille à la même vitesse jusqu'au prochain départ, qui a lieu avant la prochaine arrivée et donc $R(t)$ est le minimum de i variables aléatoires indépendantes de loi $\varepsilon(\mu/i)$;

– si $W(t) < \min_{j=1, \dots, i} R_j(t)$, à l'instant $t + W(t)$ il y a une nouvelle arrivée et donc $i + 1$ clients dans le système. Toujours d'après le paradoxe de l'autobus, la loi des caractéristiques du système ne change pas si l'on tire au sort de nouveau les temps de services des $i + 1$ clients, suivant la loi $\varepsilon(\mu)$. Alors, $R(t)$ a la même loi que le minimum de $i + 1$ variables aléatoires indépendantes de loi $\varepsilon(\mu/(i + 1))$, à moins qu'il n'y ait une autre arrivée avant le premier départ. Le cas échéant, on tire de nouveau au sort les $i + 2$ temps de service suivant la même loi $\varepsilon(\mu)$, etc.

En conclusion, dans tous les cas la loi de $R(t)$ suit conditionnellement à $\{X^{\text{PS}}(t) = i\}$, la loi $\varepsilon(\mu)$. Comme $R(t)$ suit pour tout t la loi $\varepsilon(\lambda)$, comme pour les systèmes précédent, on obtient le théorème suivant.

THÉORÈME 8.7.– Le processus X^{PS} a la même loi que le processus de congestion X de la file $M_\lambda/M_\mu/1$. Il admet donc la même loi stationnaire, à condition que $\rho < 1$. En particulier, d'après la formule de Little le temps de séjour moyen à l'équilibre est le même que celui de la file $M/M/1$.

⚠ Attention, cela ne signifie pas pour autant que les processus X et X^{PS} ont presque sûrement la même trajectoire ! Cela n'est pas le cas, l'identité précédente n'est vraie qu'en loi.

8.5. La file $M/M/\infty$

La file $M/M/\infty$ est évidemment une chimère théorique puisqu'aucun système n'aura un nombre infini de serveurs. En revanche, il existe de nombreuses situations où cet objet apparaît, par exemple dans le théorème 10.14.

Soit $(X^\infty(t), t \geq 0)$ le processus comptant le nombre de clients dans le système (et donc en service). Comme précédemment, on vérifie aisément pour tout $i \in \mathbb{N}$ que :

- le processus $(X^\infty(t), t \geq 0)$ séjourne en i durant un temps de loi $\varepsilon(i\mu + \lambda)$;
- il quitte l'état i pour l'état $i + 1$ avec probabilité $\lambda/(\lambda + i\mu)$ et si $i \geq 1$, pour l'état $i - 1$ avec probabilité $\lambda/(\lambda + i\mu)$.

Donc, si l'on admet que $T'_{n+1} - T'_n$ est indépendant du passé de T'_n (une propriété qui sera démontrée par la suite), on a :

$$\begin{aligned} \mathbf{E} [T'_{n+1} - T'_n] &= \mathbf{E} [\sigma_{n+1}] + \mathbf{E} [(T_{n+1} - T'_n) \mathbf{1}_{X(T'_n)=0}] \\ &= \frac{1}{\mu} + \frac{1}{\lambda} \mathbf{P} (X(T'_n) = 0), \end{aligned}$$

d'après le paradoxe de l'autobus. Ainsi, à l'équilibre, le temps moyen entre deux départs vaut :

$$\frac{1}{\mu} + \frac{1}{\lambda} \pi(0) = \frac{1}{\lambda},$$

autant que le temps moyen entre deux arrivées !

Le résultat suivant précise ce résultat et montre qu'à l'équilibre le processus des départs est en fait, comme celui des arrivées, un processus de Poisson d'intensité λ .

THÉORÈME 8.8 (Théorème de Burke). – A l'état stationnaire, le processus des départs est un processus de Poisson d'intensité λ . De plus, pour tout $s \geq 0$, les départs après s sont indépendants des arrivées avant s , autrement dit, $\sigma(D_u; u \geq s)$ est indépendant de $\sigma(A(u); u \leq s)$.

Démonstration. Dans tous les cas, le processus de congestion $(X(t), t \geq 0)$ est de naissance et de mort. Il est par conséquent réversible, de probabilité stationnaire π . En particulier, en supposant que X est issu de la loi π , il a pour tout $T \geq 0$ la même loi que son processus retourné $(\bar{X}^T(t), t \geq 0)$.

Il est clair que pour tout $n \in \mathbf{N}$, T'_n (respectivement T_n) est le n^{e} instant de saut vers le bas (respectivement vers le haut) du processus X . Par ailleurs, pour tout $T \geq 0$ les instants de sauts vers le bas de X jusqu'à T correspondent aux instants de sauts vers le haut du processus retourné : le processus original descend de 1 quand le processus retourné remonte d'autant. Plus précisément, si l'on note $\tilde{T}'_n = T'_n \wedge T$, $n \in \mathbf{N}$ les points du processus $(A(t), t \geq 0)$ restreint à $[0, T]$, on a pour commencer :

$$\begin{aligned} \tilde{T}'_1 &= \inf \{t \leq T; X(t) = X(t^-) - 1\} \\ &= T - \sup \{s \leq T; X(T-s) = X((T-s)^-) - 1\} \\ &= T - \sup \{s \leq T; \bar{X}^T(s) = \bar{X}^T(s^-) + 1\}. \end{aligned}$$

Le dernier *supremum* est le dernier instant de saut vers le haut du processus retourné avant $T \geq 0$. Comme X est réversible, ce *supremum* égale en loi le temps écoulé depuis le dernier instant de saut vers le haut de X (c'est-à-dire la dernière arrivée) avant $T \geq 0$. D'après le théorème 7.3, on a donc l'égalité en loi suivante :

$$\tilde{T}'_1 \stackrel{\mathcal{L}}{=} T - T_{A(T)} \stackrel{\mathcal{L}}{=} T \wedge Y_1,$$

où Y_1 est une variable aléatoire de loi $\varepsilon(\lambda)$. De la même façon, on montre que :

$$\tilde{T}'_2 - \tilde{T}'_1 \stackrel{\mathcal{L}}{=} (T - \tilde{T}'_1) \wedge Y_2,$$

où Y_2 est indépendante de Y_1 et de loi $\varepsilon(\lambda)$ et ainsi de suite.

Ceci montre que $(A(t), t \geq 0)$ égale en loi un processus de Poisson sur $[0, T]$. Comme ceci est vrai pour tout $T \geq 0$, $(A(t), t \geq 0)$ est bien un processus de Poisson.

Finalement, la propriété d'indépendance découle, naturellement, de celle du processus des arrivées. \square

8.7. Réseaux de files d'attente

Dans cette section, nous présentons succinctement les principaux résultats sur le modèle le plus simple de réseau de files d'attente : un ensemble de files d'attente dans lequel les clients qui sortent d'une file sont susceptibles d'en rejoindre une autre pour recevoir un nouveau service.

8.7.1. Réseaux de Jackson ouverts

On considère le système suivant :

- un processus de Poisson de paramètre $\lambda > 0$, qui est interprété comme le processus des arrivées de l'extérieur dans le système ;

- un ensemble de N files d'attente, où la i^e file est de type $M_{\mu_i}/S_i/\infty$ -FIFO : les clients qui entrent dans la i^e file demandent des temps de service indépendants et identiquement distribués de loi $\varepsilon(\mu_i)$ à un groupe de S_i serveurs. La loi des interarrivées dans la file n'est pas connue, car elle dépend *a priori* des autres files, comme on va le voir ;

- une matrice markovienne P de taille $N + 1$: pour tout $i \in \llbracket 0, N \rrbracket$, $P_{ij} \in [0, 1]$ et $\sum_{j=0}^N P_{ij} = 1$. La matrice P est appelée matrice de routage du système : dès qu'un client a fini son service dans la file i , il fait un tirage au sort indépendant de tous les autres paramètres pour décider de la prochaine file dans laquelle il va demander un service. Pour tout $j \in \llbracket 0, N \rrbracket$, il rejoint alors la file j avec probabilité P_{ij} . La « file 0 » représente ici l'extérieur du système : les clients faisant le chemin de 0 vers j arrivent directement de l'extérieur dans la file j et ceux qui vont de i à 0 sortent du système après avoir visité la file i . Notons que si $P_{ii} > 0$ pour un certain i , un client est susceptible de se remettre en queue dans la file i juste après avoir reçu un service dans cette même file. On suppose que P vérifie les deux conditions :

$$P_{00} = 0, \tag{8.21}$$

et pour tout $i \in \llbracket 1, N \rrbracket$, il existe $n \in \mathbb{N}$ et un n -uplet $\{i_1, i_2, \dots, i_n\}$ d'éléments de $\llbracket 1, N \rrbracket$ contenant i et tel que :

$$P_{0i_1} P_{i_1 i_2} \dots P_{i_{n-1} i_n} P_{i_n 0} > 0. \quad (8.22)$$

La condition (8.22) garantit donc que toute file i fait partie d'un chemin possible, allant de l'extérieur à l'extérieur.

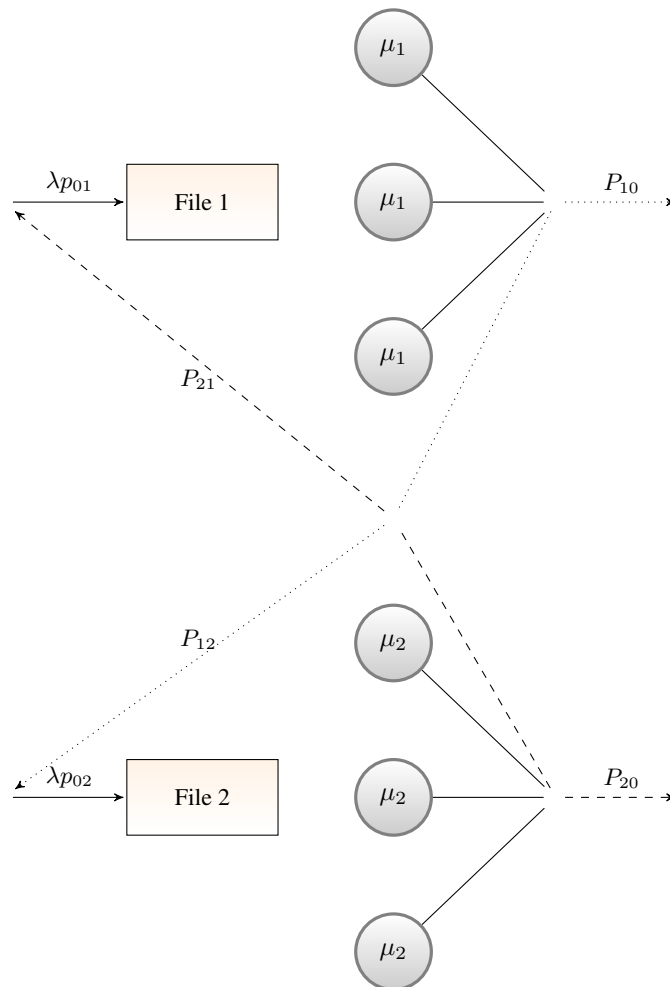


Figure 8.2 – Un réseau de Jackson à deux files

Le système est décrit par le processus :

$$(X(t), t \geq 0) = ((X_1(t), X_2(t), \dots, X_N(t)), t \geq 0),$$

où pour tout i , $X_i(t)$ compte le nombre de clients dans la i^{e} file à l'instant t . Comme le trafic dans chaque file dépend de l'état des autres files, il est facile de voir que chaque processus $(X_i(t), t \geq 0)$ n'est pas markovien. C'est le cas en revanche du processus $(X(t), t \geq 0)$, comme nous le montrons dans le lemme suivant. On appelle (voir les notations dans l'annexe A) pour tout $i \in \mathbf{N}$, $\mathbf{e}_i = (0, \dots, 0, 1, 0, \dots, 0)$ et pour tout $k \in \mathbf{N}$, $\mu_i(k) = \mu_i(k \wedge S_i)$.

LEMME 8.9.— Le processus $(X(t), t \geq 0)$ décrivant le réseau de Jackson ouvert est markovien de générateur infinitésimal A^{10} défini pour tout $x = (x(1), \dots, x(N))$ par :

$$\begin{cases} A^{10}(x, x + \mathbf{e}_j - \mathbf{e}_i) & = \mu_i(x(i))P_{ij}; \\ A^{10}(x, x + \mathbf{e}_j) & = \lambda P_{0j}; \\ A^{10}(x, x - \mathbf{e}_i) & = \mu_i(x(i))P_{i0}, \end{cases},$$

tous les autres coefficients $A^{10}(x, y)$ sont nuls et les coefficients diagonaux $A^{10}(x, x)$ égalent l'opposée de la somme des $A^{10}(x, y)$ pour $y \neq x$.

Démonstration. Supposons que le processus $(X(t), t \geq 0)$ se trouve dans l'état x à t . Alors, le processus peut quitter directement x seulement pour les états suivants :

- $x + \mathbf{e}_j - \mathbf{e}_i$ si $x(i) > 0$ et si l'un des clients en service dans la file i termine son service, puis se rend dans la file j ;
- $x - \mathbf{e}_i$ si $x(i) > 0$ et si l'un des clients en service dans la file i termine son service, puis sort du système ;
- $x + \mathbf{e}_j$ si un client entre de l'extérieur vers la file j .

D'après le théorème 6.6, le processus d'entrée de l'extérieur vers la file j est Poissonien d'intensité λP_{0j} . Le temps résiduel avant le prochain point de ce processus suit donc la loi $\varepsilon(\lambda P_{0j})$. Ensuite, le temps résiduel avant la prochaine fin de service de chacun des $x(i) \wedge S_i$ services en cours dans la file i suit, comme vu précédemment, la loi $\varepsilon((x(i) \wedge S_i)\mu_i) = \varepsilon(\mu_i(x(i)))$. En notant pour tout i l'événement :

$$B_i = \{\text{Le premier service qui se termine est un service de la file } i\},$$

la probabilité que l'on quitte effectivement x pour $x - \mathbf{e}_i + \mathbf{e}_j$ (respectivement $x - \mathbf{e}_i$) est donc donnée par $\mathbf{P}(B_i \cap \{\text{le client quitte } i \text{ pour } j\}) = \mathbf{P}(B_i) P_{ij}$ (respectivement $\mathbf{P}(B_i) P_{i0}$). Ceci conclut la preuve d'après les propriétés des lois exponentielles vues précédemment. \square

Nous aurons besoin du résultat technique suivant pour caractériser l'état stationnaire de $(X(t), t \geq 0)$.

LEMME 8.10.– Le système suivant :

$$\lambda_j = \lambda P_{0j} + \sum_{i=1}^N \lambda_i P_{ij}, \quad (8.23)$$

d'inconnues $(\lambda_1, \lambda_2, \dots, \lambda_N)$ et appelé équation du trafic, admet une unique solution dans $(\mathbf{R}^+)^N$.

Démonstration. Comme la matrice P est markovienne, il existe naturellement une unique chaîne de Markov $(M_n, n \in \mathbf{N})$ à valeurs dans $\llbracket 0, N \rrbracket$ et de matrice de transition P . Pour tout couple (i, j) d'éléments de $\llbracket 1, N \rrbracket$, il existe d'après (8.22) deux familles finies $\{i_1, \dots, i_n\}$ et $\{j_1, \dots, j_p\}$ d'éléments de $\llbracket 1, N \rrbracket$, comprenant respectivement i et j , telles que :

$$P_{0i_1} \dots P_{i_{n-1}i_n} P_{i_n 0} > 0 \text{ et } P_{0j_1} \dots P_{j_{p-1}j_p} P_{j_p 0} > 0.$$

Donc, avec les notations du chapitre 3, il existe un entier $q < n + p$ tel que la probabilité que $(M_n, n \in \mathbf{N})$ passe de i à j en q pas vérifie :

$$p^{(q)}(i, j) \geq P_{0i_1} \dots P_{i_{n-1}i_n} P_{i_n 0} P_{0j_1} \dots P_{j_{p-1}j_p} P_{j_p 0} > 0.$$

La chaîne $(M_n, n \in \mathbf{N})$ est donc irréductible sur l'espace d'état fini $\llbracket 0, N \rrbracket$. Elle est donc récurrente positive et d'après le théorème 3.15 il existe à un coefficient multiplicatif près une unique mesure stationnaire ν sur $\llbracket 0, N \rrbracket$ qui représentée en ligne, vérifie l'équation matricielle :

$$\nu = \nu P. \quad (8.24)$$

Par ailleurs, si $\lambda \{\lambda_1, \dots, \lambda_N\}$ est solution de (8.23), comme P est markovienne, d'après (8.21), on a :

$$\begin{aligned} \sum_{i=1}^N \lambda_i P_{i0} &= \sum_{i=1}^N \lambda_i \left(1 - \sum_{j=1}^N P_{ij} \right) \\ &= \sum_{i=1}^N \lambda_i - \sum_{j=1}^N \sum_{i=1}^N \lambda_i P_{ij} \\ &= \sum_{i=1}^N \lambda_i - \sum_{j=1}^N (\lambda_j - \lambda P_{0j}) \\ &= \lambda \sum_{i=1}^N P_{0i} \\ &= \lambda. \end{aligned} \quad (8.25)$$

Ceci montre bien que $\{\lambda_1, \dots, \lambda_N\}$ est solution de (8.23) si et seulement si $\Lambda = (\lambda, \lambda_1, \dots, \lambda_N)$ est solution de (8.24). L'unique solution de (8.23) est donc l'unique mesure invariante de première composante λ . \square

Pour tout $i \in \llbracket 1, N \rrbracket$, à la condition que $\lambda_i < \mu_i S_i$ le processus de congestion de la file $M_{\lambda_i}/M_{\mu_i}/S_i$ admet la probabilité invariante π^i suivante :

$$\pi^i(0) = \left(\sum_{k=0}^{\infty} \frac{(\lambda_i)^k}{\prod_{j=1}^k \mu_i(j)} \right)^{-1}; \quad (8.26)$$

$$\pi^i(k) = \frac{(\lambda_i)^k}{\prod_{j=1}^k \mu_i(j)} \pi^i(0), \quad i \geq 1. \quad (8.27)$$

Nous pouvons donc énoncer le résultat suivant.

THÉORÈME 8.11.— Soit $\{\lambda_1, \dots, \lambda_N\}$ l'unique solution de l'équation du trafic (8.23). On suppose que la condition de stabilité :

$$\lambda_i < \mu_i C_i, \quad \text{pour tout } i = 1, \dots, N, \quad (8.28)$$

est satisfaite. Soit π^i , la mesure de probabilité sur \mathbf{N} définie par (8.26) et (8.27). Alors, le processus $(X(t), t \geq 0)$ décrivant le réseau de Jackson ouvert admet sur \mathbf{N}^N l'unique probabilité stationnaire π^{JO} définie pour tout $x = (x(1), \dots, x(N)) \in \mathbf{N}^N$ par :

$$\pi^{JO}(x(1), \dots, x(N)) = \prod_{i=1}^N \pi^i(x(i)). \quad (8.29)$$

Démonstration. Dans ce qui suit, on fait la convention que $\bar{A}_{xy} = A_{xy}^{JO} = 0$ si l'une des composantes de y est négative. Pour tout x et $y \in \mathbf{N}^N$, définissons :

$$\bar{A}(x, x) = A^{JO}(x, x) \text{ et } \bar{A}(x, y) = \frac{\pi^{JO}(x) A^{JO}(x, y)}{\pi^{JO}(y)} \text{ pour } x \neq y,$$

où π^{JO} est la probabilité définie par (8.29) et A^{JO} est le générateur infinitésimal défini dans le lemme 8.9. Pour tout i et j , on a :

$$\bar{A}(x, x + \mathbf{e}_i) = \lambda_i P_{i0}$$

$$\bar{A}(x, x - \mathbf{e}_j) = \frac{\mu_j(x(j))}{\lambda_j} \lambda P_{0j} \text{ pour tout } x \text{ tel que } x(j) \geq 1,$$

ainsi que :

$$\begin{aligned}
\bar{A}(x, x - \mathbf{e}_j + \mathbf{e}_i) &= \frac{\pi^{\text{IO}}(x - \mathbf{e}_j + \mathbf{e}_i) A^{\text{IO}}(x - \mathbf{e}_j + \mathbf{e}_i, x)}{\pi^{\text{IO}}(x)} \\
&= \frac{\pi^i(x(j) - 1) \pi^i(x(i) + 1)}{\pi^j(x(j)) \pi^i(x(i))} \mu_i(x(i) + 1) P_{ij} \\
&= \frac{\mu_j(x(j))}{\lambda_j} \frac{\lambda_i}{\mu_i(x(i) + 1)} \mu_i(x(i) + 1) P_{ij} \\
&= \frac{\lambda_i}{\lambda_j} \mu_j(x(j)) P_{ij},
\end{aligned}$$

pour tout i et j tels que $i \neq j$ et tout x tel que $x(j) \geq 1$. Pour tout $x \in \mathbf{N}^N$, formons les sommes suivantes :

$$\begin{aligned}
&\sum_{y \neq x} \bar{A}(x, y) \\
&= \sum_{j=1}^N \sum_{i=1; i \neq j}^N \bar{A}(x, x - \mathbf{e}_j + \mathbf{e}_i) + \sum_{i=1}^N \bar{A}(x, x + \mathbf{e}_i) + \sum_{j=1}^N \bar{A}(x, x - \mathbf{e}_j) \\
&= \sum_{j=1}^N \left(\frac{\mu_j(x(j))}{\lambda_j} \lambda P_{0j} + \lambda_i \sum_{i=1; i \neq j}^N \frac{\mu_j(x(j))}{\lambda_j} P_{ij} \right) + \sum_{i=1}^N \lambda_i P_{i0} \\
&= \sum_{j=1}^N \frac{\mu_j(x(j))}{\lambda_j} (\lambda_j - \lambda_j P_{jj}) + \sum_{i=1}^N \lambda_i P_{i0} \\
&= \sum_{j=1}^N \mu_j(x(j)) (1 - P_{jj}) + \lambda,
\end{aligned}$$

où l'avant-dernière égalité est une conséquence de l'équation du trafic et la dernière découle de (8.25). D'autre part :

$$\begin{aligned}
&\sum_{y \neq x} A^{\text{RJ}}(x, y) \\
&= \sum_{i=1}^N \sum_{j=1; j \neq i}^N A^{\text{RJ}}(x, x - \mathbf{e}_j + \mathbf{e}_i) + \sum_{i=1}^N A^{\text{RJ}}(x, x - \mathbf{e}_i) + \sum_{j=1}^N A^{\text{RJ}}(x, x + \mathbf{e}_j) \\
&= \sum_{i=1}^N \mu_i(x(i)) (1 - P_{ii}) + \sum_{j=1}^N \lambda P_{0j} = \sum_{i=1}^N \mu_i(x(i)) (1 - P_{ii}) + \lambda = \sum_{y \neq x} \bar{A}(x, y).
\end{aligned}$$

On a donc $\bar{A}(x, x) = A^{\text{RJ}}(x, x)$ et l'on conclut par réversibilité (voir théorème 7.19). \square

Le théorème précédent, grand classique de la théorie des files d'attente, est appelé théorème de la forme produit : si le réseau de Jackson ouvert est stable, il se comporte à l'équilibre comme un système de N files d'attente indépendantes, où la i^e file est une file $M_{\lambda_i}/M_{\mu_i}/S_i$. Ce résultat, assez contre-intuitif, a un intérêt évident pour la simulation, car il indique que l'étude à l'équilibre d'un réseau de Jackson se ramène à celle de files d'attente simples à plusieurs serveurs. A l'équilibre, tout se passe donc comme si les N files fonctionnaient indépendamment et comme des files markoviennes classiques (alors que, rappelons-le, chaque file considérée seule n'est pas une file M/M). Le théorème de Burke appliqué à chacune de N files $M_{\lambda_i}/M_{\mu_i}/S_i$; $i = 1, \dots, N$ donnerait que le débit de sortie de chaque file égale λ_i et donc le débit d'entrée dans la file j soit donné par la quantité :

$$\lambda P_{0j} + \sum_{i=1}^N \lambda_i,$$

soit λ_j d'après l'équation du trafic : le résultat est cohérent.

A tout instant t , conditionnellement à ce que $X(t) = (x_1, x_2, \dots, x_N)$, on peut montrer par les techniques de calcul habituelles sur la loi exponentielle, que le temps résiduel avant la première sortie de la file i vers l'extérieur après t suit la loi $\varepsilon(\mu_i(x(i))P_{i0})$. Par conséquent, le temps résiduel avant la prochaine sortie du réseau pris dans sa globalité vers l'extérieur suit la loi $\varepsilon\left(\sum_{i=1}^N \mu_i(x(i))P_{i0}\right)$. Il est donc naturel de définir le débit de sortie instantané à t par la variable aléatoire :

$$D(t) = \sum_{i=1}^N \mu_i(X_i(t)) P_{i0}.$$

Le débit moyen de sortie à l'équilibre est donc donné par :

$$\begin{aligned} \mathbf{E}[D(\infty)] &= \mathbf{E}\left[\sum_{i=1}^N \mu_i(X_i(\infty)) P_{i0}\right] \\ &= \sum_{(x(1), \dots, x(N)) \in \mathbf{N}^N} \sum_{i=1}^N \mu_i(x(i)) P_{i0} \pi^{\text{RJ}}(x(1), \dots, x(N)), \end{aligned}$$

où $X(\infty) = (X_1(\infty), \dots, X_N(\infty))$ est la variable aléatoire distribuée sur \mathbf{N}^N suivant la loi π^{RJ} . On a alors l'analogie suivant du théorème de Burke.

THÉORÈME 8.12.— Dans un réseau de Jackson ouvert à l'équilibre, le débit moyen de sortie égale l'intensité du processus d'entrée :

$$\mathbf{E}[D(\infty)] = \lambda.$$

8.7.2. Réseaux de Jackson fermés

Le réseau de Jackson fermé est l'analogie du réseau ouvert, en supposant maintenant que le réseau n'est pas « nourri » par un processus de Poisson exogène, et qu'aucune file n'est reliée à l'extérieur : K clients (où K est fixé) circulent donc indéfiniment dans un réseau de N files M_{μ_i}/S_i , suivant les mêmes modalités que dans le réseau ouvert. Le système est donc finalement décrit par :

- la taille K de la population du réseau ;
- les N files M_{μ_i}/S_i ;
- une matrice de routage P , markovienne et de taille N satisfaisant une propriété d'irréductibilité analogue à (8.22) : pour tout i et $j \in \llbracket 1, N \rrbracket$, il existe $n \in \mathbf{N}$ et un n -uplet $\{i_1, i_2, \dots, i_n\}$ d'éléments de $\llbracket 1, N \rrbracket$ contenant i et j et tel que :

$$P_{i_1} P_{i_1 i_2} \dots P_{i_{n-1} i_n} > 0. \quad (8.30)$$

Soit l'ensemble :

$$\mathcal{A} = \left\{ x = (x(1), \dots, x(N)) \in \mathbf{N}^N; \sum_{i=1}^N x(i) = K \right\}.$$

Le processus $(X(t), t \geq 0)$, défini comme dans la section précédente, est markovien sur \mathcal{A} , de générateur défini pour tout x par :

$$A^{\text{JF}}(x, x + \mathbf{e}_j - \mathbf{e}_i) = \mu_i(x(i))P_{ij}; \text{ si } x(i) \geq 1,$$

où tous les autres termes sont nuls sauf le terme diagonal, qui est l'opposé de la somme des autres termes de la même ligne.

Comme précédemment, l'introduction d'une chaîne de Markov sur $\{1, \dots, N\}$ de matrice de transition P permet de conclure qu'il existe, à un coefficient multiplicatif près, une unique solution $\lambda \in (\mathbf{R}^+)^N$ à l'équation matricielle :

$$\lambda = \lambda P. \quad (8.31)$$

On a alors le résultat suivant.

THÉORÈME 8.13.– Le processus $(X(t), t \geq 0)$ décrivant le réseau de Jackson fermé admet une unique probabilité stationnaire π^{JF} , définie pour tout $x \in \mathcal{A}$ par :

$$\pi^{\text{JF}}(x(1), \dots, x(N)) = C \prod_{i=1}^N \prod_{\ell=1}^{x(i)} \frac{\lambda_i}{\mu_i(\ell)},$$

où $\lambda = (\lambda_1, \dots, \lambda_N)$ est une solution quelconque de (8.31) et C est la constante de normalisation :

$$C = \left(\sum_{(x(1), \dots, x(N)) \in \mathcal{A}} \prod_{i=1}^N \prod_{\ell=1}^{x(i)} \frac{\lambda_i}{\mu_i(\ell)} \right)^{-1}.$$

Un réseau de Jackson fermé est donc beaucoup moins confortable à étudier que dans le cas ouvert. On peut en effet remarquer que, au contraire du cas ouvert, la forme précédente n'est pas une forme produit à cause de l'expression de la constante C . De plus, le calcul numérique de cette constante pour un grand réseau est très coûteux algorithmiquement.

8.8. Problèmes

EXERCICE 16.— On considère la file $M/M/1$ avec la discipline de service suivante : une proportion p des clients (les « polis ») se place normalement en queue, tandis qu'une proportion $q = 1 - p$ « d'impolis » double tout le monde et prend la première place de la file d'attente. La catégorie (« polis » ou « impolis ») d'un client est indépendante de son instant d'arrivée. Cette discipline est non préemptive, c'est-à-dire que l'on n'interrompt pas le service en cours. L'intensité du processus d'arrivée global est λ , le temps moyen de service est $1/\mu$.

On note $(X(t), t \geq 0)$ le nombre de clients dans le système (file + serveur) à l'instant t .

- 1) Donner le générateur infinitésimal de X . Est-il différent du générateur infinitésimal de la file $M/M/1$ avec discipline FIFO (*First In First Out*) ?
- 2) En déduire la condition de stabilité du système et la distribution à l'état stationnaire de X .
- 3) Quel est le temps moyen d'attente à l'état stationnaire ?
- 4) Quelles sont la nature et l'intensité du processus d'arrivée des clients « impolis » ?
- 5) On note W^p le temps d'attente d'un client poli donné, on note X le nombre de clients dans le système à son arrivée et $N(W^p)$ le nombre de clients « impolis » qui arrivent pendant son temps d'attente. Montrer qu'en loi, on a :

$$W^p \stackrel{\text{loi}}{=} \sum_{j=1}^X \eta_j + \sum_{l=1}^{N(W^p)} \sigma_l, \quad (8.32)$$

où $(\eta_j, j \geq 0)$ et $(\sigma_l, l \geq 0)$ sont deux suites indépendantes entre elles et indépendantes de X , de variables aléatoires indépendantes de loi exponentielle de paramètre μ .

- 6) Expliquer pourquoi X et $N(W^p)$ sont indépendantes conditionnellement à W^p .
- 7) On suppose dorénavant que X a la loi stationnaire identifiée dans la deuxième question, montrer que :

$$\mathbf{E} \left[\sum_{j=1}^X \eta_j \right] = \frac{\rho}{(1-\rho)\mu}.$$

8) Montrer que :

$$\mathbf{E} \left[e^{-s \sum_{i=1}^{N(W^p)} \sigma_i} \mid W^p = v \right] = e^{-\lambda q v \frac{s}{\mu+s}}.$$

9) En déduire que si X a la loi stationnaire identifiée dans la deuxième question, on a :

$$\mathbf{E} \left[e^{-\lambda q W^p \frac{s}{\mu+s}} \mathbf{E} \left[e^{-s \sum_{j=1}^X \eta_j} \mid W^p \right] \right].$$

(On ne demande pas de calculer explicitement $\mathbf{E} [e^{-s W^p}]$.)

10) En déduire, par dérivation, le temps moyen d'attente d'un client « poli ».

11) On note W^i le temps d'attente d'un client « impoli » à l'état stationnaire et W le temps d'attente d'un client quelconque à l'état stationnaire – on considère prouvé que le temps d'attente du n^{e} client impoli converge en loi vers W^i , de même pour W . Expliquer pourquoi l'on a :

$$\mathbf{E} [W] = p \mathbf{E} [W^p] + q \mathbf{E} [W^i].$$

12) En déduire le temps moyen d'attente d'un client « impoli ».

EXERCICE 17.– On considère que l'on a une station d'essence à deux pompes et un emplacement d'attente :

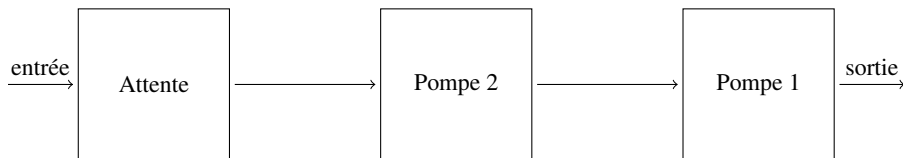


Figure 8.3 – Station d'essence à trois pompes en série

Les clients arrivent à la pompe selon un processus de Poisson d'intensité λ . Si les deux pompes sont libres, le client va en pompe 1. Si la pompe 1 est prise, le client va en pompe 2. Si les deux pompes sont occupées, le client peut se mettre dans la place d'attente. Si la place d'attente est occupée, le client passe son chemin. Un client en pompe 2 doit attendre que la pompe 1 soit libre pour sortir de la station d'essence. Dans le cas où le client à la pompe 2 finit avant celui de la pompe 1, les deux sortent en même temps de la station au moment où le client de la pompe 1 a fini. S'il y a, à ce moment-là, un client en attente, il va instantanément en pompe 1 et commence à remplir son réservoir. Dans le cas où la pompe 1 est libre mais la pompe 2 occupée, personne ne peut entrer en pompe 1. Les temps de remplissage des réservoirs suivent une loi exponentielle de paramètre μ . On modélise ce système par un processus de Markov $X = (X_1, X_2, X_3)$ où X_1 vaut 0 ou 1 et représente le nombre de clients à la

pompe 1, X_2 représente le nombre de clients à la pompe 2 et X_3 représente le nombre de clients dans la place d'attente (X_3 vaut 0 ou 1).

- 1) Quel est l'espace d'état ?
- 2) Ecrire les composantes du générateur infinitésimal ?
- 3) Existe-il une probabilité stationnaire ?
- 4) Si oui, quelles sont les équations qui la caractérise ?
- 5) Quel est le pourcentage de clients qui ne peuvent pas entrer dans la station ?
- 6) Quel est le pourcentage de clients qui ne peuvent pas entrer dans la station parce qu'elle est mal faite : le pourcentage des clients qui ne peuvent pas être servis bien que la pompe 1 soit libre.

EXERCICE 18.— Dans un magasin, les clients, en moyenne vingt par heure, arrivent à la caisse selon un processus de Poisson. Tant qu'il y a moins de deux clients en attente, il y a une seule caisse d'ouverte. Le temps de service à cette caisse est exponentiellement distribué, le temps moyen de service est de cinq minutes. A partir de trois clients en attente, une deuxième caisse ouvre. Les deux caisses partagent la même file d'attente. La deuxième caisse referme dès qu'il n'y a plus que deux clients en attente. On note X_t le nombre de clients présents à l'instant t .

- 1) Ecrire le générateur infinitésimal de X .
- 2) Trouver la probabilité stationnaire si elle existe.
- 3) Quel est le nombre moyen de clients dans le système ?
- 4) Quel est le nombre moyen de clients en attente ?
- 5) Quel est le temps moyen d'attente ?
- 6) Quel est le pourcentage du temps où la deuxième caisse est ouverte ?

EXERCICE 19.— Une agence bancaire comporte $S = 5$ employés. Le nombre d'appels est en moyenne de vingt par heure, la durée moyenne d'un appel est de six minutes. Les arrivées forment un processus de Poisson d'intensité λ et les temps de service sont indépendants et de loi exponentielle de paramètre μ . Dans les questions 1 à 5, on ne vous demande pas de démontrer les formules, juste de trouver celles qu'il faut utiliser et de faire les calculs numériques. Ecrivez bien formellement tous les calculs que vous avez à faire, il y a pas mal de résultats partiels qui servent plusieurs fois.

- 1) Dans le cas où $N = 5$ et où il n'y a pas d'attente, quel est le pourcentage de client dont l'appel échoue faute de conseiller libre ?
- 2) L'agence s'équipe d'un dispositif de mise en attente, c'est-à-dire un disque qui vous fait patienter en attendant qu'un conseiller se libère. On suppose que tous les clients attendent aussi longtemps que nécessaire. Quelle est la probabilité que la probabilité d'avoir à attendre ?
- 3) Est-elle égale à la probabilité de blocage dans le système sans attente ? Pourquoi ?

- 4) Quel est le temps moyen d'attente ?
- 5) Quelle est la probabilité que le temps d'attente soit supérieur à deux minutes ?

On supprime maintenant le système de mise en attente et on considère qu'un pourcentage $100.p$ d'appels nécessitent un reroutage vers un centre de traitement plus spécialisé. Au niveau du PABX (et des N liaisons louées), cela se traduit par une occupation de deux liaisons : une pour l'appel entrant, l'autre pour l'appel rerouté. On note X^1 le nombre d'appels directs, utilisant une seule liaison ; X^2 représente le nombre d'appels qui ont besoin de deux liaisons. N est maintenant à déterminer. X^1 est limité par S le nombre de conseillers, tandis que X^2 est seulement limité par la condition $X^1 + 2X^2 \leq N$. On considère, en effet, que le dimensionnement du centre de traitement spécialisé est tel que les appels qui lui sont destinés aboutissent et on néglige le temps nécessaire au conseiller local pour décider du reroutage. D'autre part, lorsque tous les conseillers sont occupés, les appels sont tous dirigés vers le centre de traitement spécialisé.

- 6) Ecrire le générateur infinitésimal de $X = (X^1, X^2)$.
- 7) Le processus X admet-il une probabilité stationnaire ?
- 8) Ecrire les équations permettant de la déterminer.
- 9) A N fixé, quelle est la taille de la matrice à inverser ?

EXERCICE 20.— Retrouver (8.20) avec la formule de Little.

EXERCICE 21.— Démontrer le théorème 8.13 en utilisant le théorème 7.19.

8.9. Notes et commentaires

Nous n'avons qu'effleuré l'étude des files d'attente markoviennes à buffer infini. Depuis le temps que ce sujet a intéressé nombre de chercheurs, il existe nombre d'articles dévolus à ce problème. Nous n'avons en particulier pas parlé de trafic élastique dans les files *processor sharing*. Ce cadre pourtant riche et très bien étudié dans [BON 11] permet d'obtenir une formule pour les délais, insensible à la loi du temps de service.

L'étude des réseaux à forme produit repose essentiellement sur le lemme de Kelly que l'on trouvera dans [KEL 79]. L'équation de balance locale que cela induit est souvent trop restrictive pour être satisfaite. De nombreuses versions plus faibles mais toutefois moins contraignantes que les équations globales induites par la recherche des solutions de $\pi A = 0$, existent. On en trouvera de nombreux exemples dans [CHA 99].

Résumé

– La probabilité stationnaire d'une file M/M/1 d'intensité λ en entrée et de temps de service moyen $1/\mu$ est donnée par :

$$\pi(0) = 1 - \rho, \pi(n) = \rho^n(1 - \rho) \text{ où } \rho = \lambda/\mu.$$

– Le temps d'attente à l'état stationnaire vaut 0 avec probabilité $1 - \rho$. Conditionnellement à ce qu'il soit strictement positif, le temps d'attente suit une loi exponentielle de paramètre $\mu - \lambda$. Sa valeur moyenne est $1/(\mu - \lambda)$.

– Dans la file M/M/S, la condition de stabilité est $\rho < S$. La probabilité stationnaire est donnée par (8.9).

– Les comparaisons de files d'attente ne peuvent se faire qu'à charge égale.

– Dans un réseau de Jackson ouvert, la probabilité stationnaire du système est le produit des probabilités stationnaires de chacune des files vues comme une file M/M/1 avec comme paramètres d'entrée, les composantes de la solution de l'équation d'équilibre des flux, voir lemme 8.10.

Chapitre 9

Modèles à pertes

9.1. Généralités

Un système à pertes est un système qui comporte moins de ressources que d'utilisateurs potentiels ; tout client arrivant à un moment où les ressources du système sont toutes occupées est perdu. Il s'agit donc de trouver le nombre de ressources adéquat de façon à ne pas trop pénaliser le client ou le fournisseur. Dans toute la suite, on désignera toujours par :

- $N(t)$, le nombre de clients qui se sont présentés à l'entrée du système jusqu'à l'instant t ;
- $Y(t)$, le nombre de clients qui sont effectivement rentrés dans le système entre 0 et t . Par conséquent, $X^d(t) = N(t) - Y(t)$ est le nombre de clients perdus à l'instant t ;
- $X(t)$, le nombre de clients dans le système (serveurs + salle d'attente) à l'instant t ;
- S le nombre de serveurs et K le nombre de places dans la salle d'attente.

EXEMPLE 9.1.– Le réseau téléphonique mobile de deuxième génération, plus connu sous le nom de GSM, repose techniquement sur le TDMA (*Time Division Multiple Access*).

Pour une fréquence, on divise le temps en périodes de durée égales et constantes, appelées *slots*. Les *slots* sont regroupés par paquet de huit pour former une trame. La voix d'un appel est numérisée de sorte qu'acheminer une communication revient à transporter des bits d'information. Pour acheminer une communication, les octets d'un appel sont regroupés par paquets et on attribue à un appel un *slot* fixe pendant toute la durée de la communication. Sur la figure 9.1, c'est le quatrième *slot* qui est attribué à

un appel. Un équipement émission/réception ne peut donc acheminer que huit appels simultanément. Il importe donc à l'opérateur d'en déterminer le nombre pour chaque station de base, c'est-à-dire les antennes qui décorent les toits des immeubles. En fait, pour tenir compte de la signalisation, un équipement permettra d'acheminer sept appels simultanés, deux équipements quatorze, trois équipements vingt-deux, etc.

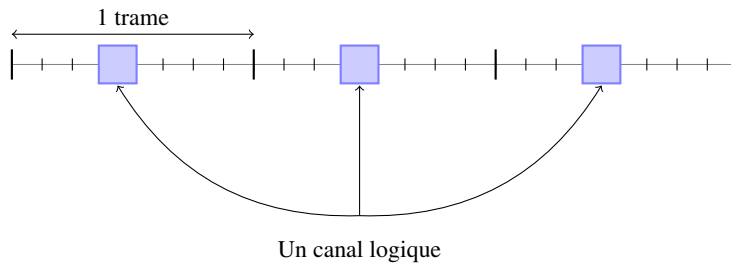


Figure 9.1 – Principe du TDMA

EXEMPLE 9.2.– Un multiplexeur est l'un des éléments-clés des réseaux de données. Il permet d'acheminer les communications qui arrivent de N ports vers N ports de sortie. Comme un même port de sortie peut être sollicité simultanément par plusieurs flux, il importe d'installer des mémoires pour stocker temporairement les données en attendant que le canal soit libre. Le problème est ici de trouver la taille des mémoires de sorte que la probabilité de perte de bits soit inférieure à un certain seuil. Ce seuil dépend de la nature du flux de données. *Grosso modo*, les flux voix et vidéo tolèrent des pertes de l'ordre de un pour mille tandis que les flux de données ne devraient tolérer aucune perte. En pratique, on considère que quelque chose qui arrive avec une probabilité inférieure à un pour un milliard ne se produira de fait pas.

Ces deux exemples sont de même nature mais le deuxième soulève deux problèmes de taille : il n'est pas évident de choisir un modèle de trafic d'entrée qui puisse refléter les différences (de débit, par exemple, pour ne citer qu'elles) qui existent entre des flux de voix, de vidéo et paroles ; par ailleurs la petitesse de la probabilité de perte interdit toute simulation et exige un modèle analytique précis. Ils nous amènent aussi à préciser et distinguer trois probabilités.

DÉFINITION 9.1.– La probabilité de congestion d'un système à $S + K$ ressources est la proportion asymptotique du temps où toutes les ressources soient occupées :

$$P_B = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \mathbf{1}_{\{X(t)=S+K\}} dt.$$

La probabilité de perte est la probabilité de blocage du point de vue des clients :

$$P_p = \lim_{t \rightarrow \infty} \frac{1}{N(t)} \sum_{j=1}^{N(t)} \mathbf{1}_{\{X(T_j^-) = S+K\}} \quad (9.1)$$

La probabilité de dépassement de seuil n'existe en fait que lorsque le nombre de ressources est infini, elle est alors définie par :

$$P_S = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \mathbf{1}_{\{X(t) > S\}} dt.$$

Ces trois probabilités sont en général distinctes. En effet, considérons une situation avec un seul serveur, des arrivées et des temps de service déterministes, de durée respectives $\rho < 1$ et 1.

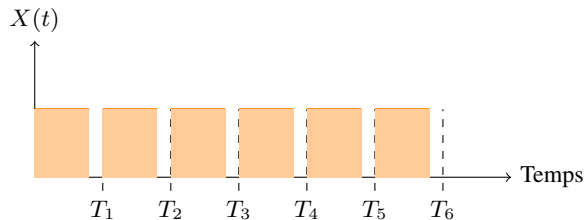


Figure 9.2 – L'évolution temporelle de la file $D/D/1/1$

D'après les définitions, la probabilité de blocage est égale au pourcentage d'aire grisée, soit ρ . La probabilité de perte est nulle puisqu'aucun client n'arrive avec le serveur occupé. En revanche, si les arrivées sont poissonniennes, les probabilités de perte et de blocage coïncident en vertu de la propriété PASTA.

THÉORÈME 9.1.– Si les arrivées suivent un processus de Poisson alors $P_B = P_p$.

Démonstration. Il suffit d'appliquer le théorème A.38 à $\psi(s) = \mathbf{1}_{\{X(s) = S+K\}}$. \square

La probabilité de perte est de loin la plus cruciale mais aussi la plus difficile à évaluer. Les probabilités de blocage et de dépassement sont beaucoup plus facilement calculables et c'est pourquoi elles sont souvent utilisées comme succédané à la probabilité de perte.

REMARQUE.— Nous pouvons retrouver cette probabilité invariante en remarquant que la file $M_\lambda/M_\mu/S/S$ est une file $M_\lambda/M_\mu/\infty$ contrainte à ne pas dépasser S clients. En d'autres termes, $(X^E(t), t \geq 0)$ est la version tronquée, contrainte à rester dans $\llbracket 0, S \rrbracket$, du processus réversible $(X^\infty(t), t \geq 0)$ représentant le nombre de clients dans la $M_\lambda/M_\mu/\infty$. Pour tout $i \in \llbracket 0, S \rrbracket$, le lemme de Kelly permet de calculer π^E :

$$\pi^E(i) = \frac{1}{\sum_{k=0}^S \pi^\infty(k)} \pi^\infty(i) = \frac{\rho^i/i!}{\sum_{k=0}^S \rho^k/k!}.$$

D'après (9.1), la probabilité de perte pour ce système vaut :

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^n \mathbf{1}_{\{S\}}(X(T_j^-)) &= \lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t \mathbf{1}_{\{S\}}(X(s^-)) ds \\ &= \pi^E(S). \end{aligned}$$

Nous avons donc établi le résultat suivant.

THÉORÈME 9.2.— La probabilité de perte de la file $M_\lambda/M_\mu/S/S$ de charge ρ est donnée par :

$$\text{Er}[\rho, S] = \frac{\rho^S/S!}{\sum_{i=0}^S \rho^i/i!}.$$

REMARQUE.— On montre que cette probabilité de perte ne dépend pas de la loi de service : cette expression est encore vraie pour la file $M/GI/S/S$.

Pour l'évaluation numérique de cette probabilité, on pourra utiliser l'approximation suivante :

$$\text{Er}[\rho, S] \approx \exp S \log \frac{\rho u}{S} + S - \rho u \sqrt{\frac{u + \rho u(1-u)^2}{S}}$$

avec u défini par :

$$u = \frac{(S + \rho + 1) - \sqrt{(S + \rho + 1)^2 - 4\rho S}}{2\rho};$$

ou alors la relation de récurrence :

$$\frac{1}{\text{Er}[\rho, S]} = 1 + \frac{S}{\rho \text{Er}[\rho, S-1]}.$$

En pratique, ayant mesuré ρ , il s'agit de trouver S tel que $\text{Er}[\rho, S]$ soit inférieur au seuil désiré pour garantir la qualité de service. On considère généralement que ce seuil

est de 0,001 pour le réseau téléphonique fixe et 0,02 pour le réseau GSM. L'algorithme est alors le suivant.

Algorithme 9.1 . Calcul du nombre de serveurs pour garantir une probabilité de perte

Données : ρ, ϵ
Résultat : S tel que $\text{Er}[\rho, S] \leq \epsilon$
 $S \leftarrow 0;$
 $x \leftarrow 1;$
tant que $x < \epsilon^{-1}$ **faire**
 $S \leftarrow S + 1;$
 $x \leftarrow 1 + \frac{S}{\rho}x;$
fin
retourner S

L'expérience montre que pour une perte de un pour mille, le nombre de serveurs nécessaire est (très) approximativement égal à $\rho + 3\sqrt{\rho}$. Si les arrivées et les temps de service étaient déterministes, ρ représenterait le nombre d'appels simultanés donc aussi le nombre de serveurs nécessaires. Le terme $3\sqrt{\rho}$ s'interprète donc comme une garantie contre les fluctuations aléatoires du trafic.

9.3. La file M/M/1/1 + K

Nous considérons maintenant une file d'attente à un serveur travaillant en FCFS, dont la salle d'attente a une capacité limitée K . Comme précédemment, le processus d'arrivées est poissonien d'intensité λ , les clients demandent des temps de service indépendants et identiquement distribués de loi $\varepsilon(\mu)$ et nous notons $\rho = \lambda/\mu$, la charge du système.

Le processus $(X^k(t), t \geq 0)$ comptant le nombre de clients dans le système est à valeurs dans $\llbracket 0, 1 + K \rrbracket$. Il satisfait la dynamique suivante :

- pour tout $i \in \llbracket 0, K \rrbracket$, le temps de séjour en i et les probabilités de saut partant de i sont clairement les mêmes que pour la file $M_\lambda/M_\mu/1/\infty$;

- une fois en $1 + K$, le système est plein et ne peut plus accepter de clients. Le processus ne peut quitter cet état que pour l'état K , après un temps de séjour égal au temps de service résiduel du client alors en service, de loi $\varepsilon(\mu)$ tant que le processus vaut $1 + K$.

Il est donc facile de voir que $(X^K(t), t \geq 0)$ est markovien et admet sur $\llbracket 0, 1 + K \rrbracket$ le générateur infinitésimal :

$$A^K = \begin{pmatrix} -\lambda & \lambda & & & \\ \mu & -(\lambda + \mu) & \lambda & (0) & \\ & & \ddots & & \\ (0) & \mu & -(\lambda + \mu) & \lambda & \\ & & \mu & -\mu & \end{pmatrix}.$$

Le processus $(X^K(t), t \geq 0)$ étant à valeurs dans un ensemble fini, il est nécessairement ergodique et admet donc une unique probabilité stationnaire π^K . Celle-ci est réversible puisque $(X^K(t), t \geq 0)$ est de naissance et de mort. On la retrouve, soit directement en résolvant le système :

$$\begin{cases} \pi^K A^K &= \mathbf{0} \\ \pi^K \mathbf{e} &= 1, \end{cases}$$

soit encore en appliquant le lemme de Kelly : $(X^K(t), t \geq 0)$ est la troncature à l'ensemble $\llbracket 0, 1 + K \rrbracket$ du processus réversible $(X(t), t \geq 0)$ comptant le nombre de clients dans la $M_\lambda/M_\mu/1/\infty$. On retrouve la probabilité donnée pour tout $i \in \llbracket 0, 1 + K \rrbracket$ par :

$$\begin{aligned} \pi^K(i) &= \frac{\pi(i)}{\sum_{k \in \llbracket 0, 1 + K \rrbracket} \pi(k)} \\ &= \frac{(1 - \rho)\rho^i}{\sum_{k \in \llbracket 0, 1 + K \rrbracket} (1 - \rho)\rho^k} \\ &= \begin{cases} \frac{\rho^i - \rho^{i+1}}{1 - \rho^{K+2}} & \text{si } \rho \neq 1, \\ \frac{1}{K + 2} & \text{si } \rho = 1, \end{cases} \end{aligned} \tag{9.3}$$

et l'on remarque au passage que l'unique probabilité invariante et la probabilité uniforme sur $\llbracket 0, 1 + K \rrbracket$ dans le cas critique $\rho = 1$.

THÉORÈME 9.3.— Dans une file $M_\lambda/M_\mu/1/1+K$, la probabilité de perte est donnée par la formule :

$$\frac{\rho^{K+1} - \rho^{K+2}}{1 - \rho^{K+2}} \text{ si } \rho \neq 1.$$

Si $\rho = 1$, la perte est donnée par $1/(K + 2)$.

Démonstration. La probabilité de perte pour ce système est donnée par la limite du taux de clients perdus, c'est-à-dire les clients trouvant un système plein à leur arrivée.

D'après la propriété PASTA et l'ergodicité de $(X^K(t), t \geq 0)$, celle-ci s'écrit :

$$\begin{aligned} \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \mathbf{1}_{\{1+K\}}(X^K(T_n-)) &= \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \mathbf{1}_{\{1+K\}}(X^K(t)) dt \\ &= \mathbf{E}[\mathbf{1}_{\{1+K\}}(X_\infty^K)] \\ &= \pi^K(1+K). \end{aligned}$$

Le résultat découle alors de (9.3). \square

Si la charge ρ est strictement inférieure à 1, la probabilité de perte peut être comparée à la probabilité de dépassement de seuil de la file $M_\lambda/M_\mu/1/\infty$ correspondante. Avec les notations de la section 8.2, la probabilité que ce dernier système ait une congestion supérieure ou égale à $1+K$ à l'équilibre s'écrit :

$$\mathbf{P}(X_\infty \geq 1+K) = \sum_{i=1+K}^{\infty} \pi(i) = (1-\rho) \sum_{i=1+K}^{\infty} \rho^i = \rho^{1+K}, \quad (9.4)$$

alors que :

$$P_{\text{PERTE}} = \pi^K(0)\rho^{1+K} < \rho^{1+K}. \quad (9.5)$$

Ainsi, il est ainsi possible d'estimer la perte d'un système à buffer fini par la probabilité de dépassement de seuil de la file à buffer infini correspondante. Ce faisant, on surestime la perte d'après (9.4) et (9.5), mais cela peut s'avérer être le seul calcul que l'on sache faire en pratique car le système à buffer infini est globalement plus simple à décrire par des lois de probabilité connues. Qui plus est, un dimensionnement pessimiste tant qu'il n'est pas exagéré, il permet de se protéger contre les fluctuations de trafic ou contre les incertitudes sur l'estimation de la charge.

Dimensionnement du buffer

Il est intéressant de se demander quelle est la taille de buffer optimale que l'on doit déployer pour garantir à l'utilisateur une qualité de service en termes de probabilité de perte de paquets. On peut déjà vérifier que la perte du système est une fonction décroissante du nombre de serveurs. En effet, la fonction définie pour tout $\rho \in \mathbf{R} + \setminus \{1\}$ par :

$$f_\rho : \begin{cases} \mathbf{R} & \longrightarrow \mathbf{R} \\ x & \longmapsto \frac{\rho^{x+1} - \rho^{x+2}}{1 - \rho^{x+2}}. \end{cases}$$

admet pour dérivée la fonction définie pour tout x par :

$$f'_\rho(x) = \frac{\rho^{x+1} \ln \rho}{(1 - \rho^{x+2})^2} (1 - \rho) > 0,$$

et les choses se passent bien sûr de la même façon pour $\rho = 1$.

Le dimensionnement optimal du buffer est donc donné par l'algorithme suivant (dans le cas $\rho \neq 1$).

Algorithme 9.2 . Calcul du nombre minimal de places dans le buffer pour garantir une probabilité de perte $\leq \varepsilon$.

Données : ρ, ε

Résultat : K tel que $P_{\text{PERTE}} \leq \varepsilon$

$K \leftarrow 1$;

tant que

$$\varepsilon < \frac{\rho^{K+1} - \rho^{K+2}}{1 - \rho^{K+2}}$$

faire

fin

retourner K

9.4. L'effet *trunk*

Appliquons nos résultats au GSM. Dans le langage du protocole, un *slot* s'appelle un TCH (*traffic channel*). On considère usuellement que la charge par mobile est de 0,025 erlang et que la probabilité de perte acceptable est 0,02. Compte tenu de ce qui précède, on obtient les résultats suivants.

Nb émetteur-récepteurs	1	2	3	4	5	6	7
Nb TCH	7	14	22	29	37	45	52
Capacité	2,9	8,2	15	21	28	35,5	42
Trafic écoulé par TCH	0,41	0,59	0,68	0,72	0,76	0,79	0,81
Nb de mobiles	116	328	596	840	1128	1424	1680

Tableau 9.1 – Mise en évidence de l'effet *trunk*

Ce tableau appelle deux commentaires. Il importe de prendre conscience de la grande variabilité des résultats. Une augmentation raisonnable du nombre de ressources induit une grande augmentation de la capacité. Par exemple, si l'on passe de un à deux émetteurs-récepteurs, la capacité est multipliée par presque trois. Pour écouler plus de trafic, on n'a pas besoin d'augmenter proportionnellement le nombre de TCH : pour écouler 42 erlangs au lieu de 21, on a besoin de 52 au lieu de $2 * 29 = 58$.

L'autre remarque est connue sous le nom d'effet *trunk* : plus on augmente le nombre de serveurs, plus la charge écoulee par chaque serveur est importante. Journalistiquement, cela s'appelle aussi « les économies d'échelle ». Par conséquent, lorsque le seuil

THÉORÈME 9.4.– Dans le modèle d’Engset, la probabilité de perte est donnée par :

$$\text{Eng}[\rho, S, M] = \frac{\rho^S C_{M-1}^S}{\sum_{j=0}^S C_{M-1}^j \rho^j}.$$

Démonstration. D’après la construction du modèle, le processus d’arrivées est un processus ponctuel de Poisson dont l’intensité varie dans le temps comme une fonction de $X(s)$: son intensité est donnée par :

$$\Lambda(t) = \int_0^t (M - X(s)) \lambda \, ds.$$

Comme les sauts de $N(t)$ sont de hauteur 1, le crochet de la martingale $N - \Lambda$ est aussi Λ . Soit le processus adapté borné, $\psi(s) = \mathbf{1}_{\{X(s)=S\}}$. D’après le théorème A.37 :

$$\begin{aligned} P_p &= \lim_{t \rightarrow \infty} \frac{1}{N(t)} \int_0^t \psi(s_-) \, dN(s) = \lim_{t \rightarrow \infty} \frac{1}{\Lambda(t)} \int_0^t \psi(s) \, d\Lambda(s) \\ &= \lim_{t \rightarrow \infty} \frac{t}{\Lambda(t)} \frac{1}{t} \int_0^t \psi(s) (M - X(s)) \lambda \, ds. \end{aligned}$$

Comme X est markovien ergodique, on a les limites presque sûres suivantes :

$$\begin{aligned} \frac{1}{t} \Lambda(t) &\xrightarrow{t \rightarrow \infty} (M - \sum_{j=0}^S j \nu(j)) \lambda \\ \text{et } \frac{1}{t} \int_0^t \psi(s) (M - X(s)) \lambda \, ds &\xrightarrow{t \rightarrow \infty} \lambda (M - S) \nu(S). \end{aligned}$$

Par conséquent :

$$\text{Eng}[\rho, S, M] = \frac{\lambda (M - S) \nu(S)}{\lambda \sum_{j=0}^S (M - j) \nu(j)} = \frac{\rho^S C_{M-1}^S}{\sum_{j=0}^S C_{M-1}^j \rho^j}.$$

En d’autres termes, la probabilité de pertes d’un système à M machines est la probabilité de blocage d’un système à $M - 1$ machines, le reste étant inchangé par ailleurs. \square

9.6. File IPP/M/S/S

Les résultats précédents peuvent laisser accroire que la perte dépend uniquement de la charge. Il n’en est rien comme l’illustre l’exemple suivant.

Un processus IPP (*Interrupted Poisson Process*) est un cas particulier de processus MMPP (voir 7.6) où le processus de phases J est à deux états : A et B . Le générateur infinitésimal de J est donc de la forme :

$$Q_J = \begin{pmatrix} -\sigma_A & \sigma_A \\ \sigma_B & -\sigma_B \end{pmatrix},$$

où $1/\sigma_A$ et $1/\sigma_B$ sont les temps moyens de séjour dans les phases A et B respectivement. Sa probabilité invariante, notée ν se calcule facilement et l'on obtient :

$$\nu(A) = \frac{\sigma_B}{\sigma_A + \sigma_B} \text{ et } \nu(B) = \frac{\sigma_A}{\sigma_A + \sigma_B}.$$

La file IPP/M/S/S est donc la file à S serveurs, sans buffer, avec des temps de service de loi exponentielle et un processus d'arrivées IPP. Le processus X comptant le nombre de serveurs occupés n'est pas markovien mais le processus (X, J) l'est. On numérote les états par ordre lexicographique et on note Λ la matrice des intensités d'arrivées :

$$\Lambda = \begin{pmatrix} \lambda_A & 0 \\ 0 & \lambda_B \end{pmatrix}.$$

L'intensité moyenne du flux d'entrée est donc :

$$\lambda = \lambda_A \frac{\sigma_B}{\sigma_A + \sigma_B} + \lambda_B \frac{\sigma_A}{\sigma_A + \sigma_B}. \quad (9.6)$$

Le générateur infinitésimal de (X, J) s'écrit donc :

$$A = \begin{pmatrix} Q_J - \Lambda & \Lambda & & & \\ \mu \text{ Id} & (Q_J - \Lambda - \mu \text{ Id}) & \Lambda & & \\ & 2\mu \text{ Id} & (Q_J - \Lambda - 2\mu \text{ Id}) & \Lambda & \\ & & \ddots & \ddots & \\ & & & S\mu \text{ Id} & (Q_J - S\mu \text{ Id}) \end{pmatrix}$$

Le processus de Markov est à espace d'états fini, évidemment irréductible, il admet une probabilité invariante π solution de l'équation $\pi A = 0$ et $\pi \cdot \epsilon = 1$. Pour simplifier les calculs, on introduit les vecteurs lignes à deux composantes $x_n = (\pi(n, A), \pi(n, B))$ pour $n = 0, \dots, S$. Les équations correspondant à $\pi A = 0$ deviennent ainsi S équations vectorielles 2 par 2 :

$$\begin{aligned} x_0(Q_J - \Lambda) + \mu x_1 &= 0 \\ x_0\Lambda + x_1(Q_J - \Lambda - \mu \text{ Id}) + 2\mu x_2 &= 0 \\ &\vdots \\ x_{S-2}\Lambda + x_{S-1}(Q_J - \lambda - (S-1)\mu \text{ Id}) + S\mu x_S &= 0 \\ x_{S-1}\Lambda + x_S(Q_J - S\mu \text{ Id}) &= 0. \end{aligned}$$

Des $(S - 1)$ premières équations, on tire :

$$\begin{aligned} x_1 &= \frac{1}{\mu} x_0 (Q_J - \Lambda) \\ x_2 &= -\frac{1}{2\mu} (x_0 \Lambda + x_1 (Q_J - \Lambda - \mu \text{Id})) \\ &\vdots \\ x_S &= -\frac{1}{S\mu} (x_{S-2} \Lambda + x_{S-1} (Q_J - \lambda - (S-1)\mu \text{Id})) \\ x_S &= x_{S-1} \Lambda (Q_J - S\mu \text{Id})^{-1}. \end{aligned}$$

Posons :

$$\begin{aligned} R_1 &= \frac{1}{\mu} (Q_J - \Lambda), \\ R_n &= -\frac{1}{n\mu} (R_{n-2} \Lambda + R_{n-1} (Q_J - \lambda - (n-1)\mu \text{Id})) \text{ pour } n = 2, \dots, S. \end{aligned} \quad (9.7)$$

On peut alors écrire :

$$x_n = x_{n-1} R_n \text{ et } x_0 (R_S - R_{S-1} \Lambda (Q_J - S\mu \text{Id})^{-1}) = 0.$$

Cela définit deux équations sur les deux composantes de x_0 . En fait, une seule est utile puisque si le système est rang 2, la seule solution est 0 ce qui est exclu. On en déduit l'algorithme de résolution suivant.

Algorithme 9.3 . Calcul de la probabilité invariante de la file IPP/M/S/S

Données : Λ, μ, S, Q_J

Résultat : π telle que $\pi A = 0$ et $\pi \cdot \epsilon = 1$

Calculer R_1, \dots, R_S selon l'équation (9.7);

$x_0(0, A) \leftarrow 1$;

Trouver $x_0(0, B)$ tel que x_0 satisfasse $x_0 (R_S - R_{S-1} \Lambda (Q_J - S\mu \text{Id})^{-1}) = 0$;

Calculer $x_n = x_{n-1} R_n$ pour $n = 1, \dots, S$;

Calculer $m = \sum_{n=0}^S x_n \cdot \epsilon$;

$\pi \leftarrow x/m$;

retourner π

THÉORÈME 9.5.– Dans une file IPP/M/S/S de probabilité invariante, la probabilité de perte est donnée par :

$$\frac{\lambda_A}{\lambda} \pi(S, A) + \frac{\lambda_B}{\lambda} \pi(S, B), \quad (9.8)$$

où λ est l'intensité moyenne donnée par (9.6).

Démonstration. On applique le théorème 7.24 à $\psi(s) = \mathbf{1}_{\{S\}}(X(s))$, il vient que la probabilité de perte est donnée par :

$$\frac{1}{\lambda} \lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t \mathbf{1}_{\{S\}}(X(s)) \lambda(J(s)) \, ds = \frac{1}{\lambda} (\lambda_A \pi(S, A) + \lambda_B \pi(S, B)),$$

d'où le résultat. \square

Fixer la charge revient à fixer λ mais suivant les valeurs relatives de λ_A , λ_B , σ_A et σ_B on obtient des probabilités de perte très différentes comme l'illustre le tableau 9.2. Nous avons choisi pour les calculs $S = 10$ serveurs et une charge de 5 Erlang. La formule d'Erlang donnerait une perte égale à 0,018.

Paramètres	Perte	Blocage
$\begin{pmatrix} \sigma_A = 1/2 & \sigma_B = 1/2 \\ \lambda_A = 10 & \lambda_B = 0 \end{pmatrix}$	0,1	0,05
$\begin{pmatrix} \sigma_B = 1/10 & \sigma_A = 9/10 \\ \lambda_A = 50 & \lambda_B = 0 \end{pmatrix}$	0,66	0,07
$\begin{pmatrix} \sigma_B = 1/100 & \sigma_A = 99/100 \\ \lambda_A = 500 & \lambda_B = 0 \end{pmatrix}$	0,96	0,01
$\begin{pmatrix} \sigma_B = 1/10 & \sigma_A = 9/10 \\ \lambda_A = 1 & \lambda_B = 5,444 \end{pmatrix}$	0,02	0,05

Tableau 9.2 – Probabilités de pertes à charge constante dans une file IPP/M/S/S

Il apparaît à la lecture de ce tableau que l'on peut augmenter la probabilité de perte en gardant une charge constante. On pourrait croire que le facteur déterminant devient alors la variance du processus d'entrée puisque plus l'on augmente la variabilité de celui-ci, plus la perte augmente. C'est certainement un critère mais ce n'est hélas pas le seul. Fixer la variance et la charge revient à imposer deux équations que doivent satisfaire les quatre paramètres. Il reste donc deux degrés de liberté que l'on peut utiliser pour faire varier la probabilité de perte comme bon nous semble.

On remarque aussi que la probabilité de blocage est désespérément loin de la probabilité de perte, il y a donc une erreur monumentale à assimiler l'une à l'autre.

9.7. Modèles d'Erlang généralisés

9.7.1. Canaux de garde

Les canaux de garde sont un système bien connu des téléphonistes qui permet de définir une priorité entre plusieurs flux sans altérer trop la probabilité de perte du flux le moins prioritaire.

Nous illustrons cette notion par la gestion des *hand-over* dans le réseau GSM. Lorsqu'un appel issu d'un téléphone mobile bouge, il arrive un moment où la BTS qui gérait cet appel perd le contact radio avec le mobile. Il est donc nécessaire de passer le relais à une autre BTS. Cette opération, appelée *hand-over*, nécessite de lourdes opérations du plan contrôle et l'on doit s'assurer qu'il y a suffisamment de ressources dans la nouvelle cellule pour acheminer ce nouvel appel.

Lorsque l'on prend en compte la mobilité des usagers, il est nécessaire d'introduire la durée de séjour dans la cellule de chaque usager. Pour pouvoir faire des calculs, on suppose que le temps de traversée d'une cellule suit une loi exponentielle de paramètre α . Le temps de communication d'un mobile vu de la BTS suit donc le minimum de la durée de communication et du temps de séjour. Compte tenu des propriétés de la loi exponentielle, cette durée de communication suit donc une loi exponentielle de paramètre $\mu + \alpha$. La probabilité qu'un usager « quitte » la cellule sur un *hand-over* est la probabilité qu'une variable aléatoire de loi exponentielle de paramètre μ se termine avant une variable aléatoire de loi exponentielle de paramètre α . On sait que cela vaut $\theta = \mu(\mu + \alpha)^{-1}$.

Considérons un ensemble de cellules de caractéristiques identiques et faisons le bilan des entrées-sorties dans une cellule donnée. Dans la figure 9.3, λ_f est le taux d'appels nouveaux ou « frais », λ_{HO} le nombre moyen d'appels en *hand-over* qui arrive dans la cellule étudiée, p la probabilité de perte dans la cellule.

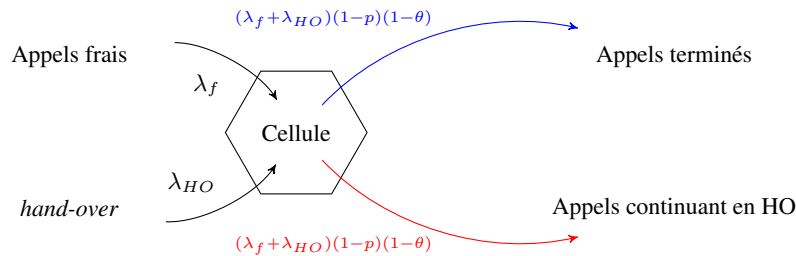


Figure 9.3 – Entrées-sorties d'une cellule

Comme le système est conçu de sorte que p soit négligeable devant 1, on a à l'équilibre :

$$\lambda_{HO} = (\lambda_f + \lambda_{HO})\theta \text{ soit } \lambda_{HO} = \frac{\theta}{1-\theta}\lambda_f.$$

La charge totale que doit traiter la BTS est donc :

$$\begin{aligned}
 \rho &= (\lambda_f + \lambda_{HO}) \times \frac{1}{\mu + \alpha} \\
 &= \lambda_f \left(1 + \frac{\theta}{1 - \theta}\right) \frac{1}{\mu + \alpha} \\
 &= \lambda_f \frac{1}{1 - \theta} \frac{1}{\mu + \alpha} \\
 &= \lambda_f \frac{\mu + \alpha}{\mu} \frac{1}{\mu + \alpha} \\
 &= \frac{\lambda_f}{\mu}.
 \end{aligned}$$

On ne peut toutefois pas appliquer la formule d'Erlang pour dimensionner le système avec *hand-over* car ceux-ci ont des besoins de qualité de service différents. Il est en effet bien plus désagréable d'avoir sa communication interrompue que de ne pouvoir en initier une. Pour donner une priorité aux appels en *hand-over*, on modifie la fonction de contrôle d'accès : on choisit $g < S$; tant que le nombre de serveurs libres est supérieur à g on accepte les appels frais et les appels en *hand-over*. Dès qu'il reste moins de g canaux disponibles, on n'accepte plus que les appels en *hand-over*. Pour calculer les probabilités de perte, on représente le système par un processus de Markov X comptant les serveurs occupés à chaque instant. C'est un processus de Markov dont les transitions peuvent se représenter par le diagramme 9.4 où $\lambda^+ = \lambda_f + \lambda_{HO}$ et $\mu^+ = \mu + \alpha$.

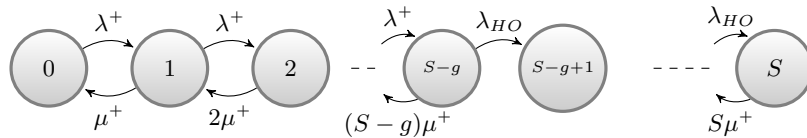


Figure 9.4 – Les transitions du processus « nombre de serveurs occupés » avec canaux de garde

A partir de là, on en déduit la probabilité invariante ν puis les probabilités de perte par la propriété PASTA. Puisque le trafic des appels frais est toujours un processus de Poisson d'intensité λ_f , le théorème A.38 implique que la probabilité de perte des appels frais est donnée par $\sum_{j=S-g}^S \nu(j)$ et celle des appels en *hand-over* par $\nu(S)$. On obtient les résultats numériques suivants.

S	30	32	34	34
g	0	2	4	3
Perte appels frais	0,8457%	1,0275%	1,0332%	0,6584%
Pertes HO	0,8457%	0,0278%	0,0008%	0,0028%

Tableau 9.3 – Pertes des appels frais et des appels en hand-over en fonction du nombre de canaux de garde

Il est tout à fait remarquable que pour de petites valeurs de g , on donne aux appels en *hand-over* une probabilité de perte bien inférieure à celle des appels frais sans pour autant trop pénaliser ces derniers.

9.7.2. Systèmes multiclassés

Dans de nombreuses situations, les clients ne réclament pas tous la même quantité de ressources. Dans ce cas là, on utilise la formule d'Erlang multiclassée. Soit un système comportant K classes de clients et S ressources. Les clients de classe i arrivent selon un processus de Poisson d'intensité λ_i et ont une durée de communication de loi exponentielle de moyenne $1/\mu_i$. On pose $\rho_i = \lambda_i/\mu_i$. Un client de classe i consomme s_i ressources. Les nombres de clients n_i de chaque classe sont donc soumis à la contrainte : $\sum_{i=1}^K n_i s_i \leq S$. On étudie le processus X qui comptabilise le nombre de ressources occupées. Son espace d'états est :

$$E = \{(n_1, \dots, n_K) \in \mathbf{N}, \sum_{j=1}^k n_j s_j \leq S\}.$$

THÉORÈME 9.6.– La probabilité invariante de ce système est donnée par :

$$\nu(n_1, \dots, n_K) = \frac{1}{G} \prod_{j=1}^k \rho^{n_j} / n_j!,$$

où G est la constante de normalisation :

$$G = \sum_{(n_1, \dots, n_K) \in E} \prod_{j=1}^k \rho^{n_j} / n_j!.$$

Démonstration. Si $S = \infty$, les composantes de X sont des processus de Markov indépendants qui évoluent respectivement comme le nombre de serveurs occupés dans une file $M_{\lambda_i}/M_{\mu_i}/\infty$. Ce sont donc des processus réversibles (voir définition 7.11) de loi invariante ν_i , loi de Poisson de paramètre ρ_i (voir l'exemple 10.3). Le processus X est donc réversible, sa probabilité invariante est le produit tensoriel de ces probabilités ν_i .

Pour S fini, X est juste la restriction à E de la dynamique précédente. En vertu du lemme de Kelly, on sait alors que la loi invariante de X est la restriction, renormalisation de la probabilité invariante dans le cas infini. \square

Soit ϵ_i le vecteur dont toutes les composantes sont nulles sauf la i^e qui vaut 1. Il y a perte d'un appel de classe i lorsque $X \in E$ mais $X + \epsilon_i$ n'appartient pas à E . La probabilité de perte p_i d'une classe donnée i est donc selon la propriété PASTA égale :

$$p_i = \sum_{n \in E, n + \epsilon_i \notin E} \nu(n_1, \dots, n_K).$$

EXEMPLE 9.3 (Interface A-bis).— Les progrès des codeurs de voix impliquent que là où un TCH par trame était nécessaire pour acheminer la communication, il n'y en a dorénavant besoin que qu'un demi. Cependant, certains appels ont toujours besoin d'un *slot* complet. Comme la théorie précédente nécessite des nombres entiers de ressources, il faut compter ici le nombre de demi-slots occupés. On a donc deux classes de clients, l'une avec $s_1 = 1$ et l'autre $s_2 = 2$. Prenons comme exemple une cellule avec trente TCH soit soixante demi-slots. Le tableau suivant donne les taux de perte de chaque classe en fonction des charges.

		Charge des appels de classe 1			
		10	20	30	40
Charge des appels de classe 2	10	0,004 %	0,24	2,47 %	8,34 %
	20	0,1 %	0,55 %	5,29 %	16,71 %
	20	2,62 %	7,48 %	13,9 %	20,62 %
	30	5,53 %	14,99 %	26,54 %	37,66 %

Tableau 9.4 – Taux de perte des différentes classes en fonction de la charge

9.8. Réseaux hiérarchiques

On a déjà vu que plus l'on augmente le nombre de serveurs moindre est la perte. Une bonne mesure de cette rentabilité peut être le coût de l'Erlang transporté : considérons un système à S serveurs, sans salle d'attente pour lequel on fixe une borne supérieure α à la probabilité de perte. Le trafic maximal $\rho_{S,\alpha}$ qui peut passer dans ce système est défini par l'équation :

$$\text{Er}[\rho_{S,\alpha}, S] = \alpha.$$

Le rendement du système est alors défini comme le rapport de $\rho_{S,\alpha}$ par S :

$$R = \frac{\rho_{S,\alpha}}{S}.$$

Comme le montre la figure 9.5, le rendement augmente quand le nombre de serveurs augmente.

Mais si l'on prend en compte le coût de l'installation d'un serveur, il vient un moment où l'augmentation du coût compense l'augmentation du gain de rendement. La solution mise en œuvre dans le réseau téléphonique classique consiste à hiérarchiser le réseau. Les abonnés sont tous reliés à un commutateur que l'on qualifiera ici de commutateur de niveau 0, les commutateurs de niveau 0 sont reliés à des commutateurs reliés à des commutateurs de niveau 1 en moindre nombre. On continue ainsi jusqu'au niveau 3 dans la hiérarchie. Pour mémoire, du temps de sa splendeur, le réseau téléphonique commuté français comportait environ 1 500 commutateurs de niveau 0, quelques centaines de commutateurs de niveau 1 et 2 et sept commutateurs de niveau 3.

L'établissement d'une communication se faisait en tentant toujours d'utiliser des jonctions de plus bas niveau possible. S'il existait une jonction entre les deux commutateurs de niveau 0, on la prenait. Sinon, le commutateur de niveau 0 envoyait la gestion de l'appel à son commutateur de niveau 1, qui lui-même essayait d'acheminer l'appel en restant à son niveau, etc.

Si le RTC en lui-même est aujourd'hui bien obsolète, il n'empêche que nombre de ses inventions ont été reprises dans des systèmes modernes, notamment les canaux de garde et les réseaux hiérarchiques.

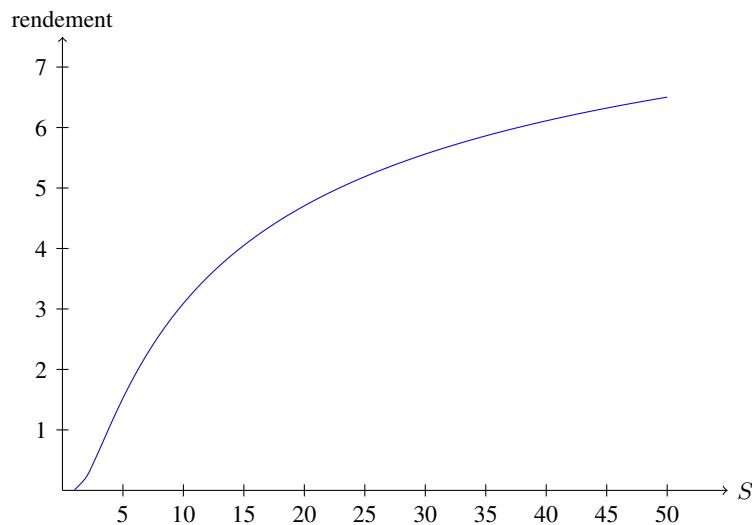


Figure 9.5 – Le rendement en fonction de S

Prenons le cas du GSM. En milieu urbain, les cellules, c'est-à-dire la zone gérée par une antenne ou station de base, sont petites et les *hand-over* fréquents pour les usagers « rapides » tels les automobilistes. On ajoute donc des antennes plus puissantes que les autres, qui couvrent une zone plus large et qui vont offrir deux services. Premièrement, prendre en charge les appels des mobiles rapides pour ainsi diminuer la fréquence des *hand-over*. Deuxièmement, acheminer les appels en débordement des microcellules qu'elles couvrent.

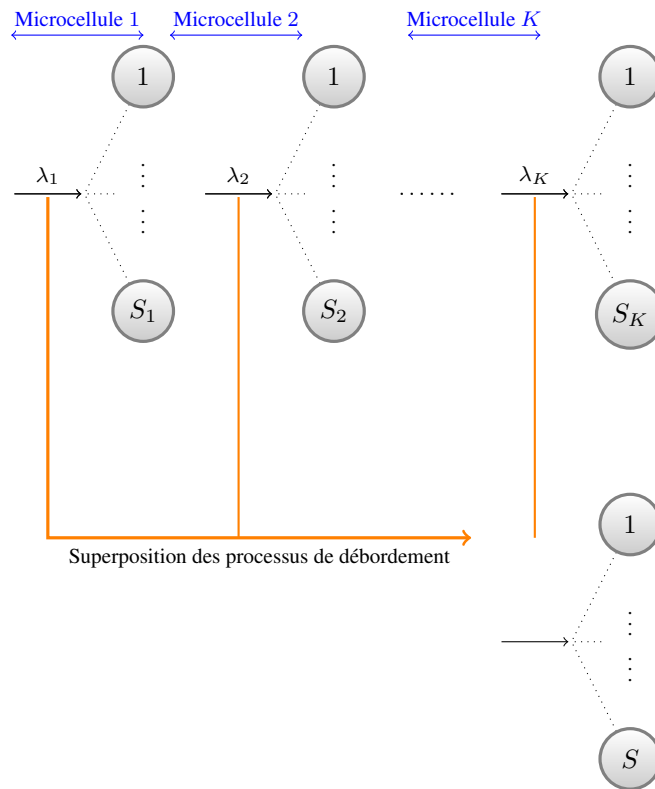


Figure 9.6 – Le trafic écoulé par la macrocellule est la somme des trafics de débordement des microcellules

L'une des questions qui se pose est celle du dimensionnement du nombre d'émetteurs-récepteurs dans la cellule de plus haut niveau. En effet, le processus des arrivées en débordement d'une microcellule n'est pas un processus de Poisson mais un MMPP (voir 7.6, exemple 7.2). Pour la microcellule numéro i , le processus des phases du processus de débordement a pour générateur infinitésimal Q_i celui du nombre de clients

Cela détermine les x_i en fonction de x_0 . Pour déterminer x_0 , nous avons deux possibilités.

– *Méthode 1.* La probabilité stationnaire du nombre de serveurs occupés dans chaque cellule est donnée par la formule d’Erlang :

$$\nu(i) = \frac{\rho^i / i!}{\sum_{j=0}^S \rho^j / j!}.$$

Comme les cellules sont indépendantes les unes des autres, la probabilité stationnaire du processus des phases est le produit tensoriel de ces trois vecteurs :

$$\nu_{mc} = \nu \otimes \nu \otimes \nu.$$

On remarque que :

$$\sum_{n=0}^L \pi(n, j) = \nu_{mc}(j).$$

On introduit alors f_n le vecteur ligne à m composantes dont seule la n^e est non nulle et vaut 1. On introduit aussi e_n le transposé de f_n . La dernière relation s’écrit alors :

$$\sum_{n=0}^L x_0 R_n \cdot e_j = \nu_{mc}(j).$$

Comme $e_j \cdot f_j$ est la matrice dont le seul coefficient non nul est le coefficient (j, j) , on a $\sum_{j=1}^m e_j \cdot f_j = Id_m$. D’autre part :

$$\nu_{mc} = \sum_{j=1}^m \nu_{mc}(j) f_j,$$

on a donc :

$$x_0 \sum_{l=0}^L R_l = \nu_{mc}. \quad (9.11)$$

– *Méthode 2.* En utilisant (9.10) dans (9.9), on obtient :

$$x_0 (R_{L-1} \Lambda_{mc} + R_L (Q_{mc} - L\theta)) = 0.$$

Cette équation détermine toutes les composantes de x_0 sauf une. Ensuite, il faut utiliser la condition de normalisation pour calculer la valeur de cette composante.

Enfin, en vertu du théorème 7.24, la probabilité de perte est donnée par :

$$\left(\nu_{mc} \Lambda_{mc} e \right)^{-1} x_L \Lambda_{mc} e, \quad (9.12)$$

où $e = \sum_{n=1}^m e_n$.

9.9. Un modèle avec renoncement

On modélise ici une limitation forcée de la charge du système, non pas par la limitation de sa capacité, mais par l'introduction d'un filtre aléatoire à l'entrée : plus le système est engorgé, moins le client a de chance d'y entrer.

Avec les notations habituelles, on considère un système $M_\lambda/M_\mu/S$, dans lequel les clients qui arrivent font un tirage au sort (indépendant d'un client à l'autre et tous les autres paramètres) pour décider de leur entrée dans le système. On note $(X^R(t), t \geq 0)$, le processus comptant le nombre de clients dans le système (et tous les paramètres seront affectés de l'exposant R pour « renoncement »). Pour le n -ème client entrant, le tirage au sort est une expérience de Bernoulli de probabilité $p(X^R(T_n^-))$, dépendant donc de la congestion à l'arrivée du client. Si la variable de Bernoulli vaut 1, le client entre et attendra dans le système jusqu'à la fin de son service. Sinon, le client ne rentre même pas dans le système et est définitivement perdu. Il est donc cohérent de supposer que la fonction $p(\cdot)$ soit décroissante et nous considérerons par exemple ici que :

$$p(n) = \frac{1}{n+1}, \quad (9.13)$$

qui est la probabilité d'entrée principalement considérée dans la littérature. Dans la suite, on appellera « client entrant », un client arrivant et entrant effectivement dans le système.

Nous allons décrire le comportement stationnaire du système par l'étude du processus $(X^R(t), t \geq 0)$. Commençons par remarquer que l'on a dans ce cas, l'analogie du paradoxe de l'autobus.

LEMME 9.7.— Soit pour tout instant t , $\tilde{W}(t)$ le temps résiduel à t avant la prochaine arrivée d'un client entrant. Pour tout $i \in \mathbf{N}$, $\tilde{W}(t)$ suit la loi $\varepsilon(\lambda p_i)$ conditionnellement à $X^R(t) = i$, c'est-à-dire que pour $x \geq 0$, on a :

$$\mathbf{P}(\tilde{W}(t) \leq x | X^R(t) = i) = 1 - e^{-\lambda p_i x}.$$

Démonstration. Notons à nouveau $T_0 = 0$ et T_1, T_2, \dots les instants d'arrivées des clients et pour tout $t \geq 0$, $N(t)$ le nombre de clients arrivés jusqu'à t et $Z(t)$, le nombre d'arrivées nécessaires après t pour voir entrer effectivement le premier client après t . Alors, il est facile de voir que :

$$\tilde{W}(t) = T_{N(t)+Z(t)} - t.$$

D'autre part, $Z(t)$ suit conditionnellement à $\{X^R(t) = i\}$ une loi géométrique de paramètre p_i , puisqu'alors, chaque client entre à partir de t avec probabilité p_i indépendamment des autres et ce jusqu'à la première entrée effective après t .

Pour tout x , on peut donc écrire :

$$\begin{aligned}
& \mathbf{P} \left(\tilde{W}(t) \geq x \mid X^R(t) = i \right) \\
&= \sum_{k \geq 1} \mathbf{P} \left(\tilde{W}(t) \geq x \mid Z(t) = k; X^R(t) = i \right) \mathbf{P} (Z(t) = k \mid X^R(t) = i) \\
&= \sum_{k \geq 1} \mathbf{P} \left(\tilde{W}(t) \geq x \mid Z(t) = k; X^R(t) = i \right) (1 - p_i)^{k-1} p_i \\
&= p_i \sum_{k \geq 1} (1 - p_i)^{k-1} \mathbf{P} (T_{N(t)+k} - t \geq x) \\
&= p_i \sum_{k \geq 1} (1 - p_i)^{k-1} \sum_{j=1}^{\infty} \mathbf{P} (T_{j+k} - t \geq x; N(t) = j).
\end{aligned} \tag{9.14}$$

D'après le paradoxe de l'autobus :

$$p_i \sum_{j=1}^{\infty} \mathbf{P} (T_{j+1} - t \geq x; N(t) = j) = p_i e^{-\lambda x}. \tag{9.15}$$

D'autre part :

$$\begin{aligned}
& p_i \sum_{k \geq 2} (1 - p_i)^{k-1} \mathbf{P} (T_k - t \geq x; N(t) = 0) \\
&= p_i \sum_{k \geq 2} (1 - p_i)^{k-1} \mathbf{P} (T_k \geq t + x; \xi_1 > t) \\
&= p_i \sum_{k \geq 2} (1 - p_i)^{k-1} \{ \mathbf{P} (\xi_1 > t + x) + \mathbf{P} (T_k \geq t + x; t < \xi_1 < t + x) \} \\
&= p_i \sum_{k \geq 2} (1 - p_i)^{k-1} e^{-\lambda(t+x)} \\
&\quad + p_i \sum_{k \geq 2} (1 - p_i)^{k-1} \int_t^{t+x} \lambda e^{-\lambda u} \int_{t+x-u}^{\infty} \lambda^{k-1} \frac{v^{k-2}}{(k-2)!} e^{-\lambda v} \mathrm{d}v \mathrm{d}u \\
&= e^{-\lambda(t+x)} (1 - p_i) + \int_t^{t+x} \lambda e^{-\lambda u} (1 - p_i) e^{-\lambda p_i(t+x-u)} \mathrm{d}u \\
&= -p_i e^{-\lambda(t+x)} + e^{-\lambda t} e^{-\lambda p_i x}. \tag{9.16}
\end{aligned}$$

Ensuite :

$$\begin{aligned}
& p_i \sum_{k \geq 2} (1 - p_i)^{k-1} \sum_{j=1}^{\infty} \mathbf{P}(T_{j+k} - t \geq x; N(t) = j) \\
&= p_i \sum_{k \geq 2} (1 - p_i)^{k-1} \sum_{j=1}^{\infty} \mathbf{P}(T_j \leq t; T_j + \xi_{j+1} > t + x) \\
&+ p_i \sum_{k \geq 2} (1 - p_i)^{k-1} \sum_{j=1}^{\infty} \mathbf{P}(T_{j+k} \geq t + x; T_j \leq t; t < T_j + \xi_{j+1} \leq t + x).
\end{aligned} \tag{9.17}$$

Or, comme on l'a vu dans la preuve du paradoxe de l'autobus :

$$\begin{aligned}
& p_i \sum_{k \geq 2} (1 - p_i)^{k-1} \sum_{j=1}^{\infty} \mathbf{P}(T_j \leq t; T_j + \xi_{j+1} > t + x) \\
&= p_i \sum_{k \geq 2} (1 - p_i)^{k-1} e^{-\lambda(t+x)} (e^{\lambda t} - 1) \\
&= (e^{-\lambda x} - e^{-\lambda(t+x)}) (1 - p_i),
\end{aligned} \tag{9.18}$$

alors que :

$$\begin{aligned}
& p_i \sum_{k \geq 2} (1 - p_i)^{k-1} \sum_{j=1}^{\infty} \mathbf{P}(T_{j+k} \geq t + x; T_j \leq t; t < T_j + \xi_{j+1} \leq t + x) \\
&= p_i \int_0^t \sum_{j=1}^{\infty} \left(\lambda^j \frac{u^{j-1}}{(j-1)!} \right) e^{-\lambda u} \int_{t-u}^{t+x-u} \lambda e^{-\lambda v} \\
&\quad \left(\int_{t+x-u-v}^{\infty} \sum_{k=2}^{\infty} \left((1 - p_i)^{k-1} \lambda^{k-1} \frac{w^{k-2}}{(k-2)!} \right) e^{-\lambda w} dw \right) dv du. \\
&= \int_0^t \lambda e^{-\lambda p_i(t+x-u)} \int_{t-u}^{t+x-u} \lambda (1 - p_i) e^{-\lambda(1-p_i)v} dv du \\
&= (e^{-\lambda t} e^{-\lambda p_i x} - e^{-\lambda(t+x)}) \int_0^t \lambda e^{\lambda u} du \\
&= e^{-\lambda p_i x} - e^{-\lambda x} - e^{-\lambda t} e^{-\lambda p_i x} + e^{-\lambda(t+x)}.
\end{aligned} \tag{9.19}$$

Probabilité de perte

Définissons le processus $(Y(t), t \geq 0)$ par récurrence sur les instants d'arrivée de la manière suivante. On fixe :

$$Y(t) = 0; t \in [0, T_1[,$$

puis pour tout $i \geq 1$:

$$Y(T_i) = \begin{cases} 1 & \text{si le client } C_i \text{ entre dans le système;} \\ 0 & \text{sinon,} \end{cases}$$

et :

$$Y(t) = Y(T_i), t \in [T_i, T_{i+1}[.$$

Il est alors facile de vérifier que le processus couple $((X^R(t), Y(t)), t \geq 0)$ est markovien sur $\mathbf{N} \times \{0, 1\}$; en effet, c'est un processus à trajectoires cadlag (on remarquera que $(Y(t), t \geq 0)$ est constant par morceaux) dont on peut écrire le générateur \tilde{A}^R comme suit :

– pour tout $i \geq 0$, le processus peut quitter l'état $(i, 0)$ pour l'état $(i + 1, 1)$ si une arrivée d'un client effectivement entrant se produit et si $i \geq 1$, pour l'état $(i - 1, 0)$ si un service se termine donc :

$$\tilde{A}^R((i, 0), (i + 1, 1)) = \lambda p_i;$$

$$\tilde{A}^R((i, 0), (i - 1, 0)) = \mu \text{ pour } i \geq 1;$$

– le processus a les mêmes transitions de $(i, 1)$ vers $(i + 1, 1)$ et $(i - 1, 1)$;
– un autre saut possible se fait de $(i, 1)$ vers $(i, 0)$, lorsqu'un client arrive et ne rentre pas donc :

$$\tilde{A}^R((i, 1), (i + 1, 1)) = \lambda p_i;$$

$$\tilde{A}^R((i, 1), (i, 0)) = \lambda(1 - p_i);$$

$$\tilde{A}^R((i, 1), (i - 1, 1)) = \mu \text{ pour } i \geq 1,$$

où l'on applique un résultat analogue au lemme 9.7 pour la deuxième transition.

Nous résolvons le système :

$$\begin{cases} \tilde{\pi}^R \tilde{A}^R &= \mathbf{0}, \\ \tilde{\pi}^R \mathbf{e} &= 1, \end{cases}$$

où \tilde{A}^R est le générateur du processus sur $\mathbf{N} \times \{0, 1\}$ et $\tilde{\pi}^R$ est une probabilité sur $\mathbf{N} \times \{0, 1\}$.

Appelons X_∞^R et Y_∞^R , les variables aléatoires limites des deux processus, si elles existent. Le cas échéant, $\tilde{\pi}^R$ doit vérifier pour tout $i \in \mathbf{N}$ que :

$$\begin{aligned}\pi^R(i) &= \mathbf{P}(X_\infty = i) = \mathbf{P}(X_\infty = i; Y_\infty = 0) + \mathbf{P}(X_\infty = i; Y_\infty = 1) \\ &= \tilde{\pi}^R(i, 0) + \tilde{\pi}^R(i, 1).\end{aligned}\quad (9.22)$$

En particulier, une unique solution au système précédent existe nécessairement et donc le processus $((X^R(t), Y(t)), t \geq 0)$ est ergodique, à la condition de stabilité (9.20). En effet, les séries mises en jeu dans le calcul de $\tilde{\pi}^R$ sont alors nécessairement sommables d'après (9.22). Remarquons ensuite que pour tout $i \in \mathbf{N}$:

$$\mu\pi^R(i+1) - \lambda p_i \pi^R(i) = 0,$$

ce qui donne avec (9.22) que :

$$\mu\tilde{\pi}^R(i+1, 0) + \mu\tilde{\pi}^R(i+1, 1) - \lambda p_i \tilde{\pi}^R(i, 0) - \lambda p_i \tilde{\pi}^R(i, 1) = 0. \quad (9.23)$$

On lit par ailleurs sur le générateur \tilde{A}^R que :

$$-\lambda p_i \tilde{\pi}^R(i, 0) - \mu \tilde{\pi}^R(i, 0) + \lambda \tilde{\pi}^R(i, 1) - \lambda p_i \tilde{\pi}^R(i, 1) + \mu \tilde{\pi}^R(i+1, 0) = 0,$$

ce qui en combinant avec (9.23) implique que pour tout $i \geq 0$, on a :

$$\begin{aligned}\tilde{\pi}^R(i+1, 0) &= (\rho p_i + 1) \tilde{\pi}^R(i, 0) + \rho(p_i - 1) \tilde{\pi}^R(i, 1); \\ \tilde{\pi}^R(i+1, 1) &= -\tilde{\pi}^R(i, 0) + \rho \tilde{\pi}^R(i, 1).\end{aligned}$$

On a donc la relation de récurrence matricielle :

$$\begin{pmatrix} \tilde{\pi}^R(i+1, 0) \\ \tilde{\pi}^R(i+1, 1) \end{pmatrix} = \begin{pmatrix} \rho p_i + 1 & \rho(p_i - 1) \\ -1 & \rho \end{pmatrix} \begin{pmatrix} \tilde{\pi}^R(i, 0) \\ \tilde{\pi}^R(i, 1) \end{pmatrix} \quad (9.24)$$

$$= M_i \begin{pmatrix} \tilde{\pi}^R(i, 0) \\ \tilde{\pi}^R(i, 1) \end{pmatrix}. \quad (9.25)$$

Pour tout $i \geq 1$, notons :

$$A_i = \prod_{j=0}^{i-1} M_j$$

et A_i^1 (respectivement A_i^2) la première (respectivement deuxième) ligne de A_i et rappelons les relations (9.22) et $\sum_{i \in \mathbf{N}} (\tilde{\pi}^R(i, 0) + \tilde{\pi}^R(i, 1)) = 1$. La probabilité $\tilde{\pi}^R$ est

donc complètement définie par :

$$\begin{aligned} \begin{pmatrix} \tilde{\pi}^R(i, 0) \\ \tilde{\pi}^R(i, 1) \end{pmatrix} &= A_i \begin{pmatrix} \tilde{\pi}^R(0, 0) \\ \tilde{\pi}^R(0, 1) \end{pmatrix} \\ &= A_i \left\{ \tilde{\pi}^R((0, 0)) \begin{pmatrix} 1 \\ -1 \end{pmatrix} + \begin{pmatrix} 0 \\ \pi^R(0) \end{pmatrix} \right\}, i \geq 0; \\ \tilde{\pi}^R(0, 0) &= \left[1 - \sum_{i \in \mathbf{N}} \left\{ A_i^1 \begin{pmatrix} 0 \\ \pi^R(0) \end{pmatrix} + A_i^2 \begin{pmatrix} 0 \\ \pi^R(0) \end{pmatrix} \right\} \right] \\ &\quad \cdot \left[\sum_{i \in \mathbf{N}} \left\{ A_i^1 \begin{pmatrix} 1 \\ -1 \end{pmatrix} + A_i^2 \begin{pmatrix} 1 \\ -1 \end{pmatrix} \right\} \right]^{-1}. \end{aligned}$$

On peut donc calculer numériquement $\tilde{\pi}^R$ en estimant les valeurs des séries de la formule précédente. La probabilité de perte est alors donnée, d'après la propriété PASTA et l'ergodicité du processus $((X(t), Y(t)), t \geq 0)$, par la formule :

$$\begin{aligned} P_{\text{PERTE}}^R &= \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \mathbf{1}_{\{0\}}(Y(T_n^-)) \\ &= \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \sum_{i=0}^{\infty} \mathbf{1}_{\{(i, 0)\}}((X^R(T_n^-), Y(T_n^-))) \\ &= \sum_{i=0}^{\infty} \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \mathbf{1}_{\{(i, 0)\}}((X^R(t), Y(t))) dt \\ &= \sum_{i=0}^{\infty} \tilde{\pi}^R((i, 0)). \end{aligned}$$

9.10. Un centre d'appels avec mise en attente et clients impatientes

On a vu que le modèle d'Erlang modélise par exemple un centre d'appels à s postes, où les appels des clients n'aboutissent que si l'une des lignes est libre au moment de leur prise d'appel. Comme l'on va le voir ici, on peut enrichir le modèle précédent dans le cas, où les clients peuvent être mis en attente. Il est alors cohérent de supposer que les clients sont susceptibles de s'impatienter et de raccrocher avant d'avoir vu leur appel aboutir.

Le modèle choisi est donc celui d'une file d'attente avec capacité infinie et clients impatientes : on considère plus précisément une file $M_\lambda/M_\mu/S/S+M_\alpha$, c'est-à-dire que les temps d'interarrivées et les durées d'appels sont exponentiels et les temps de patience des clients $(D_n, n \in \mathbf{Z})$ avant d'accéder à un opérateur sont indépendants

et identiquement distribués de loi $\varepsilon(\alpha)$, où $\alpha > 0$. Soulignons bien ici, comme dans les modèles avec impatience étudiés au chapitre 4, que l'impatience des clients dure jusqu'à l'accès à un opérateur et tout client qui n'a pas raccroché avant cet instant, poursuit alors son appel jusqu'à son terme : avec la terminologie de la section 4.6, la patience court jusqu'au début du service. Dans toute cette section, les clients sont servis en FCFS.

Notons $(X^1(t), t \geq 0)$, le processus comptant le nombre de clients dans le système à chaque instant (et ajoutons l'exposant ¹ pour « impatience » à tous les paramètres du système). A t , $X^1(t)$ compte les clients en service et les clients en attente, aussi bien ceux qui atteindront un serveur un jour que ceux qui s'impatieront avant d'y parvenir. Ce processus peut quitter l'état $i \in \mathbb{N}$ pour visiter les états suivants :

- l'état $i + 1$, si une arrivée se produit;
- l'état $i - 1$, si $i > S$ et l'un des i clients en attente quitte la salle d'attente, soit parce qu'un serveur s'est libéré, soit parce que le temps de patience du client a expiré.

Comme les temps résiduels de patience des $i - S$ clients en attente sont indépendants et suivent à tout instant la loi $\varepsilon(\alpha)$, il est donc aisé de vérifier que $(X^1(t), t \geq 0)$ est markovien, de générateur infinitésimal donné par :

- $A^1(i, i + 1) = \lambda$ pour tout $i \geq 1$;
- $A^1(i, i - 1) = i\mu$ pour tout $i \in \llbracket 1, S \rrbracket$;
- $A^1(i, i - 1) = S\mu + (i - S)\alpha$ pour tout $i \geq S + 1$.

On calcule alors, comme habituellement, la probabilité stationnaire π^1 de $(X^1, t \geq 0)$:

$$\begin{aligned} \pi^1(i) &= \frac{\rho^i}{i!} \pi^1(0) \text{ pour tout } i \in \llbracket 0, S \rrbracket; \\ \pi^1(i) &= \frac{\lambda^{i-S} \rho^S}{\prod_{j=1}^{i-S} (S\mu + j\alpha) S!} \pi^1(0) \text{ pour tout } i \geq S + 1; \\ \pi^1(0) &= \left(\sum_{i=0}^S \frac{\rho^i}{i!} + \sum_{i=1}^{\infty} \frac{\lambda^i \rho^S}{\prod_{j=1}^i (S\mu + j\alpha) S!} \right)^{-1}. \end{aligned}$$

Estimation de la probabilité de perte

Comme dans le modèle d'Erlang, on cherche à dimensionner le système en déterminant la valeur optimale de S pour garantir une probabilité de perte cible. Le calcul exact de cette probabilité est fastidieux et repose sur des arguments techniques qui dépassent le cadre de cet exposé. Nous pouvons néanmoins en donner une estimation heuristique.

Dans ce système, le client C_n est perdu si et seulement s'il trouve dans le système un temps d'attente W_n supérieur à son temps de patience D_n . La probabilité de perte est donc donnée par :

$$P_{\text{PERTE}}^I = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \mathbf{1}_{\{0, D_n\}}(W_n)$$

D'après les résultats de la section 4.6, il existe un temps d'attente W_∞ conjointement stationnaire avec les temps de patience. A supposer que les variables aléatoires limites W_∞ et D soient effectivement indépendantes (comme le sont W_n et D_n pour tout n), le théorème 2.7 nous donne donc que :

$$\begin{aligned} P_{\text{PERTE}}^I &= \mathbf{P}((W_\infty, D) \in \{(x, y); x \geq y\}) \\ &= \int_{\mathbf{R}_+} \int_0^x \alpha e^{-\alpha y} dy P_{W_\infty}(dx) \\ &= 1 - \mathcal{L}_{W_\infty}(\alpha), \end{aligned}$$

où \mathcal{L}_{W_∞} est la transformée de Laplace de W_∞ .

9.11. Problèmes

EXERCICE 22.— On considère le concentrateur suivant.

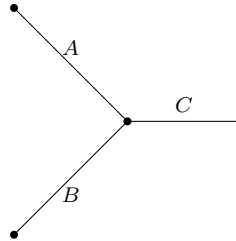


Figure 9.7 – Concentrateur simple

Les clients de classe 1 sont ceux qui empruntent les liaisons A et C. Les clients de classe 2 sont ceux qui empruntent les liaisons B et C. On note S_A, S_B, S_C les capacités de chacun des liens. On pose $X_t = (X_t^1, X_t^2)$, avec X_t^i le nombre d'appels de classe i en cours. Les arrivées des appels de classe i forment un processus de Poisson d'intensité $\lambda_i, i = 1, 2$. Les durées des appels de classe i sont exponentiellement distribuées de moyenne $1/\mu_i$. On pose $\rho_i = \lambda_i/\mu_i$.

- 1) On suppose dans un premier temps que $S_A = S_B = S_C = \infty$. Ecrire le générateur infinitésimal de X .
- 2) Quelle est sa probabilité stationnaire ?
- 3) Montrer qu'il est réversible.
- 4) Décrire l'espace d'états S lorsque les capacités sont finies.
- 5) En déduire la probabilité stationnaire π de X lorsque les capacités sont toutes finies.
- 6) Montrer que la probabilité p_i de blocage (et donc de pertes) des appels de classe i est de la forme :

$$p_i = 1 - \frac{\sum_{(n_1, n_2) \in S_i} \pi(n_1, n_2)}{\sum_{(n_1, n_2) \in S} \pi(n_1, n_2)},$$

où S_i est un sous-ensemble de S que l'on précisera.

- 7) Application numérique : $S_A = S_B = 2$, $S_C = 3$, $\rho_1 = \rho_2 = 2$. Calculer p_1 .
- 8) Comparer cette probabilité de perte à la quantité :

$$1 - (1 - E[\rho_1, S_A])(1 - E[\rho_1, S_C]).$$

Pourquoi considère-t-on cette quantité ?

9.12. Notes et commentaires

Les calculs explicites pour le dimensionnement avec *hand-over* sont issus d'un cours de X. Lagrange, ceux pour l'interface A-bis ont été faits par N. Dailly. Le dimensionnement des réseaux hiérarchiques est un ancien problème. Les anciennes méthodes reposaient sur le faisceau équivalent de Wilkinson et la méthode Kuczura (voir [IVE 01]). L'approche par les processus MMPP date de [MEI 89]. Son application aux réseaux radio mobiles est inspirée de [LAG 96].

Résumé

– La probabilité de perte n'est égale à la probabilité de blocage que si le processus d'arrivée est un processus de Poisson. Dans les autres cas, on peut se référer au théorème A.34.

– Dans la file $M/M/S/S$, la probabilité de perte est donnée par la formule d'Erlang :

$$\text{Er}[\rho, S] = \frac{\rho^S / S!}{\sum_{i=0}^S \rho^i / i!}.$$

– Dans le modèle d'Engset, la probabilité de perte est donnée par :

$$\text{Eng}[\rho, S, M] = \frac{\rho^S C_{M-1}^S}{\sum_{j=0}^S C_{M-1}^j \rho^j}.$$

où S est le nombre de serveurs et M le nombre de clients.

– La probabilité de perte ne dépend pas uniquement de la charge, voir le cas de la file IPP/M/S/S.

TROISIÈME PARTIE
Modélisation spatiale

Chapitre 10

Processus ponctuels spatiaux

10.1. Motivation

Dans les communications radio, la distance entre l'émetteur et le récepteur joue un rôle crucial. Pour évaluer les performances de protocoles radio-cellulaires, il est d'usage de considérer que les points d'accès ou station de base sont répartis de façon régulière selon un schéma hexagonal, voir figure 10.1.

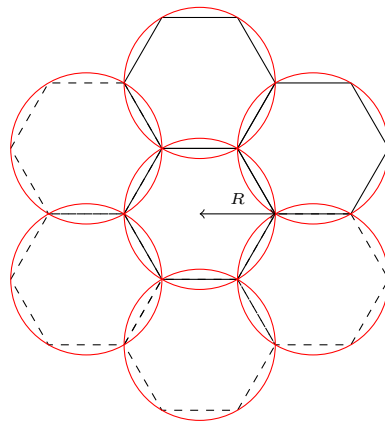


Figure 10.1 – Réseau hexagonal de stations de base

Les mobiles sont, quant à eux, modélisés par un *continuum* : un appel peut être émis d'un point x quelconque avec une probabilité indépendante de x . Cette approche

très macroscopique empêche des calculs vraiment précis et réalistes. Depuis quelques années, sous l'impulsion des travaux de F. Baccelli, les modèles issus de la géométrie stochastique rencontrent de plus en plus d'attention. Ils permettent de représenter plus finement la réalité et de faire des calculs plus rigoureusement.

10.2. Géométrie stochastique

La notion de configuration est précisée dans l'exemple A.1. Rappelons-en la définition, voir la section A.1.2 pour les notations.

DÉFINITION 10.1.– Une configuration est un ensemble de points d'un ensemble E localement fini : il y a un nombre fini de points dans tout ensemble borné. On note \mathfrak{N}_E l'ensemble des configurations de E .

EXEMPLE 10.1 (Processus de Bernoulli).– Le processus ponctuel de Bernoulli est un processus où E est un ensemble fini $E = \{x_1, \dots, x_n\}$. Chacun de ces points est « allumé » indépendamment des autres et avec probabilité p . Si l'on introduit A_1, \dots, A_n des variables aléatoires indépendantes de loi de Bernoulli de paramètre p , on peut écrire :

$$N = \sum_{i=1}^n A_i \delta_{x_i}.$$

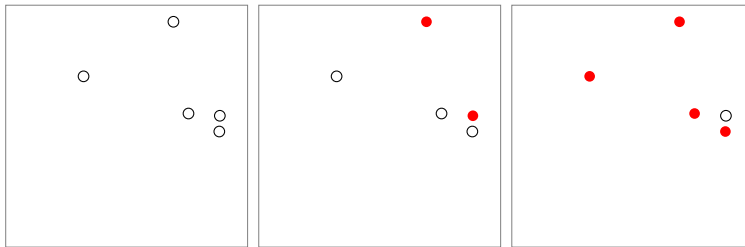


Tableau 10.1 – A gauche, l'ensemble E ; au milieu et à droite, deux réalisations possibles ; en plein, les points « allumés »

EXEMPLE 10.2 (Processus binomial).– Le nombre de points est fixé à n et l'on se donne μ , une mesure de probabilités sur \mathbf{R}^2 . Les atomes sont tirés au hasard indépendamment les uns des autres selon μ .

On peut aisément calculer que :

$$\mathbf{P}(N(A) = k) = \binom{n}{k} \mu(A)^k (1 - \mu(A))^{n-k}$$

et que pour des ensembles disjoints A_1, \dots, A_n , on a :

$$\mathbf{P}(N(A_1) = k_1, \dots, N(A_n) = k_n) = \frac{(k_1 + \dots + k_n)!}{k_1! \dots k_n!} \mu(A_1)^{k_1} \dots \mu(A_n)^{k_n}. \quad (10.1)$$

10.3. Processus de Poisson

Le processus ponctuel mathématiquement le plus riche est le processus de Poisson spatial dont on reconnaîtra qu'il généralise le processus de Poisson sur la droite réelle introduit dans le chapitre 6.

DÉFINITION 10.2.– Soit μ une mesure de Radon sur E polonais, c'est-à-dire que $\mu(\Lambda) < \infty$ pour tout ensemble compact $\Lambda \subset E$. Le processus de Poisson d'intensité μ est défini par sa transformée de Laplace. Pour toute fonction $f : E \rightarrow \mathbf{R}^+$, on a :

$$\mathbf{E} \left[\exp\left(-\int f \, dN\right) \right] = \exp\left(-\int_E (1 - e^{-f(s)}) \, d\mu(s)\right).$$

Pour préciser que la mesure intensité est μ , on indexera souvent l'espérance par μ . Par dérivation, on déduit immédiatement de la définition la formule de Campbell.

THÉORÈME 10.1 (Formule de Campbell).– Soit $f \in L^1(E, \mu)$, on a :

$$\mathbf{E}_\mu \left[\int f \, dN \right] = \int_E f \, d\mu.$$

Soit $f \in L^2(E \times E, \mu \otimes \mu)$, on a :

$$\mathbf{E}_\mu \left[\sum_{x \neq y \in N} f(x, y) \right] = \iint_{E \times E} f(x, y) \, d\mu(x) \, d\mu(y).$$

REMARQUE.– En particulier, pour $f = \mathbf{1}_A$ où A est un compact de E , on voit que $\mathbf{E}[N(A)] = \mu(A)$. Si $\mu = \lambda \, dx$ alors λ représente le nombre moyen de clients par unité de surface.

THÉORÈME 10.2.– Soit μ une mesure de Radon sur E polonais. Le processus de Poisson d'intensité μ est la probabilité sur \mathfrak{N}_E telle que :

- pour tout ensemble compact $\Lambda \subset E$, $N(\Lambda)$ suit une loi de Poisson de paramètre $\mu(\Lambda)$;
- pour Λ_1 et Λ_2 deux sous-ensembles de $(E, \mathfrak{B}(E))$ disjoints, les variables aléatoires $N(\Lambda_1)$ et $N(\Lambda_2)$ sont indépendantes.

De cette deuxième définition, on déduit immédiatement le résultat « d'uniformité » suivant.

THÉORÈME 10.3.– Soit N un processus de Poisson d'intensité μ . Soit $\Lambda \subset E$ un ensemble compact. Conditionnellement à $N(\Lambda) = n$, les atomes sont répartis selon un processus binomial pour $\mu_\Lambda(A) = \mu(A \cap \Lambda)/\mu(\Lambda)$.

Démonstration. Soit A_1, \dots, A_m une partition de Λ . Soit (k_1, \dots, k_m) tels que $k_1 + \dots + k_m = n$. On a :

$$\begin{aligned}
 & \mathbf{P}(N(A_i) = k_i, i = 1, \dots, m \mid N(\Lambda) = n) \\
 &= \frac{\mathbf{P}(N(A_i) = k_i, i = 1, \dots, m, N(\Lambda) = n)}{\mathbf{P}(N(\Lambda) = n)} \\
 &= \frac{\mathbf{P}(N(A_i) = k_i, i = 1, \dots, m)}{\mathbf{P}(N(\Lambda) = n)} \\
 &= \frac{\exp(-\sum_{i=1}^m \mu(A_i)) \prod_{i=1}^m \frac{\mu(A_i)^{k_i}}{k_i!}}{\exp(-\mu(\Lambda)) \frac{\mu(\Lambda)^n}{n!}} \quad (10.2) \\
 &= \frac{n!}{k_1! \dots k_m!} \prod_{i=1}^m \left(\frac{\mu(A_i)}{\mu(\Lambda)} \right)^{k_i}.
 \end{aligned}$$

On reconnaît alors la formule (10.1) pour μ_Λ , d'où le résultat. \square

On déduit de ce résultat un moyen de simulation d'un processus de Poisson dans tout ensemble Λ tel que $\mu(\Lambda)$ soit fini.

Algorithme 10.1 . Simulation d'une réalisation d'un P.P. (μ) sur un ensemble Λ

Données : μ, Λ

Résultat : n = réalisation d'une variable aléatoire de loi de Poisson $(\mu(\Lambda))$

pour $i = 1$ à n **faire**

 | X_i = tirage d'un point selon $\mu/\mu(\Lambda)$

fin

Sorties : $n, X_i, i = 1, \dots, n$

Par discrétisation, on peut imaginer des mesures μ quelconques, comme par exemple celles correspondant à des densités de population.

EXEMPLE 10.3 (File M/M/ ∞).— La file M/M/ ∞ est la file à arrivées poissonniennes, temps de service indépendants et identiquement distribués de loi exponentielle et une infinité de serveurs (donc sans buffer). C'est un objet d'abord théorique particulièrement simple à analyser mais aussi un modèle auquel on peut comparer d'autres situations. De par l'indépendance des interarrivées et des temps de service, d'après la deuxième caractérisation des processus de Poisson, le processus :

$$N = \sum_{n \geq 1} \delta_{(T_n, S_n)}$$

où T_n est l'instant de n^{e} arrivée et S_n le n^{e} temps de service, est un processus de Poisson d'intensité $d\mu(t, x) = \lambda dt \otimes \mu e^{-\mu x} dx$ dans $E = \mathbf{R}^+ \times \mathbf{R}^+$. Les clients en service à l'instant t sont ceux qui sont arrivés avant t et dont le service n'est pas terminé en t , c'est-à-dire que le n^{e} client est en service à l'instant t si seulement si :

$$T_n \leq t < T_n + S_n.$$

Soit \mathfrak{T}_t le domaine du plan défini par :

$$\mathfrak{T}_t = \{(x, s) \in (\mathbf{R}^+)^2, x \leq t < x + s\}.$$

Le nombre de clients à l'instant t est donc donné par :

$$X(t) = \int_{\mathfrak{T}_t} dN(x, s) = \sum_{n \geq 1} \mathbf{1}_{\mathfrak{T}_t}(T_n, S_n).$$

Dans la figure 10.2, les clients encore en service à l'instant t sont ceux qui correspondent à des points dans le trapèze ombré.

On en déduit que $X(t)$, le nombre de serveurs occupés à l'instant t suit une loi de Poisson de paramètre :

$$\int_0^t \left(\int_{t-s}^{\infty} \mu e^{-\mu x} dx \right) \lambda ds = \lambda \int_0^t e^{-\mu(t-s)} ds = \rho(1 - e^{-\mu t}),$$

où $\rho = \lambda/\mu$. Si le système n'est pas vide à l'instant 0, il faut ajouter à cette loi, la loi du nombre de clients à l'instant 0 encore en service à l'instant t . En particulier, si X_0 suit une loi de Poisson de paramètre ρ_0 , à l'instant t , le nombre de clients en service suit une loi de Poisson $\rho_0 e^{-\mu t}$ car chacun des clients a une probabilité $e^{-\mu t}$ d'être encore en service et le total est donc l'amincissement d'une variable aléatoire de loi de Poisson. En conclusion, $X(t)$ suit alors une loi de Poisson de paramètre $\rho + (\rho_0 - \rho)e^{-\mu t}$. Quelle que soit la valeur de ρ_0 , X admet comme probabilité stationnaire la loi de Poisson de paramètre ρ .

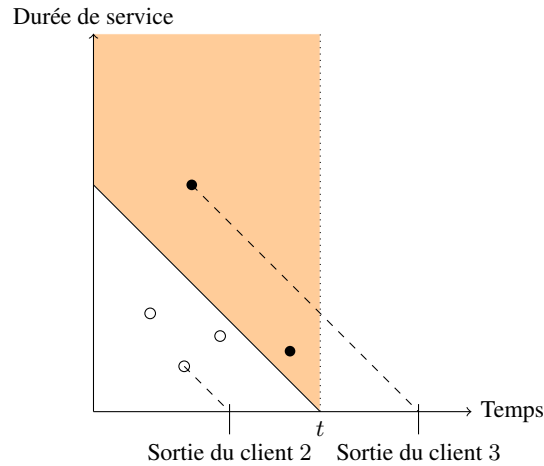


Figure 10.2 – La file $M/M/\infty$ comme un processus de Poisson dans $\mathbf{R}^+ \times \mathbf{R}^+$

REMARQUE.— Illustrons certaines différences entre un modèle poissonnien et un modèle hexagonal pour les réseaux sans fil. Dans un modèle hexagonal, le nombre moyen d'utilisateurs par unité de surface est l'inverse de l'aire d'un hexagone. Si le rayon des hexagones est R , cela donne une densité moyenne de $2/(3\sqrt{3})R^{-2}$. Dans un modèle poissonnien, le nombre moyen d'utilisateurs par unité de surface est λ . De la même manière que pour comparer deux files d'attente, la charge doit être identique, pour comparer deux systèmes spatiaux, le nombre moyen d'utilisateurs doit être identique. On doit donc choisir λ et R^2 tels que $\lambda R^2 = 2 \cdot 3^{-3/2}$.

L'un des paramètres essentiels est comme on l'a dit, la distance. Notamment, une distance courte entre ressources assure une meilleure couverture mais engendre des interférences. Dans un modèle hexagonal, si le rayon de la cellule est R , la distance entre plus proches voisins est $R\sqrt{3}$. Calculons cette quantité dans le cas d'un processus de Poisson d'intensité λdx sur \mathbf{R}^2 . Soit $x \in \mathbf{R}^2$ un point quelconque du plan, on définit le domaine $D_x(N)$ par :

$$D_x(N) = d(x, N) = \inf\{\|x - y\|, y \in N\}.$$

On a de manière évidente :

$$\mathbf{P}(d(x, N) \geq \tau) = \mathbf{P}(N(B(x, \tau)) = 0) = \exp(-\lambda\pi\tau^2).$$

Par conséquent :

$$\begin{aligned}\mathbf{E}[D_x] &= \int_0^{+\infty} \mathbf{P}(D_x \geq \tau) \, d\tau \\ &= \int_0^{+\infty} \exp(-\lambda\pi\tau^2) \, d\tau \\ &= \int_0^{+\infty} \exp(-u^2/2) \frac{1}{\sqrt{2\lambda\pi}} \, du = \frac{1}{2\sqrt{\lambda}},\end{aligned}$$

car on reconnaît la demi-intégrale de la densité gaussienne de variance λ . En utilisant le calcul de Palm, on montre que ce résultat reste vrai pour la distance en un point quelconque du processus et son plus proche voisin. En conclusion, si $\lambda R^2 = 2/3^{3/2}$, on obtient que la distance moyenne dans le modèle poissonnien est approximativement $0,8R$, soit bien moins que la distance dans le modèle hexagonal. Pour les interférences, qui sont inversement proportionnelles à la distance, le modèle poissonnien est donc plus pessimiste que le modèle hexagonal.

EXEMPLE 10.4 (Interférence moyenne).— En un point x du plan, l'interférence créée par les mobiles a pour expression :

$$I(x, N) = \sum_{y \in N} h(y)P(y)l(\|y - x\|),$$

où $P(x)$ est la puissance du signal émis par le mobile en y , l est une fonction de \mathbf{R}^+ dans \mathbf{R}^+ que l'on prend généralement de la forme :

$$l_0(r) = r^{-\gamma} \text{ ou } l_1(r) = \min(1, r^{-\gamma}). \quad (10.3)$$

La deuxième formulation donne des formules moins élégantes mais est plus réaliste (un signal ne vas pas être amplifié sous prétexte que le récepteur est très proche de l'émetteur) et évite les problèmes de divergence d'intégrale. Les variables aléatoires $(h(y), y \in N)$ sont généralement identiquement distribuées (de même loi qu'une variable aléatoire H), indépendantes entre elles et indépendantes de N . Elles représentent le facteur de perte induit par le *fading* (l'atténuation due aux mouvements locaux du récepteur) et le *shadowing* (l'atténuation du signal due aux obstacles entre l'émetteur et le récepteur). Généralement, le *fading* est modélisé par une variable aléatoire de loi exponentielle de paramètre 1. Le *shadowing* est représenté par une loi log-normale, c'est-à-dire l'exponentielle d'une variable gaussienne.

La formule de Campbell indique que :

$$\mathbf{E}_\mu [I(x)] = \mathbf{E}[H] \int P(y)l(\|y - x\|) \, d\mu(y).$$

Supposons alors que l'on se place sur le plan tout entier, que la puissance soit la même pour tous les mobiles et que μ soit proportionnelle à la mesure de Lebesgue :

$d\mu(x) = \lambda dx$. On observe immédiatement que la quantité précédente ne dépend pas de x , d'où :

$$\mathbf{E}_\mu [I(0)] = P\mathbf{E} [H] \int l(\|y\|)\lambda dy = \lambda\mathbf{E} [H] \int_0^\infty l(r)r dr.$$

Si l'on prend comme modèle de *path-loss* la fonction l_1 définie dans (10.3), pour une cellule de rayon $R > 1$, on obtient :

$$\mathbf{E}_\lambda [I(0)] = \mathbf{E} [H] \lambda \left(\pi + \frac{\pi}{\gamma - 2} (1 - R^{2-\gamma}) \right).$$

Pour R grand, cette quantité s'approche de $\mathbf{E} [H] \pi \lambda \gamma$ pour $\gamma > 2$. Il est normal que l'interférence augmente avec le nombre moyen d'utilisateurs donc d'interfereurs et avec la portée des signaux : plus γ augmente, moindre est l'atténuation due au *path-loss*.

EXEMPLE 10.5 (Loi de l'interférence).– Supposons maintenant que les variables aléatoires $(h(x), x \in \mathbf{R}^2)$ soient indépendantes de même loi. Pour s réel positif, on a :

$$\begin{aligned} \mathbf{E} \left[\exp\left(-s \int h(x)l(\|x\|) dN(x)\right) \right] \\ &= \mathbf{E} \left[\mathbf{E} \left[\exp\left(-s \int h(x)l(\|x\|) dN(x)\right) \mid N(\Lambda) \right] \right] \\ &= \mathbf{E} \left[\prod_{x \in N} \int \exp(-sl(\|x\|)y) d\mathbf{P}_H(y) \right], \end{aligned}$$

puisque conditionnellement au nombre de points dans Λ , les atomes sont indépendants les uns des autres. En notant \mathcal{L}_H la transformée de Laplace de H , on obtient :

$$\begin{aligned} \mathbf{E} \left[\exp\left(-s \int h(x)l(\|x\|) dN(x)\right) \right] &= \mathbf{E} \left[\prod_{x \in N} \mathcal{L}_H(sl(\|x\|)) \right] \\ &= \mathbf{E} \left[\exp\left(\int_\Lambda \ln \mathcal{L}_H(sl(\|x\|)) dN(x)\right) \right] \\ &= \exp\left(-\int_\Lambda 1 - e^{\ln \mathcal{L}_H(sl(\|x\|))} d\mu(x)\right) \\ &= \exp\left(\int_\Lambda (\ln \mathcal{L}_H(sl(\|x\|)) - 1) d\mu(x)\right). \end{aligned}$$

Pour le *fading* de Rayleigh, H est la loi exponentielle de paramètre 1. Si l'on suppose que $\mu = \lambda dx$ et que le *path-loss* est donné par l_0 , tous les calculs sont faisables et l'on obtient la formule suivante :

$$\mathcal{L}_{I(0)}(s) = \exp\left(-\pi \lambda s^\delta \frac{\pi \delta}{\sin(\pi \delta)}\right),$$

où $\delta = 2/\gamma$. On sait alors que cela correspond à une loi stable d'exposant caractéristique δ , voir [SAM 94].

La plupart des propriétés du processus de Poisson réel se transfèrent au processus de Poisson spatial.

THÉORÈME 10.4 (Superposition).— Soit N^1 et N^2 deux processus de Poisson indépendants d'intensité respective μ^1 et μ^2 , leur superposition N définie par :

$$\int f \, dN = \int f \, dN^1 + \int f \, dN^2$$

est un processus de Poisson d'intensité $\mu^1 + \mu^2$.

DÉFINITION 10.3.— Soit N un processus de Poisson d'intensité μ et $p : E \rightarrow [0, 1]$. Le processus de Poisson aminci de paramètre (μ, p) est le processus où un atome du processus de Poisson N en x est conservé avec probabilité $p(x)$.

THÉORÈME 10.5 (Amincissement).— Un processus de Poisson aminci de paramètres (μ, p) est un processus de Poisson d'intensité μ_p définie par :

$$\mu_p(A) = \int_A p(x) \, d\mu(x).$$

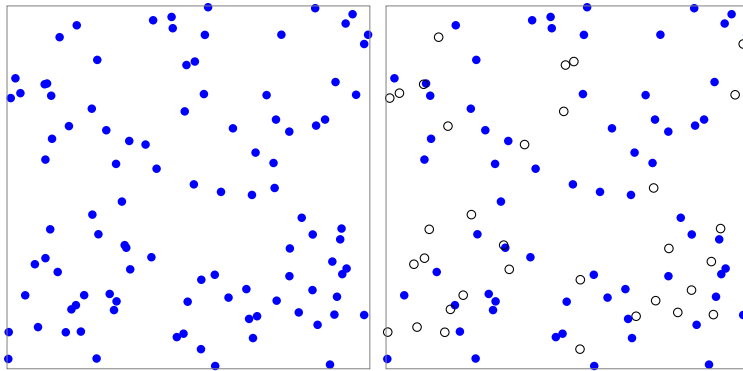


Tableau 10.2 – Une réalisation et l'un de ses amincissements avec $p = 2/3$; en plein, les points qui sont gardés

REMARQUE.— Ce résultat est intéressant dans le cadre de la modélisation. Si les utilisateurs sont représentés par les points d'un processus de Poisson ponctuel, seuls

ceux qui émettent à un instant donné concerne l'opérateur. On peut supposer que chaque utilisateur a une probabilité p d'émettre à un instant quelconque et ce indépendamment des autres mobiles. Le théorème 10.5 nous indique alors que les utilisateurs actifs sont dispersés dans le plan selon un processus de Poisson d'intensité λp .

Le théorème 10.5 est un cas particulier du théorème du déplacement.

DÉFINITION 10.4.— Soit $(\Omega', \mathcal{A}', \mathbf{P}')$ un espace probabilisé et (F, \mathcal{F}) un espace polonais. Un déplacement est une application mesurable Θ de $\Omega' \times E \rightarrow F$ telle que les variables aléatoires $(\Theta(\omega', x), x \in E)$ soient indépendantes dans leur ensemble. Pour $A \in \mathcal{F}$, on note :

$$\theta(x, A) = \mathbf{P}'(\omega' : \Theta(\omega', x) \in A).$$

Ainsi, $\theta(x, A)$ représente la probabilité que le point x soit déplacé dans A . Plus mathématiquement, si l'on note $\Theta(\omega', \cdot) * \mu$ la mesure image de μ par l'application $\Theta(\omega', \cdot)$, on a :

$$\begin{aligned} \mathbf{E}_{\mathbf{P}'} [\Theta^* \mu(A)] &= \mathbf{E}_{\mathbf{P}'} \left[\int \mathbf{1}_{\{\Theta(\omega', x) \in A\}} d\mu(x) \right] \\ &= \int \mathbf{P}'(\Theta(\omega', x) \in A) d\mu(x) = \int \theta(x, A) d\mu(x). \end{aligned}$$

Cela revient à dire que :

$$\mathbf{E}_{\mathbf{P}'} \left[\int \mathbf{1}_A d\Theta^* \mu \right] = \int \int_A \theta(x, dy) d\mu(x),$$

donc pour une fonction f positive, on obtient :

$$\mathbf{E}_{\mathbf{P}'} \left[\int f d\Theta^* \mu \right] = \int \int f(y) \theta(x, dy) d\mu(x). \quad (10.4)$$

DÉFINITION 10.5.— Un déplacement est dit conservatif lorsque pour tout $\Lambda \subset E$ compact, la propriété suivante est satisfaite :

$$\mathbf{E}_{\mathbf{P}'} [\Theta^* \mu(\Lambda)] = \int_{\Lambda} \int_F \theta(x, dy) d\mu(x) = \mu(\Lambda).$$

Cela signifie qu'en moyenne, la « masse totale » du processus ponctuel est conservée.

Soit Θ un déplacement tel que $\int_{\Lambda} \int_F e^{-f(y)} \theta(x, dy) d\mu(x) = \mu(\Lambda)$ et N un processus ponctuel, le processus ponctuel « déplacé » N^{Θ} est défini par :

$$N^{\Theta}(\omega') = \sum_{x \in N} \delta_{\Theta(\omega', x)}.$$

THÉORÈME 10.6 (Déplacement).— Soit N un processus de Poisson d'intensité μ sur E et Θ un déplacement conservatif de E dans F . Le processus N^Θ est un processus de Poisson d'intensité μ^Θ définie par :

$$\mu^\Theta(A) = \int_E \theta(x, A) \, d\mu(x).$$

Démonstration. On suppose dans un premier temps que f est à support compact, noté Λ . On sait que conditionnellement à $N(\Lambda)$, les atomes de N sont indépendants distribués selon $\mu/\mu(\Lambda)$. Par conséquent, on peut écrire :

$$\mathbf{E} \left[F \exp\left(-\int_\Lambda dN\right) \right] = \sum_{n=0}^{\infty} \frac{e^{-\mu(\Lambda)} \mu(\Lambda)^n}{n!} \int_{E^n} \prod_{j=1}^n e^{-f(x_j)} \frac{d\mu(x_j)}{\mu(\Lambda)}.$$

D'après la construction de N^Θ , l'aléa du déplacement est indépendant de N , on a :

$$\begin{aligned} \mathbf{E} \left[\exp\left(-\int f \, dN^\Theta\right) \right] &= \mathbf{E}_{\mathbf{P}' } \left[\sum_{n=0}^{\infty} \frac{e^{-\mu(\Lambda)}}{n!} \int_{E^n} \prod_{j=1}^n e^{-f(\Theta(\omega', x_j))} \, d\mu(x_j) \right] \\ &= \sum_{n=0}^{\infty} \frac{e^{-\mu(\Lambda)}}{n!} \mathbf{E}_{\mathbf{P}' } \left[\int_{E^n} \prod_{j=1}^n e^{-f(\Theta(\omega', x_j))} \, d\mu(x_j) \right]. \end{aligned}$$

Par définition d'un déplacement, les variables aléatoires $(\Theta(\omega', x_j), j = 1, \dots, n)$ sont indépendantes. En utilisant (10.4), on obtient :

$$\begin{aligned} \mathbf{E} \left[\exp\left(-\int f \, dN^\Theta\right) \right] &= \sum_{n=0}^{\infty} \frac{e^{-\mu(\Lambda)}}{n!} \left(\mathbf{E}_{\mathbf{P}' } \left[\int_E e^{-f(\Theta(\omega', x))} \, d\mu(x) \right] \right)^n \\ &= \sum_{n=0}^{\infty} \frac{e^{-\mu(\Lambda)}}{n!} \left(\int e^{-f} \, d\mu^\Theta \right)^n \\ &= \exp \left(-\mu(\Lambda) + \int_\Lambda \int_F e^{-f(y)} \theta(x, dy) \, d\mu(x) \right). \end{aligned}$$

Comme Θ est conservatif, on obtient :

$$\mathbf{E} \left[\exp\left(-\int f \, dN^\Theta\right) \right] = \exp \left(-\int_F (1 - e^{-f(y)}) \int_\Lambda \theta(x, dy) \, d\mu(x) \right)$$

donc N^Θ est bien un processus de Poisson d'intensité μ^Θ .

On obtient le cas général pour f à support quelconque, par troncature (on applique le résultat précédent à $f_\Lambda = f \mathbf{1}_\Lambda$) et passage à la limite (en considérant une suite croissante de compacts $(\Lambda_n, n \geq 1)$ telle que $\cup_n \Lambda_n = E$. Notons que l'existence d'une telle suite est assurée par le caractère polonais de E . \square

Preuve du théorème 10.5. On considère $F = E \cup \Delta$ où Δ est un point externe à E . Avec probabilité $p(x)$, l'atome x reste en x , avec la probabilité complémentaire, il est envoyé sur Δ . La restriction à E du processus ainsi obtenu est bien l'amincissement du processus initial. Ce déplacement est bien conservatif puisque l'on garde le même nombre d'atomes. Le théorème 10.5 est alors une conséquence directe du théorème 10.6. \square

En appliquant le théorème 10.6 à la fonction ($x \in \mathbf{R}^d \mapsto rx$) où $r \in \mathbf{R}^+$, on obtient un résultat dit de *scaling* très utile dans de nombreuses applications.

COROLLAIRE 10.7.— Soit N un processus de Poisson d'intensité μ sur \mathbf{R}^d . Soit $r > 0$, on appelle $N^{(r)}$ le processus dilaté défini par :

$$N^{(r)} = \sum_{x \in N} \delta_{rx}.$$

Le processus $N^{(r)}$ est un processus de Poisson d'intensité $\mu^{(r)}$ où : $\mu^{(r)}(A) = \mu(A/r)$ pour tout $A \in \mathfrak{B}(E)$.

COROLLAIRE 10.8.— Soit N un processus de Poisson d'intensité λdx sur \mathbf{R}^d . Le processus des modules est indépendant du processus des arguments. Le premier est un processus de Poisson d'intensité $2\lambda\pi r dr$, le deuxième un processus de Poisson d'intensité $(2\pi)^{-1} \mathbf{1}_{[0, 2\pi]}(\theta) d\theta$.

Démonstration. Le théorème 10.6 implique que le processus :

$$\hat{N} = \sum_{x \in N} \delta_{\|x\|, \text{Arg}(x)}$$

est un processus de Poisson d'intensité $\lambda r \mathbf{1}_{[0, 2\pi]}(\theta) dr d\theta$. D'où le résultat. \square

EXEMPLE 10.6 (Protocole OFDMA).— Illustrons ces résultats sur le cas particulier du protocole OFDMA. Dans ce protocole, le temps et la bande de fréquence sont découpés en morceaux appelées sous-porteuses. A une communication, sont attribuées une ou plusieurs sous-porteuses. En pratique, l'allocation se fait pour une durée de quelques slots d'où les affectations en rectangle dans la figure 10.3.

Pour une communication radio, il est impératif que le rapport signal sur bruit plus interférence (SINR) soit suffisamment grand pour que la communication puisse être établie. Notons P_e la puissance émise, A le facteur d'atténuation due au *fading* et au *shadowing*, γ le coefficient de *path-loss* et η la valeur minimale du SINR tolérable pour établir une communication. La formule de capacité de Shannon stipule que pour un

utilisateur souhaitant émettre à un débit C , le nombre de sous-porteuses nécessaires est donné par :

$$Q(x) = \left\lceil \frac{C}{W \log_2 \left(1 + \frac{A P_e}{\|x\|^\gamma} \right)} \right\rceil \mathbf{1}_{\{A P_e \|x\|^{-\gamma} > \eta\}},$$

où W est la largeur de bande de chaque sous-porteuse, P_e est la puissance du signal et le facteur A dépend des effets de *fading* et *shadowing*. Le rapport SINR minimal varie entre 0,001 (excellent canal) et 0,1 (conditions plus mauvaises). On supposera que l'affectation se fait à chaque slot. Il s'agit de déterminer la probabilité que le nombre de sous-porteuses disponibles soit supérieur au nombre de sous-porteuses demandées.

On suppose que le point d'accès est situé à l'origine du plan et que les utilisateurs sont spatialement répartis selon un processus de Poisson d'intensité $\lambda \, dx$. Un modèle simple pour représenter les utilisateurs actifs est d'assigner à chacun une probabilité d'activité p dépendant éventuellement de sa position et de considérer que l'activité ou l'oisiveté d'un mobile est indépendante de celle des autres. En vertu du théorème 10.2, le processus ponctuel des utilisateurs actifs est encore un processus de Poisson d'intensité $\lambda \int p(x) \, dx$. Le nombre total de sous-porteuses demandées est alors :

$$M = \int_{\Lambda} Q(x) \, dN(x),$$

où Λ est le domaine qui représente la cellule couverte par le point d'accès en $(0, 0)$. En supposant que A et P_e sont déterministes et indépendants de x , en vertu de la formule de Campbell :

$$\mathbf{E}[M] = \lambda \int_{\Lambda} Q(x) p(x) \, dx.$$

Si A et P_e dépendent de x et d'un aléa indépendant de N , on a la même expression quitte à prendre l'espérance du terme de droite.

Il est souvent impossible de calculer explicitement les lois des variables aléatoires construites à partir d'un processus ponctuel comme celle de M . L'analyse stochastique nous donne des outils pour obtenir tout de même des informations sur ces distributions (on se reportera à la section 10.4 pour les notations et les preuves).

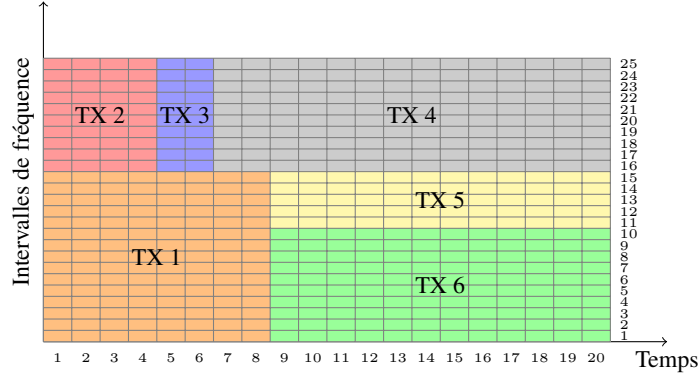


Figure 10.3 – Découpage et allocation dans l'espace temps-fréquence pour le protocole OFDMA

THÉORÈME 10.9.– Soit N un processus de Poisson d'intensité μ sur E . Soit $F \in \text{Dom } D$ telle que $\mathbf{E}[F] = 0$. On a alors l'inégalité suivante :

$$\begin{aligned} d_{\text{VT}}(\mathbf{P}_F, \mathcal{N}(0, 1)) \leq \mathbf{E} \left[\left| 1 - \int_E D_x F D_x L^{-1} F \, d\mu(x) \right| \right] \\ + \int_E \mathbf{E} [|D_x F|^2 |D_x L^{-1} F|] \, d\mu(x), \end{aligned}$$

où d_{VT} représente la distance en variation totale entre deux mesures de probabilité $\mathcal{N}(0, 1)$ la mesure gaussienne centrée réduite sur \mathbf{R} .

COROLLAIRE 10.10.– Soit N_1, \dots, N_K des processus de Poisson spatiaux indépendants d'intensité respective $\lambda_i \, dx$ sur un domaine borné Λ . Soit F une fonctionnelle de la forme :

$$F = \sum_{i=1}^K \sum_{n=1}^{N^i(\Lambda)} Y_n^i$$

où $(Y_n^i, i, n \in \mathbf{N})$ sont des variables indépendantes dont la loi ne dépend que de l'indice en exposant, c'est-à-dire que Y_n^i et Y_m^i ont la même loi mais ce n'est pas nécessairement le cas pour Y_n^i et Y_n^j . Soit :

$$c^2 = \sum_{i=1}^K \lambda_i(\Lambda) \int_{\mathbf{R}} m^2 \, d\mathbf{P}_{Y_1^i}(m),$$

où $\lambda_i(\Lambda) = \int_{\Lambda} \lambda_i \, dx$ est l'aire de Λ multipliée par λ_i . On obtient :

$$d_{\text{VT}}(\text{loi}((F - E[F])/c), \mathcal{N}(0, 1)) \leq \frac{1}{c^3} \left(\sum_{i=1}^K \lambda_i(\Lambda) \int_{\mathbf{R}} m^3 \, d\mathbf{P}_{Y_1^i}(m) \right).$$

En particulier :

$$\left| \mathbf{P}\left(\frac{F - E[F]}{c} \geq x\right) - \int_x^\infty e^{u^2/2} \frac{du}{\sqrt{2\pi}} \right| \leq \frac{1}{c^3} \left(\sum_{i=1}^K \lambda_i \int_\Lambda dx \int_{\mathbf{R}} m^3 d\mathbf{P}_{Y_1^i}(m) \right).$$

Démonstration. Comme les processus de Poisson individuels sont indépendants les uns des autres, le processus :

$$\sum_{i=1}^K \sum_{(x,m) \in N^i} \delta_{x,m}$$

est un processus de Poisson sur $\mathbf{R}^+ \times \mathbf{R}$ d'intensité :

$$\sum_{i=1}^K \lambda_i dx \otimes d\mathbf{P}_{Y_1^i}(m).$$

Dans ce cas, l'opérateur gradient est défini par :

$$D_{x,m}F(\omega) = F(\omega + \delta_{x,m}) - F(\omega)$$

La fonctionnelle F se réécrit :

$$F = \int_\Lambda \int_{\mathbf{R}} m d\left(\sum_{i=1}^K \sum_{(x,m) \in N^i} \delta_{x,m} \right)$$

où Λ est le domaine du plan sur lequel on travaille. Par conséquent, on a $D_{x,m}F = m$. Le résultat découle du théorème 10.9. \square

EXEMPLE (Suite de l'exemple 10.6).— On suppose dans un premier temps que le *fading* est constant. C'est une hypothèse peu réaliste mais qui permet d'illustrer la complexité du problème. Dans ce cas, pour chaque mobile, Q ne dépend que de sa position. On a donc un ensemble de couronnes concentriques autour de la station de base dans chacune desquelles le nombre de sous-porteuses demandées est le même. En effet, on a :

$$Q(x) = k \iff k-1 < \frac{C}{W \log_2 \left(1 + \frac{AP_e}{\|x\|^\gamma} \right)} \leq k \iff R_{k-1} < x < R_k,$$

où $R_0 = 0$, $R_k = \left(\frac{AP_e}{(2^{C_0/(kW)} - 1)I} \right)^{1/\gamma}$. Le rayon maximum R_M est défini par la relation :

$$\frac{AP_e}{\|R_M\|^\gamma} = \eta, \text{ soit } R_M = \left(\frac{AP_e}{\eta} \right)^{1/\gamma}.$$

De même, le nombre maximum de sous-porteuses demandées est $N_M = Q(R_M)$. Il est instructif d'étudier les variations du nombre moyen de sous-porteuses demandées en fonction des variations de certains paramètres. Fixons $C = 200$ kb/s, $W = 250$ kHz, $A P_e = 20\,000$ et $\lambda = 0,01$ mobile par mètre carré.

D'après la deuxième caractérisation du processus de Poisson, le nombre d'utilisateurs dans la couronne k noté N_k suit une loi de Poisson de paramètre $\lambda\pi(R_k^2 - R_{k-1}^2)$. Par ailleurs, les variables aléatoires ($N_k, k \geq 1$) sont indépendantes puisque les couronnes sont disjointes. Le nombre de sous-porteuses demandées est donc à chaque slot $N = \sum_k k N_k$. Par conséquent :

$$\mathbf{E}[N] = \lambda\pi \sum_k k(R_k^2 - R_{k-1}^2) \text{ et } \text{Var}(N) = \lambda\pi \sum_k k^2(R_k^2 - R_{k-1}^2).$$

Une telle variable aléatoire suit une loi dite de Poisson composé. On sait facilement en calculer la transformée de Laplace mais cela ne donne pas d'expression explicite de la probabilité que N soit égale à k . On peut aussi en calculer numériquement la loi par convolution des lois de chacune des variables aléatoires de Poisson « dilatées » qui la composent. Néanmoins, la taille des calculs est rapidement prohibitive, on a donc recours à des approximations. Pour une variable aléatoire de loi de Poisson de paramètre λ , on sait que quand λ tend vers l'infini, la loi de cette variable convenablement renormalisée tend vers une loi gaussienne. Les inconvénients de cette démarche sont de deux ordres. D'une part, on ne sait pas ce que signifie λ « grand », c'est-à-dire à partir de quelles valeurs de λ l'approximation est « valable ». D'autre part, l'approximation est, suivant les valeurs, de λ tantôt pessimiste, tantôt optimiste ainsi que le montre la figure 10.4, obtenue pour $\gamma = 3,5$ et $\lambda = 0,02$.

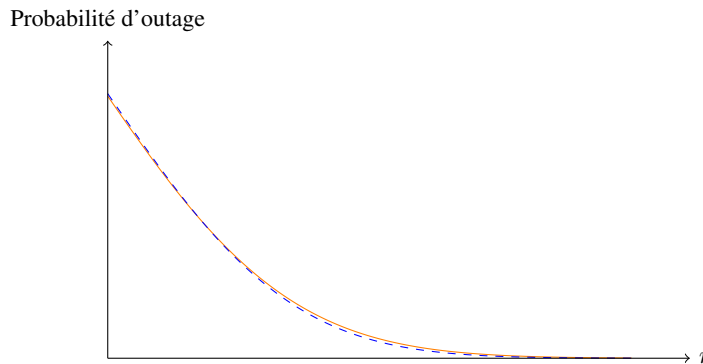


Figure 10.4 – Probabilité d'outage en fonction de r . En trait plein, la valeur exacte, en pointillés, l'approximation gaussienne

Pour répondre à la première difficulté, on peut utiliser les résultats précédents. En prenant $Y_n^k = k$ pour tout n , on remarque que le c^2 du théorème 10.10 est égal à $\text{Var}(N)$. On a donc :

$$\begin{aligned} d_{\text{VT}} \left(\text{loi} \left(\frac{N - \mathbf{E}[N]}{c}, \mathcal{N}(0, 1) \right) \right) &\leq \frac{\lambda \pi \sum_k k^3 (R_k^2 - R_{k-1}^2)}{(\lambda \pi \sum_k k^2 (R_k^2 - R_{k-1}^2))^{3/2}} \\ &= \frac{1}{\sqrt{\lambda}} \frac{\sum_k k^3 (R_k^2 - R_{k-1}^2)}{\sqrt{\pi} (\sum_k k^2 (R_k^2 - R_{k-1}^2))^{3/2}}. \end{aligned}$$

En d’autres termes, la convergence vers la loi gaussienne se fait en $1/\sqrt{\lambda}$ comme cela est attendu. Cette borne toute théorique est en fait très pessimiste comme le prouve la figure 10.4.

Néanmoins, l’avantage de ce résultat est qu’il fournit une borne explicite. Le tableau 10.3 montre une grande sensibilité du nombre moyen de sous-porteuses demandées aux variations de γ .

η	γ			
	2,5	3	3,5	4
0,1	$2 \cdot 10^3$	334	98	38
0,01	$9 \cdot 10^5$	10^5	$3 \cdot 10^3$	56
0,001	$5 \cdot 10^6$	$5 \cdot 10^5$	10^5	555

Tableau 10.3 – Nombre moyen de sous-porteuses demandées en fonction de η et γ

Les mêmes calculs montreraient une aussi grande sensibilité aux variations de λ , le nombre moyen d’usagers actifs par unité de surface. Or ces deux paramètres sont particulièrement délicats à estimer, notamment le coefficient γ qui dépend grandement de l’environnement (rural, urbain, semi-urbain, pavillonnaire, gratte-ciels, etc.). Il est donc d’une importance cruciale d’avoir des dimensionnements qui soient robustes aux variations de ces paramètres.

THÉORÈME 10.11. – Soit N un processus de Poisson d’intensité μ sur E et Λ un compact de E . Soit $F : \mathfrak{N}_\Lambda \rightarrow \mathbf{R}$ telle que :

$$D_x F(N_\Lambda) \leq \beta, \quad (\mu \otimes \mathbf{P}) - \text{p.p. et } \int_E |D_x F(N_\Lambda)|^2 d\mu(x) \leq \alpha^2, \quad \mathbf{P} - \text{p.p.}$$

Pour tout $r > 0$, on a l’inégalité suivante :

$$\mathbf{P}(F(N_\Lambda) - \mathbf{E}[F(N_\Lambda)] > r) \leq \exp \left(-\frac{r}{2\beta} \ln \left(1 + \frac{r\beta}{\alpha^2} \right) \right).$$

EXEMPLE (Suite de l'exemple 10.6).– Dans le cas qui nous intéresse, le nombre maximal de sous-porteuses demandées est borné et ne dépend que des caractéristiques du réseau (puissance émise, largeur de bande des sous-porteuses, etc.). Par conséquent, $D_x N \leq N_M$ et si E est la cellule, on peut choisir $\alpha^2 = N_M^2 \mu(E) = \lambda \pi R_M^2 N_M^2$ dans le théorème 10.11. On obtient donc :

$$\mathbf{P}(N - \mathbf{E}[N] > r) \leq \exp\left(-\frac{r}{2N_M^2} \log\left(1 + \frac{r}{\lambda \pi R_M^2}\right)\right).$$

Par rapport à l'approximation gaussienne, cette formule ne nécessite pas de calculer la variance de N , seule la moyenne suffit.

On constate dans le tableau 10.4, que le dimensionnement par l'approximation gaussienne est optimiste, ce qui est absolument rédhibitoire. Le surdimensionnement induit par l'inégalité de concentration est raisonnable dans la plupart des cas. Lorsqu'il est franchement grand, cela peut être vu comme une garantie contre les imprécisions dans les mesures de λ et γ . On pourrait tout à fait envisager des classes d'utilisateurs avec différents débits et intégrer au modèle des considérations de *fading* et de *shadowing*. Les calculs exacts deviennent impossibles tandis que le calcul de la borne de concentration reste aussi aisé.

γ	Exact	Gaussienne	Concentration	Erreur
2,5	2060	2056	2357	0,14
2,6	1420	1417	1682	0,18
2,7	1010	1006	1243	0,23
2,8	738	735	948	0,29
2,9	553	549	745	0,35
3	423	420	600	0,42
3,5	146	142	276	0,9
4	69	66	177	1,57

Tableau 10.4 – Dimensionnement par les trois méthodes pour différentes valeurs de γ (pour $\lambda = 0,01$); la dernière colonne contient le surdimensionnement relatif entre celui obtenu par l'inégalité de concentration et le calcul exact

10.4. Analyse stochastique

Abordons maintenant la preuve des théorèmes 10.9 et 10.11.

LEMME 10.12 (Théorème de Girsanov).– Soit N et N' deux processus de Poisson ponctuels, d'intensité respectives μ et μ' . Supposons que $\mu' \ll \mu$ et notons $p = d\mu'/d\mu$. Soit Λ un compact de E . Si, de plus que $p \in L^1(\mu_\Lambda)$, alors pour toute

fonction F bornée, on a :

$$\mathbf{E} [F(N'_\Lambda)] = \mathbf{E} \left[F(N_\Lambda) \exp \left(\int \ln p \, dN_\Lambda + \int_\Lambda (1-p) \, d\mu \right) \right].$$

Démonstration. Vérifions cette identité pour les fonctionnelles exponentielles F de la forme $\exp(-\int f \, dN)$ avec f à support compact. En vertu de la définition 10.2, on a :

$$\begin{aligned} \mathbf{E} \left[\exp \left(- \int f \, dN_\Lambda \right) \exp \left(\int_\Lambda \ln p \, dN_\Lambda + \int (1-p) \, d\mu \right) \right] \\ = \mathbf{E} \left[\exp \left(- \int (f - \ln p) \, dN_\Lambda \right) \right] \exp \left(\int_\Lambda (1-p) \, d\mu \right). \end{aligned}$$

D'après la définition 10.2, il vient :

$$\begin{aligned} \mathbf{E} \left[\exp \left(- \int f \, dN_\Lambda \right) \exp \left(\int_\Lambda \ln p \, dN_\Lambda + \int (1-p) \, d\mu \right) \right] \\ = \exp \left(- \int (1 - \exp(-f + \ln p)) \, d\mu + \int_\Lambda (1-p) \, d\mu \right). \end{aligned}$$

Après simplification, on obtient :

$$\begin{aligned} \mathbf{E} \left[\exp \left(- \int f \, dN_\Lambda \right) \exp \left(\int_\Lambda \ln p \, dN_\Lambda + \int (1-p) \, d\mu \right) \right] \\ = \exp \left(- \int_\Lambda (1 - e^{-f}) p \, d\mu \right) = \mathbf{E} [F(N'_\Lambda)]. \end{aligned}$$

Il en résulte que les mesures sur \mathfrak{N}_E , $\mathbf{P}_{N'_\Lambda}$ et $Rd\mathbf{P}_{N_\Lambda}$ où :

$$R = \exp \left(\int \ln p \, dN_\Lambda + \int (1-p) \, d\mu \right)$$

ont la même transformée de Laplace. Par conséquent, elles sont égales et le résultat s'ensuit pour toute fonction F bornée. \square

Dans ce qui suit, pour une configuration η :

$$\eta + \delta_x = \begin{cases} \eta, & \text{si } x \in \eta, \\ \eta \cup \{x\}, & \text{si } x \notin \eta. \end{cases}$$

De même :

$$\eta - \delta_x = \begin{cases} \eta \setminus \{x\}, & \text{si } x \in \eta, \\ \eta, & \text{si } x \notin \eta. \end{cases}$$

Comme μ est supposée diffuse, $\mathbf{E}[N(\{x\})] = \mu(\{x\}) = 0$ donc pour x fixé, presque sûrement, η ne contient pas x .

DÉFINITION 10.6.— Soit N un processus de Poisson d'intensité μ et $F : \mathfrak{N}_E \rightarrow \mathbf{R}$ une fonction mesurable telle que $\mathbf{E}[F^2] < \infty$. On définit $\text{Dom } D$ comme l'ensemble des variables aléatoires de carré intégrable telles que :

$$\mathbf{E} \left[\int_E |F(N + \delta_x) - F(N)|^2 d\mu(x) \right] < \infty.$$

Pour $F \in \text{Dom } D$, on pose :

$$D_x F(N) = F(N + \delta_x) - F(N).$$

EXEMPLE.— Par exemple, pour f déterministe appartenant à $L^2(\mu)$, $F = \int f dN$ appartient à $\text{Dom } D$ et $D_x F = f(x)$ car :

$$F(N + \delta_x) = \sum_{y \in N \cup \{x\}} f(y) = \sum_{y \in N} f(y) + f(x).$$

De même, si $F = \max_{y \in N} f(y)$ alors :

$$D_x F(N) = \begin{cases} 0 & \text{si } f(x) \leq F(N) \\ f(x) - F & \text{si } f(x) > F(N). \end{cases}$$

Dans les deux cas, si f est bornée, DF l'est aussi.

L'une des formules essentielles pour le processus de Poisson est la suivante.

THÉORÈME 10.13.— Soit N un processus de Poisson d'intensité μ . Pour tout champ aléatoire $F : \mathfrak{N}_E \times E \rightarrow \mathbf{R}$ tel que :

$$\mathbf{E} \left[\int_E |F(N, x)| d\mu(x) \right] < \infty$$

alors :

$$\mathbf{E} \left[\int_E F(N, x) d\mu(x) \right] = \mathbf{E} \left[\int_E F(N \setminus x, x) dN(x) \right]. \quad (10.5)$$

Démonstration. D'après la première définition du processus de Poisson, pour f à support compact et Λ un compact de E , pour tout $t > 0$, on a :

$$\mathbf{E} \left[\exp\left(-\int (f + t \mathbf{1}_\Lambda) dN\right) \right] = \exp\left(-\int_E 1 - e^{-f(x) - t \mathbf{1}_\Lambda(x)} d\mu(x)\right).$$

D'après le théorème de dérivation sous le signe somme, on a d'une part :

$$\frac{d}{dt} \mathbf{E} \left[\exp(-\int (f + t \mathbf{1}_\Lambda) dN) \right] \Big|_{t=0} = -\mathbf{E} \left[e^{-\int f dN} \int \mathbf{1}_\Lambda dN \right],$$

et d'autre part :

$$\begin{aligned} \frac{d}{dt} \exp(-\int_E 1 - e^{-f(x)-t \mathbf{1}_\Lambda(x)} d\mu(x)) \Big|_{t=0} \\ = -\mathbf{E} \left[\int e^{-\int f dN + f(x)} \mathbf{1}_\Lambda(x) d\mu(x) \right]. \end{aligned}$$

Comme $\int f dN - f(x) = \int f d(N - \delta_x)$, la relation (10.5) est vraie pour les fonctions de la forme $\mathbf{1}_\Lambda e^{-\int f dN}$. On admettra que cela suffit à ce que le résultat soit vrai pour toutes les fonctions F telles que les deux membres soient bien définis. \square

Soit μ une mesure de Radon sur E polonais et Λ un compact de E . Introduisons maintenant, le processus de Poisson-Glauber, noté \mathfrak{N}^Λ dont la dynamique est la suivante :

- $\mathfrak{N}^\Lambda(0) = \eta \in \mathfrak{N}_\Lambda$;
- chaque atome de η a une durée de vie indépendante de celle des autres atomes, de loi exponentielle de paramètre 1 ;
- des atomes naissent à des instants suivant un processus de Poisson d'intensité $\mu(\Lambda)$; à son apparition, chaque atome est localisé indépendamment de tous les autres selon $\mu/\mu(\Lambda)$; il est aussi affecté de manière indépendante d'une durée de vie de loi exponentielle de paramètre 1.

Algorithme 10.2. Les k premières transitions d'une trajectoire du processus de Glauber associé à un processus de Poisson d'intensité μ , sur un domaine Λ , de condition initiale N

Données : μ, Λ, N, k

Résultat : $T_1, \dots, T_k =$ instants de sauts. $N_j = \mathfrak{N}^\Lambda(T_j)$, $j = 1, \dots, k$

$T_0 \leftarrow 0$;

$N_0 \leftarrow N$;

pour $n = 1$ à k **faire**

$L_n \leftarrow$ tirage d'une $\varepsilon(\mu(\Lambda) + N_{n-1}(\Lambda))$;

$T_n \leftarrow T_{n-1} + L_n$;

$r \leftarrow \mu(\Lambda) / (\mu(\Lambda) + N_{n-1}(\Lambda))$;

$U \leftarrow$ tirage d'une $U([0, 1])$;

si $U \leq r$ **alors**

$X \leftarrow$ tirage d'une variable aléatoire de loi $\mu/\mu(\Lambda)$;

$N_n \leftarrow N_{n-1} + \delta_X$;

sinon

$\kappa \leftarrow$ tirage d'une $U(\{1, \dots, N_{n-1}(\Lambda)\})$;

$N_n \leftarrow N_{n-1} - \delta_{x_\kappa}$;

fin

fin

retourner T_1, \dots, T_k

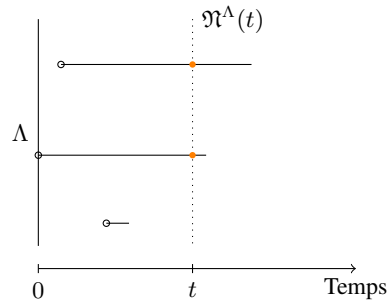


Figure 10.5 – Réalisation d'une trajectoire de \mathfrak{N}^Λ

A chaque instant, $\mathfrak{N}^\Lambda(t)$ est une configuration de E (voir figure 10.5). On remarque d'abord que le nombre total d'atomes de $\mathfrak{N}^\Lambda(t)$ suit exactement la même dynamique que le nombre de serveurs occupés dans une file $M/M/\infty$ de paramètre $\mu(\Lambda)$ et 1.

THÉORÈME 10.14.— Supposons que $\mathfrak{N}^\Lambda(0)$ soit un processus de Poisson ponctuel d'intensité ν . A chaque instant, la loi de $\mathfrak{N}^\Lambda(t)$ est celle d'un processus de Poisson d'intensité $e^{-t}\nu_\Lambda + (1 - e^{-t})\mu_\Lambda$ où ν_Λ est la restriction de ν à Λ . En particulier, si $\nu_\Lambda = \mu_\Lambda$, la loi de $\mathfrak{N}^\Lambda(t)$ ne dépend pas de t et est égale à μ_Λ . On note $\mathbf{E}_{\mu_\Lambda}[X]$ l'espérance d'une variable aléatoire X calculée sous cette condition.

Démonstration. Pour deux parties disjointes A et B de Λ , par construction les processus \mathfrak{G}_A et \mathfrak{G}_B sont indépendants et suivent la même dynamique qu'une file $M/M/\infty$ de paramètres respectifs $(\mu(A), 1)$ et $(\mu(B), 1)$. Le résultat découle des propriétés de la file $M/M/\infty$ établies dans l'exemple 10.3. \square

Comme tous les temps de séjour sont de loi exponentielle, \mathfrak{N}^Λ est un processus de Markov à valeurs dans \mathfrak{N}_E . Loin de nous l'idée de développer la théorie générale des processus de Markov dans les espaces de mesures. On peut toutefois étudier son générateur infinitésimal et son semi-groupe.

THÉORÈME 10.15.— Soit Λ un compact de E . Le générateur infinitésimal de \mathfrak{N}^Λ est donné par :

$$-\mathfrak{L}_\Lambda F(N) = \int_\Lambda (F(N + \delta_x) - F(N)) d\mu(x) + \int (F(N - \delta_x) - F(N)) dN(x), \quad (10.6)$$

pour F bornée de \mathfrak{N}_Λ dans \mathbf{R} .

Démonstration. On raisonne de la même manière que pour les processus de Markov. A un instant t , il peut y avoir soit un départ, soit une naissance. Lors d'un départ, on choisit l'atome tué uniformément parmi les atomes existants. Le taux de mort est donc $\eta(\Lambda)$ et chaque atome a une probabilité $\eta(\Lambda)^{-1}$ d'être tué. Par conséquent, la transition η vers $\eta - \delta_x$ a lieu au taux 1 pour tout $x \in \eta$. Le taux de naissance est $\mu(\Lambda)$ et la position du nouvel atome est réparti selon la mesure $\mu_\Lambda/\mu(\Lambda)$ donc la transition η vers $\eta + \delta_x$ a lieu au taux $d\mu_\Lambda(x)$ pour chaque $x \in \Lambda$. On en déduit (10.6). \square

THÉORÈME 10.16.— Le semi-groupe \mathfrak{P}^Λ est ergodique. De plus, \mathfrak{L}_Λ est inversible de L_0^2 dans L_0^2 où L_0^2 est le sous-espace de L^2 des variables aléatoires d'espérance nulle et l'on a :

$$\mathfrak{L}_\Lambda^{-1}F = \int_0^\infty \mathfrak{P}_t^\Lambda F dt. \quad (10.7)$$

Pour tout $x \in E$ et tout $t > 0$, on a :

$$D_x \mathfrak{P}_t^\Lambda F = e^{-t} \mathfrak{P}_t^\Lambda D_x F, \quad (10.8)$$

ainsi que :

$$\mathbf{E}_{\mu_\Lambda} \left[\int_\Lambda |D_x(\mathfrak{L}_\Lambda^{-1}F(N))|^2 d\mu(x) \right] \leq \mathbf{E}_{\mu_\Lambda} \left[\int_\Lambda |D_x F(N)|^2 d\mu(x) \right]. \quad (10.9)$$

Démonstration. Notons (x_1, \dots, x_n) les atomes de $\mathfrak{N}^\Lambda(0)$ et (Y_1, \dots, Y_n) des variables aléatoires indépendantes de loi exponentielle de paramètre 1. On pose :

$$\mathfrak{N}^\Lambda(0)[t] = \sum_{i=1}^n \mathbf{1}_{\{Y_i \geq t\}} \delta_{x_i},$$

la mesure constituée des atomes de $\mathfrak{N}^\Lambda(0)$ survivants en t . La loi de $\mathfrak{N}^\Lambda(t)$ est celle de la superposition indépendante d'un processus de Poisson d'intensité $(1 - e^{-t})\mu_\Lambda$ et de $\mathfrak{N}^\Lambda(0)[t]$. D'après le lemme 10.12, pour tout $F \in L^1$, on sait que :

$$\mathbf{E}_{(1-e^{-t})\mu_\Lambda} [F(N_\Lambda)] = \mathbf{E}_{\mu_\Lambda} [F(N_\Lambda) \exp(\ln(1 - e^{-t})N(\Lambda) + e^{-t}\mu(\Lambda))].$$

Par conséquent, pour toute fonction F bornée et tout $\eta \in \mathfrak{N}_\Lambda$, on a l'identité suivante :

$$\begin{aligned} \mathfrak{P}_t^\Lambda F(\eta) &= \mathbf{E} [F(\mathfrak{N}^\Lambda(t)) \mid \mathfrak{N}^\Lambda(0) = \eta] \\ &= \mathbf{E} [F(\eta[t] + N_\Lambda) \exp(\ln(1 - e^{-t})N(\Lambda) + e^{-t}\mu(\Lambda))]. \end{aligned} \quad (10.10)$$

Posons :

$$R(t) = \exp(\ln(1 - e^{-t})N(\Lambda) + e^{-t}\mu(\Lambda)).$$

On a d'une part $R(t) \leq e^{\mu(\Lambda)}$ et d'autre part, d'après la définition 10.2, $\mathbf{E} [R(t)] = 1$ et ce pour tout $t \geq 0$. Comme $\mathfrak{N}^\Lambda(0)$ a un nombre fini d'atomes, $\mathfrak{N}^\Lambda(0)[t]$ tend presque sûrement vers la mesure nulle quand t tend vers l'infini. Par convergence dominée, on en déduit que :

$$\mathfrak{P}_t^\Lambda F(\eta) \xrightarrow{t \rightarrow \infty} \mathbf{E}_{\mu_\Lambda} [F(N_\Lambda)],$$

c'est-à-dire que \mathfrak{P}^Λ est ergodique.

La propriété (10.7) est une relation bien connue entre semi-groupe et générateur infinitésimal. Formellement, sans se soucier des convergences d'intégrale, on a :

$$\begin{aligned} \mathfrak{L}_\Lambda \left(\int_0^\infty \mathfrak{P}_t^\Lambda F \, dt \right) &= \int_0^\infty \mathfrak{L}_\Lambda \mathfrak{P}_t^\Lambda F \, dt \\ &= - \int_0^\infty \frac{d}{dt} \mathfrak{P}_t^\Lambda F \, dt \\ &= F - \mathbf{E} [F] = F, \end{aligned}$$

d'après l'ergodicité de \mathfrak{P}^Λ et puisque F est centrée.

Notons $\mathfrak{N}^\Lambda(t, N_\Lambda)$ la valeur de $\mathfrak{N}^\Lambda(t)$ lorsque la condition initiale est N_Λ . On peut écrire :

$$D_x \mathfrak{P}^\Lambda F(t) = \mathbf{E} [\mathfrak{N}^\Lambda(t, N_\Lambda + \delta_x)] - \mathbf{E} [\mathfrak{N}^\Lambda(t, N_\Lambda)].$$

Soit Y_x la durée de vie de l'atome en x . Soit $Y_x \geq t$ et l'atome est toujours vivant à t donc $\mathfrak{N}^\Lambda(t, N_\Lambda + \delta_x) = \mathfrak{N}^\Lambda(t, N_\Lambda) + \delta_x$. Soit l'atome est déjà mort à t et

$\mathfrak{N}^\Lambda(t, N_\Lambda + \delta_x) = \mathfrak{N}^\Lambda(t, N_\Lambda)$. Comme Y_x est par construction, indépendant de N_Λ et de \mathfrak{N}^Λ , il est légitime d'écrire :

$$\begin{aligned} & \mathbf{E} [F(\mathfrak{N}^\Lambda(t, N_\Lambda + \delta_x) | N_\Lambda)] - \mathbf{E} [F(\mathfrak{N}^\Lambda(t, N_\Lambda) | N_\Lambda)] \\ &= \mathbf{E} [\mathbf{1}_{\{Y_x \geq t\}} (F(\mathfrak{N}^\Lambda(t, N_\Lambda + \delta_x)) - F(\mathfrak{N}^\Lambda(t, N_\Lambda))) | N_\Lambda] \\ & \quad + \mathbf{E} [\mathbf{1}_{\{Y_x \leq t\}} (F(\mathfrak{N}^\Lambda(t, N_\Lambda)) - F(\mathfrak{N}^\Lambda(t, N_\Lambda))) | N_\Lambda] \\ &= e^{-t} \mathbf{E} [D_x F(\mathfrak{N}^\Lambda(t, N_\Lambda))], \end{aligned}$$

d'où le résultat. D'après la représentation (10.10) et l'inégalité de Jensen, on voit que :

$$|\mathfrak{P}_t^\Lambda F|^2 \leq \mathfrak{P}_t^\Lambda F^2. \quad (10.11)$$

D'après (10.7) et (10.8), on a :

$$\begin{aligned} \int_\Lambda |D_x(\mathfrak{L}_\Lambda^{-1} F(N_\Lambda))|^2 d\mu(x) &= \int_\Lambda |D_x \int_0^\infty \mathfrak{P}_t^\Lambda F(N_\Lambda) dt|^2 d\mu(x) \\ &= \int_\Lambda \left| \int_0^\infty e^{-t} \mathfrak{P}_t^\Lambda D_x F(N_\Lambda) dt \right|^2 d\mu(x). \end{aligned}$$

En vertu de l'inégalité de Jensen et de (10.11), il vient :

$$\begin{aligned} & \int_\Lambda |D_x(\mathfrak{L}_\Lambda^{-1} F(N_\Lambda))|^2 d\mu(x) \\ & \leq \int_\Lambda \int_0^\infty e^{-t} |\mathfrak{P}_t^\Lambda D_x F(N_\Lambda)|^2 dt d\mu(x) \\ & \leq \int_\Lambda \int_0^\infty e^{-t} \mathfrak{P}_t^\Lambda |D_x F(N_\Lambda)|^2 dt d\mu(x) \\ & = \int_\Lambda \int_0^\infty e^{-t} \mathbf{E} [|D_x F|^2(\mathfrak{N}^\Lambda(t)) | \mathfrak{N}^\Lambda(0) = N_\Lambda] dt d\mu(x). \end{aligned}$$

Comme $\mathfrak{N}^\Lambda(t)$ a la même loi que $\mathfrak{N}^\Lambda(0)$, si celle-ci est choisie comme celle d'un processus de Poisson d'intensité μ_Λ , lorsque l'on prend les espérances de chaque côté,

on obtient l'identité suivante :

$$\begin{aligned}
\mathbf{E}_{\mu_\Lambda} \left[\int_\Lambda |D_x(\mathfrak{L}_\Lambda^{-1}F(N_\Lambda))|^2 d\mu(x) \right] &= \mathbf{E}_{\mu_\Lambda} \left[\int_\Lambda \int_0^\infty e^{-t} |D_x F|^2(\mathfrak{N}^\Lambda(t)) dt d\mu(x) \right] \\
&= \int_\Lambda \int_0^\infty e^{-t} \mathbf{E}_{\mu_\Lambda} [|D_x F|^2(N_\Lambda)] dt d\mu(x) \\
&= \mathbf{E}_{\mu_\Lambda} \left[\int_\Lambda |D_x F(N_\Lambda)|^2 d\mu(x) \right],
\end{aligned}$$

d'où le résultat. □

THÉORÈME 10.17.— Soit F et G deux fonctions appartenant à $\text{Dom } D$. L'identité suivante est satisfaite :

$$\mathbf{E}_{\mu_\Lambda} \left[\int_\Lambda D_x F(N_\Lambda) D_x G(N_\Lambda) d\mu(x) \right] = \mathbf{E}_{\mu_\Lambda} [F(N_\Lambda) \mathfrak{L}_\Lambda G(N_\Lambda)].$$

En particulier, si G est centrée, il vient :

$$\mathbf{E}_{\mu_\Lambda} [F(N_\Lambda)G(N_\Lambda)] = \mathbf{E}_{\mu_\Lambda} \left[\int_\Lambda D_x F(N_\Lambda) D_x(\mathfrak{L}_\Lambda^{-1}G)(N_\Lambda) d\mu(x) \right]. \quad (10.12)$$

Démonstration. Soit F et G appartenant à $\text{Dom } D$, d'après (10.5) deux fois et la définition de \mathfrak{L}_Λ , on a :

$$\begin{aligned}
\mathbf{E}_{\mu_\Lambda} \left[\int_\Lambda D_x F(N_\Lambda) D_x G(N_\Lambda) d\mu(x) \right] &= \mathbf{E}_{\mu_\Lambda} \left[\int_\Lambda (F(N_\Lambda) - F(N_\Lambda - \delta_x))(G(N_\Lambda) - G(N_\Lambda - \delta_x)) dN_\Lambda(x) \right] \\
&= \mathbf{E}_{\mu_\Lambda} [G(N_\Lambda)\mathfrak{L}_\Lambda F] + \mathbf{E}_{\mu_\Lambda} \left[\int G(N_\Lambda)(F(N_\Lambda + \delta_x) - F(N_\Lambda)) d\mu(x) \right] \\
&\quad - \mathbf{E}_{\mu_\Lambda} \left[\int G(N_\Lambda - \delta_x)(F(N_\Lambda) - F(N_\Lambda - \delta_x)) dN_\Lambda(x) \right] \\
&= \mathbf{E}_{\mu_\Lambda} [G(N_\Lambda)\mathfrak{L}_\Lambda F(N_\Lambda)].
\end{aligned}$$

Le résultat s'ensuit. □

Preuve du théorème 10.11. Soit Λ un compact de E , F bornée, d'espérance nulle. D'après le théorème 10.17, nous pouvons écrire les identités suivantes :

$$\begin{aligned} \mathbf{E}_{\mu_\Lambda} \left[F(N_\Lambda) e^{\theta F(N_\Lambda)} \right] &= \mathbf{E}_{\mu_\Lambda} \left[\int D_x(\mathfrak{L}_\Lambda^{-1} F(N_\Lambda)) D_x(e^{\theta F(N_\Lambda)}) d\mu(x) \right] \\ &= \mathbf{E}_{\mu_\Lambda} \left[\int_\Lambda D_x(\mathfrak{L}_\Lambda^{-1} F(N_\Lambda)) (e^{\theta D_x F(N_\Lambda)} - 1) e^{\theta F(N_\Lambda)} d\mu(x) \right]. \end{aligned}$$

La fonction $(x \mapsto (e^x - 1)/x)$ est continue croissante sur \mathbf{R} donc on a :

$$\begin{aligned} \mathbf{E}_{\mu_\Lambda} \left[F(N_\Lambda) e^{\theta F(N_\Lambda)} \right] &= \theta \mathbf{E}_{\mu_\Lambda} \left[\int_\Lambda D_x(\mathfrak{L}_\Lambda^{-1} F(N_\Lambda)) D_x F(N_\Lambda) \frac{e^{\theta D_x F(N_\Lambda)} - 1}{\theta D_x F(N_\Lambda)} e^{\theta F(N_\Lambda)} d\mu(x) \right] \\ &\leq \theta \alpha^2 \frac{e^{\theta\beta} - 1}{\theta\beta} \mathbf{E}_{\mu_\Lambda} \left[e^{\theta F(N_\Lambda)} \right]. \end{aligned}$$

Cela implique que :

$$\frac{d}{d\theta} \log \mathbf{E}_{\mu_\Lambda} \left[e^{\theta F(N_\Lambda)} \right] \leq \alpha^2 \frac{e^{\theta\beta} - 1}{\beta},$$

donc :

$$\mathbf{E}_{\mu_\Lambda} \left[e^{\theta F(N_\Lambda)} \right] \leq \exp \left(\frac{\alpha^2}{\beta} \int_0^\theta (e^{\beta u} - 1) du \right).$$

Pour $x > 0$, pour tout $\theta > 0$, on en déduit que :

$$\begin{aligned} \mathbf{P}(F(N_\Lambda) > x) &= \mathbf{P}(e^{\theta F(N_\Lambda)} > e^{\theta x}) \\ &\leq e^{-\theta x} \mathbf{E} \left[e^{\theta F(N_\Lambda)} \right] \leq e^{-\theta x} \exp \left(\frac{\alpha^2}{\beta} \int_0^\theta (e^{\beta u} - 1) du \right). \end{aligned} \quad (10.13)$$

Ce résultat est vrai pour tout θ donc on peut optimiser en θ . A x fixé, on cherche la valeur de θ qui annule la dérivée du terme de droite par rapport à θ . En reportant dans (10.13), on obtient le résultat. \square

Preuve du théorème 10.9. On peut toujours supposer que $F \in \text{Dom } D$ est d'espérance nulle. On note respectivement ϕ et ϕ_c la fonction de répartition (respectivement fonction de répartition inverse) d'une gaussienne centrée réduite :

$$\phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-u^2/2} du, \quad \phi_c(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-u^2/2} du.$$

Pour $r > 0$, considérons ψ_r , solution de l'équation différentielle :

$$x.y(x) - y'(x) = \phi_c(r) - \mathbf{1}_{[r, \infty[}(x), \text{ pour tout } x \in \mathbf{R}.$$

On peut choisir :

$$\psi_r(x) = \begin{cases} \sqrt{2\pi}e^{x^2/2}\phi(x)\phi_c(r) & \text{si } x \leq r, \\ \sqrt{2\pi}e^{x^2/2}\phi_c(x)\phi(r) & \text{si } x \geq r. \end{cases}$$

Notons que :

$$|\psi'_r(x)| \leq \phi(r) \text{ et } |\psi''_r(x)| \leq 2. \quad (10.14)$$

On utilise alors la méthode de Stein. Il suffit de remarquer que d'après (10.12), on a :

$$\begin{aligned} \phi_c(r) - \mathbf{P}(F \geq r) &= \mathbf{E}_{\mu_\Lambda} [\mathbf{1}_{[r, \infty[}(F) - \phi_c(r)] \\ &= \mathbf{E}_{\mu_\Lambda} [F\psi_r(F) - \psi'_r(F)] \\ &= \mathbf{E}_{\mu_\Lambda} \left[\int D_x \psi_r(F) D_x \mathfrak{L}_\Lambda^{-1} F \, d\mu(x) \right] - \mathbf{E}_{\mu_\Lambda} [\psi'_r(F)]. \end{aligned} \quad (10.15)$$

D'après la formule de Taylor avec reste intégral, il s'ensuit que :

$$\begin{aligned} D_x \psi_r(F) &= \psi_r(F(N_\Lambda + \delta_x)) - \psi_r(F(N_\Lambda)) = \psi'_r(F(N_\Lambda)) D_x F(N_\Lambda) \\ &+ (D_x F(N_\Lambda))^2 \int_0^1 \psi''_r(F(N_\Lambda) + (1-u)F(N_\Lambda + \delta_x))(1-u) \, du. \end{aligned} \quad (10.16)$$

En reportant (10.16) dans (10.15) et en utilisant (10.9), d'après (10.14), il vient :

$$\begin{aligned} |\phi_c(r) - \mathbf{P}(F \geq r)| &\leq \phi(r) \mathbf{E}_{\mu_\Lambda} \left[\left| 1 - \int D_x F D_x \mathfrak{L}_\Lambda^{-1} F \, d\mu(x) \right| \right] \\ &+ 2 \mathbf{E}_{\mu_\Lambda} \left[\int_0^1 (1-u) \, du \int |D_x F(N_\Lambda)|^2 |D_x \mathfrak{L}_\Lambda^{-1} F| \, d\mu(x) \right]. \end{aligned}$$

Le résultat en découle. \square

10.5. Problèmes

EXERCICE 23.— On continue l'étude du protocole OFDMA. On suppose maintenant que l'on prend en compte le *fading* de Rayleigh. Cela signifie que le nombre de sous-porteuses réclamées par un utilisateur en position x devient :

$$N(x) = \min \left(N_M, \left\lceil \frac{C}{W \log_2 \left(1 + \frac{K F_x}{\|x\|^\gamma} \right)} \right\rceil \right) \text{ si } K F_x \|x\|^{-\gamma} > \text{SNR}_{\min},$$

où F_x est le coefficient du *fading* de Rayleigh. Du point de vue de la modélisation, cela signifie qu'à chaque atome du processus de Poisson l'on ajoute une marque représentant le *fading* de Rayleigh de cet utilisateur. On suppose généralement que :

- F_x et F_y sont des variables aléatoires indépendantes si x et y sont différents ;
- pour tout $x \in \mathbf{R}^2$, F_x suit une loi exponentielle de paramètre 1.

Le modèle est donc maintenant un processus ponctuel marqué où les réalisations sont de la forme :

$$\tilde{\xi} = \sum_{x \in \xi} \delta_{x, m}$$

avec $d\mathbf{P}_M(m) = \exp(-m) \mathbf{1}_{\mathbf{R}^+}(m) dm$.

1) Exprimer sous forme d'intégrale le nombre moyen de sous-porteuses demandées à chaque instant dans la cellule. On utilisera le même principe de découpage que précédemment pour calculer explicitement cette intégrale. On rappelle que la fonction gamma est définie par :

$$\Gamma(z) = \int_0^{\infty} e^{-m} m^{z-1} dz.$$

2) Pour $k = 1, \dots, N_M$, quelle est la loi du nombre d'utilisateurs qui demandent k sous-porteuses à un instant donné ?

3) Vérifier ce résultat sur les simulations.

4) Calculer par simulation la probabilité d'outage pour des valeurs de S allant de S_{\min} à S_{\max} avec un pas de 10.

5) Que vaut la borne de l'inégalité de concentration dans ce nouveau modèle ?

6) On fixe un seuil de perte égal à 0,01, calculer le nombre de ressources nécessaires pour obtenir une perte inférieure à ce seuil par simulation et en utilisant la borne obtenue par inégalité de concentration.

7) Calculer par simulation la probabilité d'outage pour λ , $\lambda + 10\%$, $\lambda + 20\%$. Comparer à la borne de concentration obtenue pour λ .

8) Même question en faisant varier γ de 10 %, 20 %.

9) Vos conclusions.

EXERCICE 24.- On considère X comme le processus représentant le nombre de serveurs occupés dans une file $M_\lambda/M_\mu/\infty$. On dit que $f : \mathbf{N} \rightarrow \mathbf{R}$ est c -lipschitzienne si :

$$f(n+1) - f(n) \leq c, \text{ pour tout } n.$$

1) Exprimer $X(t)$ sous forme d'une intégrale par rapport à un processus de Poisson sur $\mathbf{R}^+ \times \mathbf{R}^+$.

C	200 kb/s
W	250 kHz
K	10^6
γ	2.8
R	300 m
λ	$0,01 \text{ m}^{-2}$
SINR_{\min}	0,1
p	0,01
S_{\min}	30
S_{\max}	100

Tableau 10.5 – Valeurs numériques

On introduit D l'opérateur de différence associé à ce processus de Poisson. Soit f c -lipschitzienne.

2) Pour tout $(s, z) \in \mathbf{R}^+ \times \mathbf{R}^+$, montrer que :

$$|D_{s,z}f(X(t))| \leq c \mathbf{1}_{[0,t]}(s)$$

et que :

$$\int |D_{s,z}f(X(t))|^2 \lambda \, ds \, d\mathbf{P}_\sigma(z) \leq \lambda c^2 t.$$

3) En déduire une inégalité de concentration pour $f(X(t))$.

10.6. Notes et commentaires

La référence pour ce chapitre est bien entendu [BAC 09a, BAC 09b]. Pour de nombreux résultats sur les interférences dans un cadre poissonnien, on pourra consulter [HAE 08]. Les résultats d'analyse stochastique sont inspirés de [HOU 02] et [WU 00]. L'introduction de la dynamique de Glauber comme processus d'Ornstein-Uhlenbeck est nouvelle. On trouvera une autre approche du calcul de Malliavin pour le processus de Poisson principalement dans [PRI 09].

Résumé

- Un processus de Poisson ponctuel permet de représenter des mobiles à un instant donné ou des points d'accès de manière mathématiquement utilisable.
- La formule de Campbell permet de calculer facilement les espérances de fonctionnelles d'un processus de Poisson. Elle permet de faire le lien avec le modèle dit « fluide ».
- De la même manière qu'en dimension 1, la superposition de deux processus de Poisson indépendants est un processus de Poisson.
- Le théorème du déplacement stipule que lorsque l'on bouge les atomes d'un processus de Poisson indépendamment les uns des autres avec les mêmes statistiques, le résultat est encore un processus de Poisson dont on sait calculer l'intensité.
- L'analyse stochastique permet d'établir l'inégalité de concentration. Cette identité est utile pour borner des probabilités et ainsi définir des dimensionnements robustes.

Annexe A

Cuisine et dépendance

Les concepts mathématiques et les théorèmes ont leur existence et leur intérêt propre. Dans le cadre de la modélisation, les outils mathématiques deviennent des ustensiles au service de la résolution d'un problème concret. Pour faire de la bonne cuisine, encore faut-il connaître les ingrédients disponibles. Nous passons en revue dans ce chapitre les notions de base, rarement décrites parce que supposées connues, indispensables dans le reste de cet ouvrage. Il s'agit en particulier de définir les concepts qui permettent de mesurer la dépendance entre aléas.

A.1. Espace de probabilités, processus

A.1.1. *Dénombrabilité*

Le concept de dénombrabilité joue un rôle primordial en probabilités ne serait-ce que parce que la propriété de mesurabilité est stable par union dénombrable. Il est donc intéressant de clarifier certaines notions et résultats relatifs à ce concept.

DÉFINITION A.1.– Un ensemble est fini de cardinal n s'il est en bijection avec l'ensemble $\{1, \dots, n\}$.

DÉFINITION A.2.– Un ensemble est dit dénombrable s'il est en bijection avec \mathbf{N} , l'ensemble des entiers naturels.

EXEMPLE.– $2\mathbf{N}$, l'ensemble des entiers pairs, est dénombrable. $\mathcal{P}(\mathbf{N})$, l'ensemble des parties de \mathbf{N} , ne l'est pas. Il est en bijection avec l'ensemble des réels \mathbf{R} . En effet,

pour une partie A de \mathbf{N} , on peut définir « la suite indicatrice » $\chi_A(n)$ par :

$$\chi_A(n) = \begin{cases} 1 & \text{si } n \in A, \\ 0 & \text{sinon.} \end{cases}$$

Puis à cette suite, on peut associer le réel x_A défini par : $x_A = \sum_{n \geq 1} 2^{-n} \chi_A(n)$. Réciproquement, à un réel de $[0, 1[$, on peut associer son développement dyadique propre :

$$x = \sum_{n \geq 1} x_n 2^{-n}$$

avec $x_n \in \{0, 1\}$ pour tout n . Cela définit un ensemble en prenant :

$$A = \bigcup_{n : x_n = 1} \{n\}.$$

Par conséquent, $\mathcal{P}(\mathbf{N})$ est en bijection avec $\{0, 1\}^{\mathbf{N}}$ $\mathcal{P}(\mathbf{N})$ n'est pas dénombrable.

THÉORÈME A.1. – Si X et Y sont deux ensembles dénombrables alors $X \times Y$ et $X \cup Y$ sont dénombrables.

Tout est basé sur le résultat essentiel qui stipule de $\mathbf{N} \times \mathbf{N}$ est en bijection avec \mathbf{N} . La bijection est construite comme indiquée sur la figure A.1. Par exemple, l'élément $(2, 0)$ est envoyé sur 3 et l'élément $(0, 2)$ est envoyé sur 5.

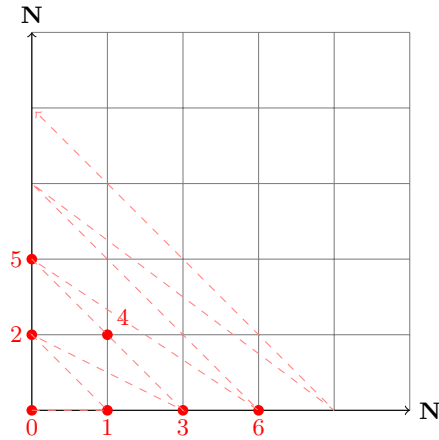


Figure A.1 – Bijection de $\mathbf{N} \times \mathbf{N}$ avec \mathbf{N}

COROLLAIRE A.2.– L'ensemble des entiers relatifs, \mathbf{Z} , est dénombrable.

Le théorème suivant est loin d'être trivial.

THÉORÈME A.3.– S'il existe une injection de l'ensemble X dans l'ensemble Y et une injection de Y dans X alors il existe une bijection de X dans Y .

COROLLAIRE A.4.– L'ensemble des rationnels, \mathbf{Q} est dénombrable.

Démonstration. Il existe évidemment une injection de \mathbf{N} dans \mathbf{Q} et par construction de \mathbf{Q} , il existe une injection de \mathbf{Q} dans $\mathbf{Z} \times \mathbf{Z}$, qui d'après le théorème précédent est en bijection avec \mathbf{N} . Par conséquent, \mathbf{Q} est en bijection avec \mathbf{N} . \square

REMARQUE.– La plupart des processus que nous rencontrerons prennent leurs valeurs dans des espaces au plus dénombrables. Grâce à l'étymologie, cela revient à dire que l'on peut numéroter les éléments de ces ensembles. Soit E un ensemble au plus dénombrable et soit $i_x \in \mathbf{N}$ l'indice correspondant à $x \in E$. Inversement, x_i le i -ème élément de E est l'élément x de E tel que $i_x = i$. Lorsque E est un sous-ensemble discret de \mathbf{R} , la numérotation est évidente. Elle demande souvent un peu d'astuce lorsque par exemple E est un espace produit comme dans le cas de la file MMPP/M/S/S, mais cette difficulté relève de la représentation informatique des données plus que des mathématiques à proprement parler.

EXEMPLE.– Soit $E = \{0, 1\} \times \{a, b, c\}$. Cet ensemble comporte six éléments. On peut décider d'un ordre en utilisant l'ordre lexicographique :

$$(0, a) < (0, b) < (0, c) < (1, a) < (1, b) < (1, c).$$

Ce qui donne $i_{(0, c)} = 3$ et $x_4 = (1, a)$.

Pour ce qui est des mathématiques, une fonction de E dans \mathbf{R} est complètement caractérisée par son « graphe » $((x, f(x)), x \in E)$ mais comme E est dénombrable, on identifie cet ensemble au vecteur colonne dont la i_x -ème composante est $f(x)$. Une mesure sur un espace dénombrable est caractérisée par sa valeur sur les singletons donc s'identifie aussi à un vecteur $\pi = (\pi(i_x), x \in E)$. Ce vecteur est classiquement écrit en ligne car mesures et fonctions sont en dualité (penser au théorème de Riesz), l'intégrale de f sous la mesure π s'écrit :

$$\int f d\pi = \sum_{x \in E} f(x)\pi(x) = (\pi(x_1), \dots, \pi(x_n), \dots) \begin{pmatrix} f(x_1) \\ \vdots \\ f(x_n) \\ \vdots \end{pmatrix},$$

où le vecteur ligne et le vecteur colonne sont multipliés selon les règles du produit matriciel. Il est parfois plus pratique de raisonner en termes de fonctions, notamment pour les preuves mathématiques, parfois plus agréable d'utiliser la notation vectorielle, en particulier pour les calculs à l'ordinateur ou à la main. Dans le cas où l'on raisonne avec des fonctions, il est plus naturel de noter les états par x, y , etc. Dans le cas où l'on fait référence à des vecteurs, il est d'usage de numéroter les états par i, j , etc.

A.1.2. Espaces polonais

Les espaces polonais sont une catégorie d'ensembles topologiques suffisamment générale pour de nombreuses applications et qui permettent de faire des probabilités de façon rigoureuse sans douleur.

DÉFINITION A.3.– Un espace polonais E est un espace muni d'une distance d qui satisfait les deux propriétés suivantes :

- il est complet : toute suite de Cauchy converge ;
- il est séparable : il existe une suite $(x_n, n \geq 1)$ d'éléments de E dense dans E , c'est-à-dire telle que pour tout $x \in E$, tout $\epsilon > 0$, il existe x_n tel que $d(x_n, x) < \epsilon$.

EXEMPLE.– Les espaces $\mathbf{R}, \mathbf{R}^n, \mathbf{C}^n$ sont des espaces polonais pour la distance usuelle. Des familles dénombrables denses sont respectivement $\mathbf{Q}, (\mathbf{Q})^n, (\mathbf{Q} + i\mathbf{Q})^n$.

EXEMPLE.– Plus généralement, nous serons amenés à manipuler des suites infinies de réels. Dès que la taille d'un « vecteur » devient infinie, se posent des problèmes de sommabilité ou de « bornitude ». Si E est un espace dénombrable, on considère donc souvent des sous-espaces de l'ensemble des suites de réels indexées par E , ensemble que l'on note \mathbf{R}^E . Les deux sous-ensembles de \mathbf{R}^E les plus importants pour nous seront $l^\infty(E)$ et $l^2(E)$.

DÉFINITION A.4.– L'ensemble $l^\infty(E)$ est l'ensemble des suites de réels indexées par E dénombrable, qui sont bornées :

$$u = (u(x), x \in E) \in l^\infty(E) \iff \sup_{x \in E} |u(x)| < \infty.$$

On note alors $\|u\|_\infty = \sup_{x \in E} |u(x)|$. L'espace $(l^\infty(E), \|\cdot\|_\infty)$ est un espace vectoriel normé complet, autrement dit, un espace de Banach. Soit ϕ une bijection entre \mathbf{N} et E et $E_n = \phi(\{0, \dots, n\})$. L'ensemble $\bigcup_{n \geq 0} \mathbf{Q}^{E_n}$ est évidemment dénombrable et dense donc $(l^\infty(E), \|\cdot\|_\infty)$ est polonais.

DÉFINITION A.5.– Soit E un ensemble dénombrable et π une mesure sur E . L'ensemble $l^2(E, \pi)$ est l'ensemble des suites de réels indexées par E dénombrable qui

sont de carré sommable pour π :

$$u = (u(x), x \in E) \in l^2(E) \iff \sum_{x \in E} |u(x)|^2 \pi(x) < \infty.$$

On note alors $\|u\|_2 = (\sum_{x \in E} |u(x)|^2 \pi(x))^{1/2}$. L'espace $(l^2(E, \pi), \|\cdot\|_2)$ est un espace vectoriel muni d'un produit scalaire :

$$\langle u, v \rangle_{l^2(E, \pi)} = \sum_{x \in E} u(x)v(x)\pi(x),$$

qui généralise le produit scalaire des vecteurs de \mathbf{R}^n . De plus, $(l^2(E), \|\cdot\|_2)$ est complet pour ce produit scalaire, c'est donc un espace de Hilbert. L'ensemble $\bigcup_{n \geq 0} \mathbf{Q}^{E_n}$ est aussi dense dans $(l^2(E, \pi), \|\cdot\|_2)$ donc $(l^2(E, \pi), \|\cdot\|_2)$ est polonais.

EXEMPLE.– L'espace des fonctions continues sur $[0, T]$ à valeurs dans \mathbf{R} est aussi un espace polonais pour la distance induite par la norme uniforme :

$$d(f, g) = \sup_{0 \leq t \leq T} |f(t) - g(t)|.$$

Il est bien connu que cet espace est complet. D'après le théorème de Weierstrass, les polynômes à coefficients rationnels forment une famille dénombrable dense. En effet, l'ensemble des polynômes à coefficients de degré au plus k est en bijection avec \mathbf{Q}^{k+1} donc la réunion de ces ensembles est une réunion dénombrable d'ensembles dénombrables donc elle est dénombrable.

EXEMPLE.– L'espace $D([0, T], \mathbf{R})$ des fonctions continues à droite avec des limites à gauche (cadlag) est généralement muni de la distance :

$$d(f, g) = \inf_{\phi \in \mathfrak{H}_T} \sup_{0 \leq t \leq T} |f(t) - g(\phi(t))|,$$

où \mathfrak{H}_T est l'ensemble des homéomorphismes (bijections continues, d'inverse continu) de $[0, T]$ dans lui-même. On démontre que cet espace est complet et que la famille de polynômes à coefficients rationnels définie dans le cas des fonctions continues est encore dense.

EXEMPLE A.1.– Le dernier exemple dont nous aurons besoin est celui des configurations sur un espace lui-même polonais E .

DÉFINITION A.6.– Une configuration est un ensemble de points d'un ensemble E localement fini : il y a un nombre fini de points dans tout ensemble borné. On note \mathfrak{N}_E l'ensemble des configurations de E .

On note ξ , l'élément générique de l'ensemble des configurations. Chaque ξ est un ensemble, c'est-à-dire $\xi = \{x_1, x_2, \dots\}$, mais on l'identifie aussi à une mesure

ponctuelle : $\xi = \sum_{x \in \xi} \delta_x$. Suivant les contextes, l'une ou l'autre des représentations est la plus agréable. La configuration vide notée \emptyset correspond à la mesure nulle, à ne pas confondre avec la mesure de Dirac en 0. Une configuration a au plus un nombre dénombrable d'éléments, on peut donc indexer ceux-ci par les entiers mais il n'y a pas d'ordre préférentiel : x_1 peut représenter n'importe lequel des atomes de ξ . C'est pourquoi il est préférable de garder une notation ensembliste tant que possible. Pour $f : E \rightarrow \mathbf{R}$, on pose (sans se soucier de la convergence de la série pour le moment) :

$$\int f \, d\xi = \sum_{x \in \xi} f(x).$$

Pour A un ensemble de E , $\xi(A)$ est le nombre de points de $\xi \cap A$. On note aussi $|\xi| = \xi(E)$ le nombre de points total de ξ , éventuellement infini. Enfin, il est très fréquent que l'on soit obligé de se restreindre à des sous-ensembles de E . On note ξ_B la restriction de ξ à B : $\xi_B(A) = \xi(A \cap B)$.

EXEMPLE.— L'ensemble $\{(n, 0), n \in \mathbf{N}^*\}$ est une configuration. En revanche, l'ensemble $\{(1/n, 0), n \in \mathbf{N}^*\}$ n'en est pas une car il y a un nombre infini de points dans l'intervalle $[0, 1]$. Voir figure A.2.



Figure A.2 – Exemple et contre-exemple de configurations

DÉFINITION A.7.— Soit E un espace polonais, un processus ponctuel N sur E est une variable aléatoire à valeurs dans l'ensemble \mathfrak{N}_E des configurations de E .

DÉFINITION A.8.— Soit E un espace polonais et \mathfrak{N}_E l'ensemble des configurations de E . La suite de configurations $(\xi_n, n \geq 1)$ converge vers la configuration ξ au sens vague si pour toute fonction f continue à support compact de E dans \mathbf{R} , on a :

$$\int f \, d\xi_n \xrightarrow{n \rightarrow +\infty} \int f \, d\xi.$$

Cela veut presque dire que les positions des atomes des ξ_n tendent vers celles des atomes de ξ mais il peut y avoir fuite de masse à l'infini :

$$\delta_n \xrightarrow{n \rightarrow +\infty} 0 \text{ car pour } f \text{ à support compact } \int f \, d\delta_n \xrightarrow{n \rightarrow +\infty} 0.$$

THÉORÈME A.5.— L'espace \mathfrak{N}_E muni de la convergence vague est polonais.

Pour construire la distance sur \mathfrak{N}_E , on considère une suite dense $(x_n, n \geq 1)$ d'éléments de E . On note $B_{n,q} = \{x \in E, d_E(x, x_n) \leq 1/q\}$, la boule ouverte de centre x_n et de rayon $1/q$. La famille dénombrable de fonctions $(f_{n,q} = \mathbf{1}_{B_{n,q}}, n \geq 1, q \geq 1)$ engendre la tribu $\mathcal{B}(E)$. Pour deux configurations ω et η de \mathfrak{N}_E , on pose :

$$d(\omega, \eta) = \sum_{n \geq 1} \sum_{q \geq 1} 2^{-(n+q)} \zeta(|\int f_{n,q}(d\omega - d\eta)|),$$

où $\zeta(x) = x/(1+x)$. On peut montrer que cette distance engendre la même topologie que celle de la convergence vague, c'est-à-dire que les suites convergentes sont les mêmes dans les deux cas. Pour construire une famille dénombrable dense, il suffit de s'inspirer du schéma de ce que l'on a fait pour les fonctions continues. E est polonais donc il existe une famille dénombrable dense $\mathcal{Q} = (x_n, n \in \mathbf{N})$. L'ensemble des mesures ponctuelles simples finies dont les atomes appartiennent à \mathcal{Q} joue le même rôle que les polynômes à coefficients rationnels.

Les espaces polonais possèdent des propriétés sympathiques pour les probabilistes. Par exemple, les variables aléatoires définies sur un espace polonais sont automatiquement « tendues » : pour tout $\epsilon > 0$, il existe un compact K tel que $\mathbf{P}(X \in K^c) \leq \epsilon$. Par ailleurs, les tribus sont dites à base dénombrable : il existe une famille dénombrable d'ensembles qui engendrent la tribu (penser aux intervalles à extrémités rationnelles dans le cas de \mathbf{R}). On peut alors réduire certaines preuves à des vérifications sur ces ensembles particulièrement simples et appliquer ce qu'il est convenu d'appeler un « principe de classe monotone », voir le lemme A.8.

A.1.3. Processus

Un processus est une famille de variables aléatoires indexée par $\mathfrak{T} = \mathbf{N}$ ou $\mathfrak{T} = \mathbf{R}^+$, plus rarement par \mathbf{Z} ou \mathbf{R} . Ces variables aléatoires sont supposées être définies sur le même espace Ω et prendre leur valeur dans le même espace polonais E , ce qui revient à dire que l'on a une application X comme suit :

$$\begin{aligned} X : \Omega \times \mathfrak{T} &\longrightarrow E \\ (\omega, t) &\longmapsto X(\omega, t). \end{aligned}$$

On peut tout aussi bien ne garder comme variable que l'aléa ω et considérer X comme une variable aléatoire à valeurs dans les applications de \mathfrak{T} dans E , ensemble noté usuellement $E^{\mathfrak{T}}$:

$$\begin{aligned} X : \Omega &\longrightarrow E^{\mathfrak{T}} \\ \omega &\longmapsto (t \mapsto X(\omega, t)). \end{aligned}$$

Dans cette description, la valeur de $X(\omega)$ s'appelle la trajectoire de X , c'est l'ensemble des éléments de E obtenus en fixant ω et en faisant varier t dans \mathfrak{T} .

La première description est la plus naturelle car elle correspond à l'idée que l'on se fait d'un système dynamique : une suite de valeurs indexées par le temps. La deuxième description plus abstraite permet de construire le cadre mathématique rigoureux.

On sait bien qu'en probabilité, seul compte l'espace d'arrivée des variables aléatoires, l'espace Ω est souvent mal défini. Pour se simplifier la vie, il est courant de considérer que Ω est déjà l'espace des valeurs. Dans le cas de la deuxième présentation, l'aléa devient directement une trajectoire. Les valeurs à l'instant t deviennent les applications coordonnées :

$$\begin{aligned} X_t : E^{\mathfrak{T}} &\longrightarrow E \\ \omega &\longmapsto X_t(\omega) = \omega(t). \end{aligned}$$

A partir du moment, où l'on prend pour Ω l'espace des trajectoires et que l'on veut munir cet espace d'une probabilité, on mesure la difficulté qu'il y a à travailler sur un espace « compliqué », ce qui est un aspect occulté mais implicite dans la présentation avec un Ω quelconque.

Dans le cas où l'ensemble d'indices est dénombrable, $E^{\mathbb{N}}$ hérite assez naturellement de la topologie produit de celle de E qui en fait un espace polonais. La loi de X est alors déterminée par les lois marginales fini-dimensionnelles. Cela signifie que deux processus X et Y tels que :

$$\text{Loi}(X_0, \dots, X_n) = \text{Loi}(Y_0, \dots, Y_n) \text{ pour tout } n \geq 0,$$

ont la même loi. En revanche, il ne suffit pas que X_n ait la même loi que Y_n . Par exemple, si les Y_n sont des copies indépendantes de X et $X_n = X$ pour tout n alors $\mathbf{P}(X_1 \neq X_2) = 0$ et $\mathbf{P}(Y_1 \neq Y_2) > 0$, ce qui empêche l'égalité en loi des deux processus.

La situation est plus complexe si l'espace d'indices a la puissance du continu car $\mathcal{E}^{\mathbb{R}^+}$ n'est pas naturellement équipé d'une topologie. Il y a deux choix essentiels d'espace dont les propriétés sont présentées plus bas : soit l'espace des fonctions continues, soit celui des fonctions continues à droite avec des limites à gauche. Dans les deux cas, la loi d'un processus est déterminée par les lois marginales fini-dimensionnelles : deux processus X et Y ont même loi si et seulement si :

$$\text{Loi}(X_{t_1}, \dots, X_{t_n}) = \text{Loi}(Y_{t_1}, \dots, Y_{t_n}) \text{ pour tout } t_1, \dots, t_n \in \mathfrak{T}.$$

A.1.4. Tribus

DÉFINITION A.9.– Une tribu \mathcal{E} d'un ensemble E est un sous-ensemble de parties de E qui satisfait les trois propriétés suivantes :

- $\emptyset \in \mathcal{E}$;
- si $A \in \mathcal{E}$ alors $A^c \in \mathcal{E}$;
- si $(A_n, n \geq 1)$ est une famille dénombrables d'éléments de \mathcal{E} alors $\bigcup_{n \geq 1} A_n$ est un élément de \mathcal{E} .

Les ensembles d'une tribu sont souvent appelés ensembles mesurables parce que ce sont ceux dont on pourra calculer la (mesure de) probabilité.

EXEMPLE.– Les exemples classiques de tribus sont la tribu « grossière » : $\mathcal{E} = \{\emptyset, E\}$ et la tribu de toutes les parties de E : $\mathcal{E} = \mathcal{P}(E)$. Nous en verrons d'autres exemples ci-dessous avec notamment le concept de tribu engendrée.

REMARQUE.– Pour une variable aléatoire discrète, on voudra calculer des quantités comme $\mathbf{P}(X = i)$, c'est-à-dire calculer la probabilité du singleton $\{i\}$ sous la loi de X . Il faut donc que tous les singletons soient des ensembles mesurables. Dans un espace dénombrable, tous les ensembles s'écrivent comme réunion au plus dénombrable des singletons qui les composent. Par stabilité d'une tribu par réunion dénombrable, on en déduit que si les singletons sont mesurables, tous les ensembles le sont.

La situation est totalement différente pour un espace non dénombrable. On veut toujours que les singletons soient des mesurables mais cela ne suffit pas à couvrir toutes les parties de E . De plus, on montre que, sauf à renoncer à l'axiome du choix, l'on ne peut pas raisonnablement construire de mesure définie sur l'ensemble des parties d'un espace non dénombrable. La bonne notion est celle de tribu engendrée.

DÉFINITION A.10.– Soit $\mathcal{C} \subset \mathcal{P}(E)$, on note $\sigma(\mathcal{C})$ la plus petite tribu qui contient \mathcal{C} .

EXEMPLE A.2.– Pour un ensemble de $A \in E$, $\sigma(A) = \{\emptyset, A, A^c, E\}$.

DÉFINITION A.11.– Pour un ensemble polonais E , on note $\mathfrak{B}(E)$ la tribu des boréliens de E définie comme la tribu engendrée par les ouverts de la topologie de E .

Pour une application X de E dans \mathbf{R}^d , on note $\sigma(X)$ la plus petite tribu qui rend X mesurable de $(\mathbf{R}^d, \sigma(X))$ dans $(\mathbf{R}^d, \mathfrak{B}(\mathbf{R}^d))$.

LEMME A.6.– Si X est mesurable de $(E, \{\emptyset, A, A^c, E\})$ dans (F, \mathcal{F}) un autre espace polonais alors X est de la forme : il existe f_1 et f_2 dans F tels que :

$$X(\omega) = \begin{cases} f_1 & \text{si } \omega \in A, \\ f_2 & \text{si } \omega \in A^c. \end{cases} \quad (\text{A.1})$$

Démonstration. En effet, soit $\omega_1 \in E$. Comme X est mesurable $B = X^{-1}(\{X(\omega_1)\})$ vaut l'un des quatre ensembles $\{\emptyset, A, A^c, E\}$. Comme $\omega_1 \in B$, B ne peut être vide. Si $B = E$, cela signifie que pour tout $\omega \in B = E$, $X(\omega) = X(\omega_1)$ donc X est constante. La variable aléatoire X est bien de la forme (A.1) avec $f_1 = f_2 = X(\omega_1)$. Si $B = A$ alors $X(\omega) = X(\omega_1)$ pour tout $\omega \in A$. Soit $\omega_2 \in A^c$, l'ensemble $C = X^{-1}(\{X(\omega_2)\})$ ne peut être égal qu'à A^c et donc $X(\omega) = X(\omega_2)$ pour tout $\omega \in A^c$. Par conséquent, X est bien de la forme (A.1) avec $f_i = X(\omega_i)$. \square

Plus généralement, on a le théorème suivant.

THÉORÈME A.7.— Soit $X : E \longrightarrow (F, \mathcal{F})$ et $Y : E \longrightarrow (\mathbf{R}, \mathfrak{B}(\mathbf{R}))$ où E et F sont deux espaces polonais. La variable aléatoire Y est $\sigma(X)$ mesurable si et seulement s'il existe $\psi : F \longrightarrow \mathbf{R}$ mesurable telle que $Y = \psi(X)$.

Démonstration. La tribu engendrée par X contient nécessairement les ensembles de la forme $X^{-1}(A)$ pour $A \in \mathcal{F}$. Comme cet ensemble d'ensembles constitue une tribu, $\sigma(X) = \{X^{-1}(A), A \in \mathcal{F}\}$. Si Y est de la forme $\mathbf{1}_B$ alors $Y^{-1}(\{1\}) = B$ appartient à $\sigma(X)$ donc il existe $C \in \mathcal{F}$ tel que $B = X^{-1}(C)$. Par conséquent, on a $Y = \mathbf{1}_C(X)$.

Soit maintenant Y étagée :

$$Y = \sum_{i=1}^n \alpha_i \mathbf{1}_{A_i}$$

avec $A_i \cap A_j = \emptyset$ et $\alpha_i - \alpha_j \neq 0$ pour $i \neq j$. Puisque $A_i = Y^{-1}(\{\alpha_i\})$, par le même raisonnement, on construit C_1, \dots, C_n ensembles \mathcal{F} mesurables tels que $Y = (\sum_{i=1}^n \alpha_i \mathbf{1}_{C_i})(X)$. Soit Y variable aléatoire positive \mathcal{F} mesurable, on sait qu'il existe Y_n étagée qui converge presque sûrement vers Y . On a précédemment construit ψ_n tel que $Y_n = \psi_n(X)$. Comme Y_n converge vers Y , ψ_n converge vers quelque chose sur l'image de E par X . Malheureusement, rien ne garantit que cet ensemble soit mesurable. Pour éviter cet écueil, on pose $\psi = \limsup_n \psi_n$. Comme toute limite supérieure de fonctions mesurables est mesurable, cette fonction est mesurable. De plus, aux endroits où ψ_n converge vers ψ , la limite supérieure aussi. En conclusion, on a bien construit ψ mesurable telle que $Y = \psi(X)$. \square

L'un des principes les plus utiles pour établir des formules en les réduisant à des vérifications sur des ensembles est celui de classe monotone, que nous énonçons maintenant.

DÉFINITION A.12.— Un π -système est un ensemble de parties de E polonais stable par intersection finie.

DÉFINITION A.13.– Un λ -système \mathcal{D} est un ensemble de parties de E polonais stable par limite croissante (la réunion d'une famille croissante d'éléments de \mathcal{D} appartient à \mathcal{D}) et différence (si $B \subset A$ avec $A, B \in \mathcal{D}$ alors $A \setminus B \in \mathcal{D}$.)

THÉORÈME A.8.– Soit \mathcal{C} un π -système et \mathcal{D} un λ -système d'un espace polonais E . Si $\mathcal{C} \subset \mathcal{D}$ alors $\sigma(\mathcal{C}) \subset \mathcal{D}$.

A.2. Espérance conditionnelle

Pour (E, \mathcal{E}) polonais, on note $L^2(E, \mathcal{E}, \mathbf{P})$ l'espace des variables aléatoires de carré intégrable : les variables aléatoires telles que $\mathbf{E}[X^2] < \infty$. Pour \mathcal{F} sous-tribu de \mathcal{E} , la théorie des espaces de Hilbert stipule alors que l'on peut définir la projection orthogonale sur $L^2(E, \mathcal{F}, \mathbf{P})$.

DÉFINITION A.14.– Soit $X \in L^2(E, \mathcal{E}, \mathbf{P})$, on note $\mathbf{E}[X | \mathcal{F}]$, l'espérance conditionnelle de X sachant \mathcal{F} , définie comme la projection orthogonale de X sur l'espace de Hilbert $L^2(E, \mathcal{F}, \mathbf{P})$.

Par définition d'une projection orthogonale, cela signifie que :

$$\mathbf{E}[X | \mathcal{F}] \text{ est } \mathcal{F}\text{-mesurable et } \mathbf{E}[ZX] = \mathbf{E}[Z\mathbf{E}[X | \mathcal{F}]], \quad (\text{A.2})$$

pour toute variable aléatoire Z bornée, \mathcal{F} -mesurable. Par analogie avec le cas non conditionnel, on introduit la probabilité conditionnelle sachant (une tribu) \mathcal{F} , par :

$$\mathbf{P}(A | \mathcal{F}) = \mathbf{E}[\mathbf{1}_A | \mathcal{F}].$$

Ainsi définie, l'espérance conditionnelle et la probabilité conditionnelle sont des variables aléatoires.

EXEMPLE.– Soit $\mathcal{F} = \{\emptyset, A, A^c, \Omega\}$ où $A \in E$. Donc \mathcal{F} est une sous-tribu de \mathcal{E} car $A \in E$. Calculons $\mathbf{E}[X | \mathcal{F}]$ pour un $X \in L^2(E, \mathcal{E}, \mathbf{P})$. D'après le lemme A.6, si Z est une variable aléatoire \mathcal{F} -mesurable alors elle est forcément du type :

$$Z = a\mathbf{1}_A + b\mathbf{1}_{A^c}, \quad a, b \in \mathbf{R},$$

par conséquent $\mathbf{E}[X | \mathcal{F}]$ sera de la forme aussi $c\mathbf{1}_A + d\mathbf{1}_{A^c}$, et il suffit de déterminer les constantes c et d . En remplaçant $\mathbf{E}[X | \mathcal{F}]$ par $c\mathbf{1}_A + d\mathbf{1}_{A^c}$ dans (A.2), pour tout a et tout b , nous obtenons :

$$\mathbf{E}[(a\mathbf{1}_A + b\mathbf{1}_{A^c})(c\mathbf{1}_A + d\mathbf{1}_{A^c})] = ac\mathbf{P}(A) + bd\mathbf{P}(A^c),$$

donc :

$$\mathbf{E}[X(a\mathbf{1}_A + b\mathbf{1}_{A^c})] = ac\mathbf{P}(A) + bd\mathbf{P}(A^c).$$

Par conséquent, en choisissant successivement $a = 0$, b quelconque puis $b = 0$, a quelconque, nous obtenons :

$$\mathbf{E}[X | \mathcal{F}] = \frac{\mathbf{E}[X \mathbf{1}_A]}{\mathbf{P}(A)} \mathbf{1}_A + \frac{\mathbf{E}[X \mathbf{1}_{A^c}]}{\mathbf{P}(A^c)} \mathbf{1}_{A^c}.$$

En particulier, pour $X = \mathbf{1}_C$, $C \in E$, on a :

$$\mathbf{P}(C | \mathcal{F}) = \mathbf{P}(C | A) \mathbf{1}_A + \mathbf{P}(C | A^c) \mathbf{1}_{A^c},$$

d'où le lien avec la notion de probabilité conditionnelle élémentaire.

Les résultats suivants se trouvent dans tous les ouvrages sur l'espérance conditionnelle, nous en omettons les preuves.

THÉORÈME A.9.— Soit X une variable aléatoire de $L^2(E, \mathcal{E}, \mathbf{P})$ et \mathcal{F} une sous-tribu de \mathcal{E} . On a les propriétés suivantes :

- 1) si $X \geq 0$ p.s. alors $\mathbf{E}[X | \mathcal{F}] \geq 0$;
- 2) $\mathbf{E}[|\mathbf{E}[X | \mathcal{F}]|] \leq \mathbf{E}[|X|]$;
- 3) pour tout $X \in L^1(E, \mathcal{E}, \mathbf{P})$, il existe une unique variable aléatoire, notée $\mathbf{E}[X | \mathcal{F}]$, qui satisfait (A.2);
- 4) si X est \mathcal{F} -mesurable et $Y \in \mathcal{E}$ mesurable telles que $XY \in L^1$ alors $\mathbf{E}[XY | \mathcal{F}] = X \mathbf{E}[Y | \mathcal{F}]$;
- 5) si $X \in L^1$ est indépendante de \mathcal{F} alors $\mathbf{E}[X | \mathcal{F}] = \mathbf{E}[X]$;
- 6) si $\mathcal{F} \subset \mathcal{G} \subset \mathcal{E}$ alors pour tout $X \in L^1$, on a :

$$\mathbf{E}[\mathbf{E}[X | \mathcal{G}] | \mathcal{F}] = \mathbf{E}[\mathbf{E}[X | \mathcal{F}] | \mathcal{G}] = \mathbf{E}[X | \mathcal{F}].$$

A.2.1. Indépendance et conditionnement

Les résultats suivants sur l'indépendance conditionnelle sont moins connus mais tout à fait essentiels pour établir de façon économique les différentes formes de la propriété de Markov.

DÉFINITION A.15.— Soit $\mathcal{F}_1, \dots, \mathcal{F}_n, \mathcal{G} \subset \mathcal{E}$ des tribus. Les tribus $\mathcal{F}_1, \dots, \mathcal{F}_n$ sont indépendantes conditionnellement à \mathcal{G} lorsque :

$$\mathbf{P}\left(\bigcap_{j=1}^n B_j \mid \mathcal{G}\right) = \prod_{j=1}^n \mathbf{P}(B_j \mid \mathcal{G}) \text{ presque sûrement,} \quad (\text{A.3})$$

pour tout $B_j \in \mathcal{F}_j$, $j = 1, \dots, n$.

Une famille infinie de tribus $(\mathcal{F}_r, r \in T)$ est indépendante conditionnellement à \mathcal{G} si (A.3) est vraie pour toute sous-famille finie.

THÉORÈME A.10.– Soit \mathcal{F} , \mathcal{G} et \mathcal{H} trois tribus de (E, \mathcal{E}) . Il y a équivalence entre les trois propriétés suivantes :

- 1) \mathcal{F} et \mathcal{H} sont indépendantes conditionnellement à \mathcal{G} ;
- 2) pour tout $H \in \mathcal{H}$, $\mathbf{P}(H | \mathcal{F} \vee \mathcal{G}) = \mathbf{P}(H | \mathcal{G})$, presque sûrement;
- 3) \mathcal{H} et $\mathcal{F} \vee \mathcal{G}$ sont indépendantes conditionnellement à \mathcal{G} .

Démonstration. Supposons que \mathcal{F} et \mathcal{H} sont indépendantes conditionnellement à \mathcal{G} . Pour $F \in \mathcal{F}$, $G \in \mathcal{G}$ et $H \in \mathcal{H}$, on a :

$$\begin{aligned} \mathbf{E}[\mathbf{P}(H | \mathcal{G}) \mathbf{1}_F \mathbf{1}_G] &= \mathbf{E}[\mathbf{P}(H | \mathcal{G}) \mathbf{P}(F | \mathcal{G}) \mathbf{1}_G] \\ &= \mathbf{E}[\mathbf{P}(H \cap F | \mathcal{G}) \mathbf{1}_G] \end{aligned}$$

par définition de l'indépendance conditionnelle. De la propriété 4 du théorème (A.9), on déduit que :

$$\mathbf{E}[\mathbf{P}(H \cap F | \mathcal{G}) \mathbf{1}_G] = \mathbf{E}[\mathbf{1}_H \mathbf{1}_{F \cap G}].$$

On note :

$$\mathcal{D} = \{M = F \cap G, \mathbf{E}[\mathbf{P}(H | \mathcal{G}) \mathbf{1}_M] = \mathbf{E}[\mathbf{1}_H \mathbf{1}_M]\}.$$

De ce qui précède, $\mathcal{C} = \{M = F \cap G, F \in \mathcal{F}, G \in \mathcal{G}\} \subset \mathcal{D}$. Il est évident que \mathcal{C} est un π -système. Par linéarité et convergence monotone, il apparaît que \mathcal{D} est un λ -système. En vertu du théorème A.8, \mathcal{D} contient la tribu engendrée par \mathcal{C} . Or $\mathcal{F} \subset \mathcal{C}$ et $\mathcal{G} \subset \mathcal{G}$ donc \mathcal{C} contient $\mathcal{F} \vee \mathcal{G}$. Ceci signifie que pour tout $M \in \mathcal{F} \vee \mathcal{G}$, on a :

$$\mathbf{E}[\mathbf{P}(H | \mathcal{G}) \mathbf{1}_M] = \mathbf{E}[\mathbf{1}_H \mathbf{1}_M].$$

Comme $\mathbf{P}(H | \mathcal{G})$ est $\mathcal{F} \vee \mathcal{G}$ mesurable, on conclut de cette identité que le point 2 est vérifié.

Réciproquement si le point 2 est vérifié, pour tout $F \in \mathcal{F}$ et tout $H \in \mathcal{H}$, on obtient :

$$\begin{aligned} \mathbf{P}(F \cap H | \mathcal{G}) &= \mathbf{E}[\mathbf{P}(F \cap H | \mathcal{F} \vee \mathcal{G}) | \mathcal{G}] \\ &= \mathbf{E}[\mathbf{1}_F \mathbf{P}(H | \mathcal{F} \vee \mathcal{G}) | \mathcal{G}] \\ &= \mathbf{E}[\mathbf{1}_F \mathbf{P}(H | \mathcal{G}) | \mathcal{G}] \\ &= \mathbf{P}(H | \mathcal{G}) \mathbf{P}(F | \mathcal{G}). \end{aligned}$$

Ce qui prouve l'indépendance de \mathcal{F} et \mathcal{H} conditionnellement à \mathcal{G} .

D'après le point 2, \mathcal{H} et $\mathcal{F} \vee \mathcal{G}$ sont indépendantes conditionnellement à \mathcal{G} si et seulement si $\mathbf{P}(H | \mathcal{F} \vee \mathcal{G}) = \mathbf{P}(H | \mathcal{G})$ pour tout $H \in \mathcal{H}$. C'est très exactement la même condition que celle qui induit que \mathcal{F} et \mathcal{H} sont indépendantes conditionnellement à \mathcal{G} . L'équivalence des points 1 et 3 en découle. \square

Le raisonnement de la première étape s'appelle un « argument de classe monotone ». Les utilisations suivantes d'un tel procédé ne seront pas détaillées, le principe en étant toujours le même.

THÉORÈME A.11.— Soit $\mathcal{G}, \mathcal{H}, \mathcal{F}_1, \dots, \mathcal{F}_n, \dots$ des tribus. Les propositions suivantes sont équivalentes :

- 1) les tribus \mathcal{H} et $\bigvee_n \mathcal{F}_n$ sont indépendantes conditionnellement à \mathcal{G} ;
- 2) pour tout entier n , les tribus \mathcal{H} et \mathcal{F}_{n+1} sont indépendantes conditionnellement à $\mathcal{G} \vee \mathcal{F}_1 \vee \dots \vee \mathcal{F}_n$.

Démonstration. Si \mathcal{H} et $\bigvee_n \mathcal{F}_n$ sont indépendantes conditionnellement à \mathcal{G} alors \mathcal{H} et toute tribu engendrée par une sous-famille finie des \mathcal{F}_j sont indépendantes conditionnellement à \mathcal{G} . Appliquons le théorème A.10 avec $\mathcal{F} = \bigvee_{j=1}^n \mathcal{F}_j$ puis $\mathcal{F} = \bigvee_{j=1}^{n+1} \mathcal{F}_j$, on obtient :

$$\mathbf{P}(H | \mathcal{G}) = \mathbf{P}(H | \mathcal{G} \vee \bigvee_{j=1}^n \mathcal{F}_j) \text{ et } \mathbf{P}(H | \mathcal{G}) = \mathbf{P}(H | \mathcal{G} \vee \bigvee_{j=1}^{n+1} \mathcal{F}_j).$$

En appliquant une nouvelle fois le théorème A.10 avec $\mathcal{F} = \mathcal{F}_{n+1}$, on en déduit le point 2.

Réciproquement, supposons que pour tout $n \geq 0$, pour tout $H \in \mathcal{H}$, on ait :

$$\mathbf{P}(H | \mathcal{G} \vee \bigvee_{j=1}^n \mathcal{F}_j) = \mathbf{P}(H | \mathcal{G} \vee \bigvee_{j=1}^{n+1} \mathcal{F}_j).$$

Par transitivité de la relation d'égalité, on a alors :

$$\mathbf{P}(H | \mathcal{G}) = \mathbf{P}(H | \mathcal{G} \vee \bigvee_{j=1}^m \mathcal{F}_j) \text{ pour tout } m.$$

Toujours d'après le théorème A.10, on en déduit que \mathcal{H} et $\bigvee_{j=1}^m \mathcal{F}_j$ sont indépendantes conditionnellement à \mathcal{G} . Par définition de l'indépendance conditionnelle pour un nombre infini de tribus, cela suffit à montrer le point 1. \square

A.2.2. Propriété de Markov

Introduisons d'abord l'opérateur de décalage θ_t . Supposons que l'on ait une famille de variables aléatoires à valeurs dans E , indexées par \mathfrak{T} dénombrable ou pas. En dehors de toute considération topologique, les trajectoires de ce processus sont des éléments de $E^{\mathfrak{T}}$, les applications de \mathfrak{T} dans E .

DÉFINITION A.16.– Pour tout $t \in \mathfrak{T}$, l'opérateur de décalage θ_t est défini par :

$$\begin{aligned} \theta_t : E^{\mathfrak{T}} &\longrightarrow E^{\mathfrak{T}} \\ (\omega_s, s \in \mathfrak{T}) &\longmapsto (\omega_{t+s}, s \in \mathfrak{T}). \end{aligned}$$

⌚ La trajectoire $\theta_t \omega$ est donc la trajectoire qui commence à l'instant t et qui est indexée par la durée entre l'instant présent et t . Une variable aléatoire $\sigma(X_u, t \leq u)$ -mesurable s'écrit donc $F(\omega) = \psi(\theta_t \omega)$ avec ψ -mesurable. Soit $\mathfrak{T} \subset \mathbf{R}$ un ensemble d'indices. Soit $(\mathcal{F}_t, t \in \mathfrak{T})$ une filtration de (E, \mathcal{E}) polonais : $t \leq t' \implies \mathcal{F}_t \subset \mathcal{F}_{t'}$.

DÉFINITION A.17.– Une famille de variables aléatoires $X = (X_t, t \in \mathfrak{T})$ à valeurs dans E , est markovienne lorsque les deux conditions suivantes sont vérifiées :

- X_t est \mathcal{F}_t mesurable pour tout $t \in \mathfrak{T}$;
- les tribus \mathcal{F}_t et $\sigma(X_s)$ sont indépendantes conditionnellement à X_t pour tout $t \leq s \in \mathfrak{T}$.

THÉORÈME A.12.– Si X est markovienne alors pour tout $t \in \mathfrak{T}$, les tribus \mathcal{F}_t et $\sigma\{X_s, s \geq t\}$ sont indépendantes conditionnellement à X_t . De plus, pour toute fonction ψ bornée de (E, \mathcal{E}) dans \mathbf{R} , on a l'identité suivante :

$$\mathbf{E}[\psi \circ \theta_t | \mathcal{F}_t] = \mathbf{E}[\psi \circ \theta_t | X_t]. \quad (\text{A.4})$$

⌚ Cette propriété signifie que passé et futur sont indépendants conditionnellement au présent. A l'instant t , pour déterminer la suite de l'évolution de X , seule compte la valeur en t , peu importe comment on y est arrivé.

Démonstration. Par un argument de classe monotone dans le cas où $\mathfrak{T} = \mathbf{R}^+$, par définition dans le cas $\mathfrak{T} = \mathbf{N}$, il faut et il suffit que l'on montre que les tribus \mathcal{F}_t et $\sigma\{X_s, s \in \{t = t_0 < t_1 < \dots < t_n\}\}$ sont indépendantes conditionnellement à X_t .

On sait que pour tout j , $\sigma(X_{t_n})$ et $\mathcal{F}_{t_{n-1}}$ sont indépendantes conditionnellement à X_{t_n} . Or $\mathcal{F}_{t_0} \vee \bigvee_{j=1}^{n-1} \sigma(X_{t_j})$ est une sous-tribu de $\mathcal{F}_{t_{n-1}}$ puisque X_{t_j} est \mathcal{F}_{t_j} (donc $\mathcal{F}_{t_{n-1}}$)-mesurable. D'après le théorème A.11 avec $\mathcal{F} = \sigma(X_{t_n})$ et $\mathcal{G} = \sigma(X_{t_{n-1}})$, les tribus \mathcal{F}_t et $\sigma(X_{t_n})$ sont indépendantes conditionnellement à $\sigma(X_{t_0}, \dots, X_{t_{n-1}})$. En appliquant de nouveau le théorème A.11 avec $\mathcal{F} = \mathcal{F}_t$ et $\mathcal{G} = \sigma(X_{t_n})$, on voit que les tribus \mathcal{F}_t et $\sigma\{X_s, s \in \{t = t_0 < t_1 < \dots < t_n\}\}$ sont indépendantes conditionnellement à X_t .

Pour montrer le deuxième point, remarquons de $\psi \circ \theta_t$ est $\bigvee_{s \geq t} \mathcal{F}_s$ -mesurable. D'autre part, comme $\mathbf{E}[\psi \circ \theta_t | X_t]$ est \mathcal{F}_t -mesurable, il suffit d'établir, pour tout

$F \in \mathcal{F}_t$, l'identité suivante :

$$\mathbf{E} [\mathbf{E} [\psi \circ \theta_t | X_t] \mathbf{1}_F] = \mathbf{E} [\psi \circ \theta_t \mathbf{1}_F].$$

Or, d'après les propriétés de l'espérance conditionnelle et la première partie de la preuve, on a :

$$\begin{aligned} \mathbf{E} [\mathbf{E} [\psi \circ \theta_t | X_t] \mathbf{1}_F] &= \mathbf{E} [\mathbf{E} [\psi \circ \theta_t | X_t] \mathbf{E} [\mathbf{1}_F | X_t]] \\ &= \mathbf{E} [\mathbf{E} [\psi \circ \theta_t \mathbf{1}_F | X_t]] \\ &= \mathbf{E} [\psi \circ \theta_t \mathbf{1}_F], \end{aligned}$$

d'où le résultat. □

A.3. Espaces de vecteurs et ordres

Fixons $p \geq 1$. Dans l'espace euclidien \mathbf{R}^p , pour tout $x, y \in \mathbf{R}^p$ et $\lambda \in \mathbf{R}$, nous notons naturellement :

$$\begin{aligned} x &= (x(1), x(2), \dots, x(p)) \\ \lambda.x &= (\lambda x(1), \dots, \lambda x(p)) \\ x + y &= (x(1) + y(1), \dots, x(p) + y(p)) \end{aligned}$$

On notera par ailleurs dans toute la suite :

- $\mathbf{e}_1 = (1, 0, 0, \dots, 0)$, le premier vecteur de la base canonique de \mathbf{R}^p ;
- $\mathbf{1} = (1, 1, \dots, 1)$;
- $X^+ = (X(1)^+, X(2)^+, \dots, X(p)^+)$ pour tout $X \in \mathbf{R}^p$;
- pour tout $X \in \mathbf{R}^p$, \bar{X} est le vecteur correspondant ordonné par ordre croissant.

On note alors $\overline{(\mathbf{R}^+)^p}$, l'ensemble des vecteurs de $(\mathbf{R}^+)^p$ dont les coordonnées sont rangées par ordre croissant.

DÉFINITION A.18.– Une relation « \leq » sur un ensemble E définit une relation d'ordre si elle est :

- (i) réflexive : pour tout $x \in E$, $x \leq x$;
- (ii) transitive : pour tout $x, y, z \in E$, $x \leq y$ et $y \leq z$ implique $x \leq z$;
- (iii) antisymétrique : pour tout $x, y \in E$, $x \leq y$ et $y \leq x$ implique $y = x$.

La relation d'ordre est alors dite totale si $x \leq y$ ou $y \leq x$ pour tout $x, y \in E$, partielle sinon.

On définit une première relation d'ordre partiel sur \mathbf{R}^p , notée « $<$ », par la relation suivante :

$$X < Y \iff X(i) \leq Y(i) \text{ pour tout } i = 1, \dots, p. \quad (\text{A.5})$$

En particulier, $(\mathbf{R}^+)^p$ admet bien sûr le point $<$ -minimal $\mathbf{0} = (0, \dots, 0)$. Par ailleurs, en raisonnant coordonnée par coordonnée, on vérifie toutes les suites $<$ -croissantes et majorées dans \mathbf{R}^p convergent.

L'ordre Schur-convexe, noté $<_c$ est un autre (quasi) ordre partiel sur \mathbf{R}^p , particulièrement utilisé en économie.

DÉFINITION A.19.– Soit u et v deux vecteurs de \mathbf{R}^p . On dit que $u <_c v$ si :

$$\left\{ \begin{array}{l} \sum_{i=1}^p u(i) = \sum_{i=1}^p v(i), \\ \sum_{i=k}^p \bar{u}(i) \leq \sum_{i=k}^p \bar{v}(i), \quad k = 2, \dots, p. \end{array} \right.$$

↳ On peut donc comparer u et v par l'ordre Schur-convexe si u et v représentent deux répartitions en p composantes de la même masse totale. Alors $u <_c v$ signifie que u répartit une masse plus grande sur un plus grand nombre de composantes que v . La relation « $<_c$ » est presque un ordre partiel dans le sens où $u <_c v$ et $v <_c u$ impliquent que u et v sont égaux à une permutation de leur coordonnées près.

Soit \mathfrak{S}_p , l'ensemble des permutations de $\llbracket 1, p \rrbracket$. Pour tout élément $x \in \mathbf{R}^p$ et tout $\gamma \in \mathfrak{S}_p$, nous notons :

$$x_\gamma = (x(\gamma(1)), x(\gamma(2)), \dots, x(\gamma(p))).$$

DÉFINITION A.20.– Soit E , un ensemble. Une fonction $F : \mathbf{R}^p \rightarrow E$ est dite symétrique si $F(x) = F(x_\gamma)$ pour tout $x \in \mathbf{R}^p$ et $\gamma \in \mathfrak{S}_p$.

– Soit $\gamma \in \mathfrak{S}_p$ et $x \in \mathbf{R}^p$. On dit que γ est une permutation ordonnante de x si x n'est pas totalement ordonné et si $\gamma(i) = j$ et $\gamma(j) = i$ pour un certain couple (i, j) tel que $i < j$ et $x(i) > x(j)$, et $\gamma(k) = k$ pour tout $k \notin \{i, j\}$.

La démonstration suivante est laissée au lecteur.

LEMME A.13.– Pour tout $x, y \in \mathbf{R}^p$, l'équivalence suivante est satisfaite :

$$x <_c y \iff -x <_c -y. \quad (\text{A.6})$$

LEMME A.14.– Pour tout $x, y \in \mathbf{R}^p$, pour toute fonction convexe et symétrique $F : \mathbf{R}^p \rightarrow \mathbf{R}$, l'équivalence suivante est satisfaite :

$$x \prec_c y \iff F(x) \leq F(y). \quad (\text{A.7})$$

Pour tout $x, y \in \mathbf{R}^p$, pour toute permutation γ ordonnant x :

$$x_\gamma - \bar{y} \prec_c x - \bar{y} \text{ en particulier, } \bar{x} - \bar{y} \prec_c x - \bar{y}. \quad (\text{A.8})$$

Nous introduisons maintenant la relation d'ordre suivante sur $\overline{(\mathbf{R}^+)^p}$, qui est une variante de l'ordre Schur-convexe.

DÉFINITION A.21.– Soit u et $v \in \overline{(\mathbf{R}^+)^p}$. On note $u \prec_* v$ si :

$$\sum_{i=k}^p u(i) \leq \sum_{i=k}^p v(i), \text{ pour tout } k \in \llbracket 1, p \rrbracket.$$

On peut vérifier en exercice que « \prec_* » définit bien un ordre partiel sur $\overline{(\mathbf{R}^+)^p}$. Par ailleurs, nous avons le résultat suivant.

LEMME A.15.– Soit $u, v \in \overline{(\mathbf{R}^+)^p}$ tels que $u \prec_* v$:

(i) pour tout $x \in \mathbf{R}$, on a :

$$[u - x.\mathbf{1}]^+ \prec_* [v - x.\mathbf{1}]^+ ;$$

(ii) pour tout $i \in \llbracket 1, p \rrbracket$ tel que $u(j) \leq v(j)$, pour tout $y \in \mathbf{R}^+$, on a :

$$\overline{u + y.e_j} \prec_* \overline{v + y.e_j}.$$

Démonstration. (i) Le résultat est trivial si $u(p) \leq x$. Sinon, pour tout $k \in \llbracket 1, p \rrbracket$ il existe un $\ell \geq k$ tel que :

$$\sum_{i=k}^p (u(i) - x)^+ = \sum_{i=\ell}^p (u(i) - x) \leq \sum_{i=\ell}^p (v(i) - x) \leq \sum_{i=k}^p (v(i) - x)^+.$$

(ii) Pour tout $k > j$, on a :

$$\begin{aligned} \sum_{i=k}^p (\overline{u + y \cdot \mathbf{e}_j})(i) &= \left(\sum_{i=k}^p u(i) \right) \vee \left(u(j) + y + \sum_{i=k+1}^p u(i) \right) \\ &\leq \left(\sum_{i=k}^p v(i) \right) \vee \left(v(j) + y + \sum_{i=k+1}^p v(i) \right) \\ &= \sum_{i=k}^p (\overline{v + y \cdot \mathbf{e}_j})(i), \end{aligned}$$

alors que pour tout $k \leq j$:

$$\sum_{i=k}^p (\overline{u + y \cdot \mathbf{e}_j})(i) = \sum_{i=k}^p u(i) + y \leq \sum_{i=k}^p v(i) + y = \sum_{i=k}^p (\overline{v + y \cdot \mathbf{e}_j})(i).$$

D'où le résultat. \square

Plusieurs SRS étudiées au chapitre 4 prennent des valeurs vectorielles avec un nombre non borné de composantes. Nous introduisons donc l'espace \mathcal{S} des suites positives presque nulles (toutes leurs composantes sont nulles à partir d'un certain rang, et jamais nulles auparavant) :

$$\mathcal{S} := \left\{ u \in (\mathbf{R}^+)^{\mathbf{N}^*}, \exists N(u) \in \mathbf{N}, u(i) = 0 \forall i > N(u) \right. \\ \left. \text{et } u(i) > 0 \forall i \leq N(u) \right\}. \quad (\text{A.9})$$

En d'autres termes, $N(u)$ est le nombre de coordonnées non nulles de u . Pour tout $u \in \mathcal{S}$, nous notons \underline{u} , la version ordonnée de u par ordre décroissant :

$$\underline{u}(i+1) \leq \underline{u}(i) \text{ pour tout } i \in \mathbf{N}.$$

Pour tout $u \in \mathcal{S}$, nous notons \underline{u} , la version ordonnée de u par ordre décroissant :

$$\underline{u}(i+1) \leq \underline{u}(i) \text{ pour tout } i \in \mathbf{N}.$$

Nous élargissons maintenant la définition des ordres partiels « $<$ » et « $<_c$ » à \mathcal{S} .

DÉFINITION A.22.— Soit $u, v \in \mathcal{S}$. Soit \tilde{u} et \tilde{v} , les versions ordonnées des restrictions respectives de u et v à leurs $N(u) \vee N(v)$ premières coordonnées.

(i) On dit que $u <_c v$ si $\tilde{u} <_c \tilde{v}$, autrement dit :

$$u <_c v \iff \begin{cases} \sum_{i \in \mathbf{N}^*} u(i) = \sum_{i \in \mathbf{N}^*} v(i); \\ \sum_{i=j}^{\infty} \underline{u}(i) \geq \sum_{i=j}^{\infty} \underline{v}(i) \text{ pour tout } j \in \mathbf{N}^*. \end{cases}$$

(ii) On dit que $u < v$ si $\tilde{u} < \tilde{v}$, autrement dit :

$$u <_c v \iff \underline{u}(i) < \underline{v}(i) \text{ pour tout } i \in \mathbf{N}^*.$$

Nous définissons de la même façon l'ensemble des suites à valeurs dans $(\mathbf{R}^+)^2$. Tout élément u de $(\mathbf{R}^2)^{\mathbf{N}^*}$ est noté :

$$u := \left((u^1(1), u^2(1)), (u^1(2), u^2(2)), \dots \right).$$

De manière équivalente, on peut écrire u sous la forme :

$$u = (u^1, u^2),$$

où :

$$u^1 = (u^1(1), u^1(2), \dots) \text{ et } u^2 = (u^2(1), u^2(2), \dots)$$

sont deux éléments de $\mathbf{R}^{\mathbf{N}^*}$.

On définit alors, de manière similaire à \mathcal{S} :

$$\mathcal{S}^2 := \left\{ u \in ((\mathbf{R}^+)^2)^{\mathbf{N}^*}, \exists N(u) \in \mathbf{N}, (u^1(i), u^2(i)) = (0, 0) \forall i > N(u) \right. \\ \left. \text{et } u^1(i) > 0, u^2(i) > 0 \forall i < N(u) \text{ et } u^1(N(u)) > 0, u^1(N(u)) = 0 \right\}, \quad (\text{A.10})$$

l'ensemble des suites dont les coefficients sont égaux à $(0, 0)$ à partir d'un certain rang $N(u) + 1$, à valeurs dans $(\mathbf{R}_+^*)^2$ jusqu'au rang $N(u) - 1$ et de première composante strictement positive au rang $N(u)$ (la deuxième étant nulle).

A.4. Processus à variation finie

En calcul stochastique, apparaissent de manière récurrente des fonctions dites à « variation finie ». Ces fonctions se révèlent être des différences de deux fonctions monotones croissantes et les fonctions croissantes jouissent de propriétés de différentiabilité intéressantes. Cela permet d'écrire une formule d'intégration par parties et une formule de changement de variables pour les fonctions à variation finie qui sont d'un usage constant en probabilités.

DÉFINITION A.23.— Soit $[a, b] \subset \mathbf{R}$, on appelle subdivision de $[a, b]$, un ensemble fini de points t_0, \dots, t_n tel que :

$$a = t_0 < t_1 < \dots < t_n = b.$$

On note $|\pi|$, le pas de la subdivision défini par $|\pi| = \sup_{t_i \in \pi} |t_{i+1} - t_i|$. On note $\Pi_{[a, b]}$ l'ensemble des subdivisions de $[a, b]$.

DÉFINITION A.24.— Soit $f : [a, b] \rightarrow \mathbf{R}$, f est à variation finie lorsque :

$$\sup_{\pi \in \Pi_{[a, b]}} \sum_{t_i \in \pi} |f(t_{i+1}) - f(t_i)| \text{ est fini.}$$

On note $\text{Var}_{[a, b]}(f)$, cette borne supérieure.

REMARQUE.— Les fonctions croissantes et les fonctions lipschitziennes sont à variation finie.

THÉORÈME A.16.— Si $[c, d] \subset [a, b]$ et f est à variation finie sur $[a, b]$ alors f est à variation finie sur $[c, d]$. De plus :

$$\text{Var}_{[a, c]}(f) + \text{Var}_{[c, b]}(f) = \text{Var}_{[a, b]}(f).$$

Démonstration. Laissez au lecteur. □

THÉORÈME A.17 (Décomposition de Jordan).— Soit f à variation finie sur $[a, b]$, il existe un unique couple de fonctions (g, h) telles que :

- $f = g - h + f(a)$;
- g et h croissantes ;
- $g(a) = h(a) = 0$;
- $\text{Var}_{[a, b]}(f) = \text{Var}_{[a, b]}(g) + \text{Var}_{[a, b]}(h)$.

Démonstration. On pose :

$$g(x) = \frac{1}{2} (f(x) - f(a) + \text{Var}_{[a, x]}(f)), \text{ et } h(x) = \frac{1}{2} (f(a) - f(x) + \text{Var}_{[a, x]}(f)).$$

Il est immédiat que :

$$g(a) = h(a) = 0, f = g - h + f(a).$$

Comme :

$$|f(x) - f(y)| \leq \text{Var}_{[x, y]}(f) = \text{Var}_{[a, y]}(f) - \text{Var}_{[a, x]}(f),$$

g et h sont croissantes. Par conséquent :

$$\text{Var}_{[a,b]}(g) = g(b) - g(a) \text{ et } \text{Var}_{[a,b]}(h) = h(b) - h(a),$$

d'où : $\text{Var}_{[a,b]}(f) = \text{Var}_{[a,b]}(g) + \text{Var}_{[a,b]}(h)$.

Il reste à montrer l'unicité. On suppose qu'il existe une autre paire (g_1, h_1) que celle construite. Soit $x < y$, g_1 est croissante donc $g_1(y) - g_1(x) \geq 0$. D'autre part :

$$g_1(y) - g_1(x) = f(y) - f(x) + h_1(y) - h_1(x) \geq f(y) - f(x),$$

puisque h_1 est croissante. On peut donc dire que :

$$\begin{aligned} g_1(y) - g_1(x) &\geq \max(0, f(y) - f(x)) \\ &= \frac{1}{2}(f(y) - f(x) + |f(y) - f(x)|). \end{aligned}$$

Pour toute subdivision de $[x, y]$, on a alors :

$$\begin{aligned} g_1(y) - g_1(x) &\geq \frac{1}{2}(f(y) - f(x) + \sum_i |f(t_{i+1}) - f(t_i)|) \\ &= \frac{1}{2}(f(y) - f(x) + \text{Var}_{[x,y]}(f)) \\ &= \frac{1}{2}(f(y) - f(x) + \text{Var}_{[a,y]}(f) - \text{Var}_{[a,x]}(f)) \\ &= g(y) - g(x). \end{aligned}$$

La fonction $\beta \equiv g_1 - g$ est donc croissante. Par ailleurs, la relation :

$$f \equiv g - h + f(a) \equiv g_1 - h_1 + f(a)$$

implique que $h_1 \equiv h + \beta$ et $\beta(a) = 0$. Enfin, la contrainte :

$$\text{Var}_{[a,b]}(f) = \text{Var}_{[a,b]}(g_1) + \text{Var}_{[a,b]}(h_1),$$

implique que :

$$g_1(b) - g_1(a) + h_1(b) - h_1(a) = g(b) + h(b).$$

Or $g_1(b) - g_1(a) + h_1(b) - h_1(a) = g(b) + h(b) + 2\beta(b)$. β est donc la fonction nulle, ce qui signifie que $g \equiv g_1$ et $h \equiv h_1$. \square

On s'intéresse maintenant aux propriétés de différentiabilité des fonctions monotones. Le théorème « massue » est le théorème de Lebesgue.

THÉORÈME A.18 (Théorème de différentiation de Lebesgue).— Soit $[a, b]$ un intervalle compact de \mathbf{R} et f une fonction croissante sur $[a, b]$. La fonction f est dérivable presque sûrement (au sens de la mesure de Lebesgue) sur $]a, b[$, f' est une fonction intégrable et :

$$\int_a^b f'(t) \, dt \leq f(b) - f(a). \quad (\text{A.11})$$

REMARQUE.— La dérivée est définie au sens usuel comme limite des taux d'accroissement mais le théorème usuel qui dit que $f(b) - f(a) = \int_a^b f'(t) \, dt$ n'est plus vrai. Pour expliquer d'où vient ce déficit apparent, faisons appel à une version modifiée du théorème de Radon-Nikodym.

THÉORÈME A.19.— Soit μ et ν deux mesures positives σ -finies. Il existe deux mesures ν_a et ν_d telles que :

$$- \nu \equiv \nu_a + \nu_d;$$

- ν_a est absolument continue par rapport à μ , c'est-à-dire qu'il existe $h \in L^1(\mu)$ telle que :

$$\nu_a(B) = \int_B h \, d\mu;$$

- ν_d est étrangère à μ , c'est-à-dire qu'il existe N mesurable tel que $\nu_d(N^c) = 0$ et $\mu(N) = 0$; autrement dit, le support de ν_d est inclus dans un ensemble μ -négligeable.

Le deuxième ingrédient est l'intégrale de Stieltjes.

DÉFINITION A.25.— Pour $g \in \mathcal{C}_c(\mathbf{R})$ et f croissante, continue à droite, on définit l'intégrale de g par rapport à f , notée $\int g \, d f$, comme la limite si elle existe :

$$\sum_i g(t_i)(f(t_{i+1}) - f(t_i)),$$

quand le pas de la subdivision, c'est-à-dire $\sup_i |t_{i+1} - t_i|$ tend vers 0.

REMARQUE.— L'existence de la limite est assurée sous l'hypothèse que f et g n'ont pas les mêmes points de discontinuité.

Enfin, le théorème de Riesz assure alors qu'il existe une mesure λ_f (de Radon, c'est-à-dire finie sur les compacts) telle que :

$$\int g \, d f = \int g \, d \lambda_f \text{ pour tout } g \in \mathcal{C}_c(\mathbf{R}).$$

THÉORÈME A.20.— Pour f et λ_f ainsi définies, on a :

$$\lambda_f(]a, b]) = f(b) - f(a) \text{ et } \lambda_f([a, b]) = f(b) - f(a^-),$$

où $f(x^-) = \lim_{y \uparrow x} f(y)$ et $f(x_+) = \lim_{y \downarrow x} f(y) = f(x)$ par continuité à droite de f .

Démonstration. On considère la suite de fonctions ($h_n, n \geq 1$) (voir figure A.3) définie par :

$$\begin{aligned} h_n(x) &= 1 \text{ pour } a + \frac{1}{n} \leq x \leq b, \\ h_n(x) &= n(x - a) \text{ pour } a \leq x \leq a + \frac{1}{n}, \\ h_n(x) &= 1 - n(x - b) \text{ pour } b \leq x \leq b + \frac{1}{n}. \end{aligned}$$

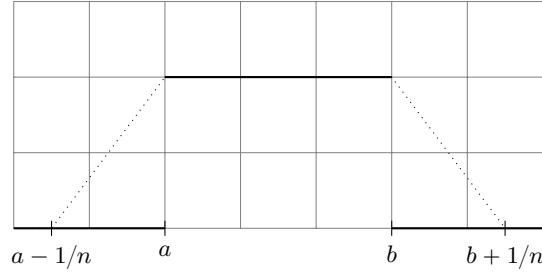


Figure A.3 – Fonction plateau approchant $\mathbf{1}_{]a,b]}$

h_n tend simplement vers $\mathbf{1}_{]a,b]}$ et $|h_n| \leq 1$ donc par convergence dominée :

$$\int h_n d\lambda_f \rightarrow \lambda_f(]a, b]).$$

En prenant comme subdivision $\{t_0 = a, t_1 = a + 1/n, t_2 = b, t_3 = b + 1/n\}$, on obtient :

$$(f(a + \frac{1}{n}) - f(a)) + (f(b) - f(a + \frac{1}{n})) + (f(b + \frac{1}{n}) - f(b)) \leq \int h_n df.$$

Comme $0 \leq h_n \leq 1$, pour toute subdivision, on a :

$$\begin{aligned} \sum h_n(t_i)(f(t_{i+1}) - f(t_i)) &\leq (f(a + \frac{1}{n} + |\pi|) - f(a)) \\ &\quad + (f(b + |\pi|) - f(a + \frac{1}{n})) + (f(b + \frac{1}{n}) - f(b)), \end{aligned}$$

donc :

$$\int h_n \, d f \leq f(b + |\pi|) - f(b) + f(b + \frac{1}{n}) - f(a).$$

Les h_n sont positives croissantes donc par convergence monotone, $\int h_n \, d f$ converge vers $\int h \, d f$ d'une part et d'autre part, les encadrements que nous venons d'obtenir montrent, compte tenu de la continuité à droite de f , que :

$$\lim_{n \rightarrow \infty} \int h_n \, d f = f(b) - f(a).$$

On a donc bien montré que $f(b) - f(a) = \lambda_f([a, b])$. L'autre identité se démontre de manière similaire en prenant une suite de fonctions qui vaut 1 sur $[a, b]$ et qui s'annule en deçà de $a - 1/n$. \square

En combinant les deux derniers théorèmes, on obtient la description suivante.

THÉORÈME A.21.— Soit f croissante, continue à droite, nulle en $-\infty$, il existe h localement intégrable et ν_f une mesure singulière par rapport à la mesure de Lebesgue telles que :

$$f(x) = \int_{-\infty}^x h(u) \, d u + \nu_f([-\infty, x]).$$

REMARQUE.— La partie dérivable n'épuise donc pas toute la masse d'où la relation (A.11). D'autre part, comme h est localement intégrable, le théorème de convergence dominée assure que la partie $\int_{-\infty}^x h(u) \, d u$ est une fonction continue de la variable x donc si f est discontinue en x , on a nécessairement $\Delta f(x) = \nu_f(\{x\})$.

THÉORÈME A.22 (Intégration par parties).— Soit f et g deux fonctions à variation finie et continues à droite sur $[0, t[$, alors :

$$f(t)g(t) - f(0)g(0) = \int_0^t f(s^-) \, d g(s) + \int_0^t g(s^-) \, d f(s) + [f, g]_t, \quad (\text{A.12})$$

où :

$$[f, g]_t = \sum_{0 \leq s \leq t} \Delta f(s) \Delta g(s) \text{ et } \Delta f(s) = f(s) - f(s^-).$$

Démonstration. On calcule l'intégrale de $d \lambda_f \otimes d \lambda_g$ sur le carré $[0, t] \times [0, t]$ de deux manières différentes. D'une part, il vient du théorème A.20 que :

$$\int_{[0, t]^2} d \lambda_f \otimes d \lambda_g = (f(t) - f(0))(g(t) - g(0)).$$

D'autre part, on décompose le carré en deux triangles (inférieur et supérieur : I et S , respectivement) et la diagonale D , voir figure A.4.

On applique aux deux triangles le théorème de Fubini. Pour le cas du triangle I , on obtient :

$$\begin{aligned} \int_I d\lambda_f \otimes d\lambda_g &= \int_0^t \left(\int_0^{s^-} d\lambda_g \right) d\lambda_f(s) \\ &= \int_0^t (g(s^-) - g(0)) d\lambda_f(s) \\ &= \int_0^t g(s^-) d\lambda_f(s) - g(0)(f(t) - f(0)). \end{aligned}$$

De même :

$$\int_S d\lambda_f \otimes d\lambda_g = \int_0^t f(s^-) d\lambda_g(s) - f(0)(g(t) - g(0)).$$

La diagonale D est de mesure de Lebesgue nulle donc l'intégrale sur D de $d\lambda_f \otimes d\lambda_g$ se réduit à l'intégrale du produit de leur partie singulière :

$$\int_D d\lambda_f \otimes d\lambda_g = \int_D d\nu_f \otimes d\nu_g = \sum_{0 \leq s \leq t} \Delta f(s) \Delta g(s),$$

en vertu de la remarque A.4. En sommant les trois égalités, on obtient (A.12). \square

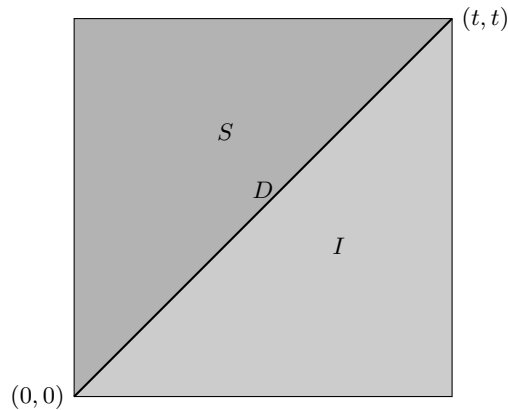


Figure A.4 – Décomposition de $[0, t]^2$

Le dernier théorème est un théorème de changement de variable. En dimension 1, le théorème de changement variable usuel stipule que :

$$F(g(t)) - F(g(0)) = \int_0^t d(F \circ g)(s) = \int_0^t F'(g(s))g'(s) \, ds. \quad (\text{A.13})$$

C'est en fait l'application de la relation $f(t) - f(0) = \int_0^t f'(s) \, ds$ à l'identité $(F \circ g)' = F' \circ g \cdot g'$. Dans le cas des fonctions à variation finie, on n'a pas la première relation et la deuxième paraît difficile à vérifier. Néanmoins, on obtient un résultat similaire à (A.13).

THÉORÈME A.23.— Soit g une fonction à variation finie, continue à droite et F une fonction de classe \mathcal{C}^1 , on a :

$$\begin{aligned} F(g(t)) - F(g(0)) &= \int_0^t F'(g(s^-))dg(s) \\ &+ \sum_{0 \leq s \leq t} F(g(s)) - F(g(s^-)) - F'(g(s^-))\Delta g(s). \end{aligned} \quad (\text{A.14})$$

REMARQUE.— En particulier, si g est continue on obtient le résultat usuel.

Démonstration. On démontre (A.14) pour $F(x) = x^n$, par récurrence sur n à partir de la formule d'intégration par parties. L'expression étant linéaire en F , (A.14) est donc vraie pour les polynômes. On approche toute fonction \mathcal{C}^1 par une suite de polynômes au sens de la norme uniforme sur $[0, t]$, les passages à la limite sont possibles, on en déduit le résultat pour F de classe \mathcal{C}^1 . \square

A.5. Martingales

A.5.1. Martingales à temps discret

DÉFINITION A.26.— Une suite $(X_n, n \geq 0)$ de variables aléatoires réelles, intégrables s'appelle une martingale (respectivement une sous-martingale ou une sur-martingale) si :

- pour tout n , X_n est \mathcal{F}_n -mesurable et intégrable ;
- $\mathbf{E}[X_{n+1} | \mathcal{F}_n] = X_n$ presque sûrement (respectivement $\mathbf{E}[X_{n+1} | \mathcal{F}_n] \geq X_n$ ou $\mathbf{E}[X_{n+1} | \mathcal{F}_n] \leq X_n$).

EXEMPLE.— L'exemple typique est celui d'une suite de variables aléatoires indépendantes. Soit $(\xi_i, i \geq 1)$ sont des variables aléatoires i.i.d., avec $\mathbf{E}[\xi_i] = 0$, alors la

suite définie par $X_0 = 0$, $X_n = \sum_{i=1}^n \xi_i$, est une $\mathcal{F}_n = \sigma(\xi_0, \dots, \xi_n)$ martingale, car :

$$\mathbf{E} \left[\sum_{i=1}^{n+1} \xi_i \mid \mathcal{F}_n \right] = \mathbf{E} [X_n + \xi_{n+1} \mid \mathcal{F}_n] = X_n + \mathbf{E} [\xi_{n+1}] = X_n,$$

car ξ_{n+1} est indépendante de \mathcal{F}_n .

DÉFINITION A.27.— Soit τ une variable aléatoire à valeurs entières. τ est un temps d'arrêt lorsque pour tout $n \in \mathbf{N}$:

$$(\omega : \tau(\omega) = n) \in \mathcal{F}_n.$$

THÉORÈME A.24.— Soit $(X_n, n \geq 0)$ une martingale et τ un temps d'arrêt. Alors $(X_n^\tau, n \geq 1)$, où X_n^τ est définie par $X_{\tau \wedge n}$, est une martingale.

Démonstration. Nous avons :

$$X_{\tau \wedge n} = \sum_{m=0}^{n-1} X_m \mathbf{1}_{\{\tau=m\}} + X_n \mathbf{1}_{\{\tau \geq n\}},$$

par conséquent $X_{\tau \wedge n}$ est \mathcal{F}_n -mesurable. En ce qui concerne l'intégrabilité de $X_{\tau \wedge n}$, comme $|X_n|$ est une sous-martingale, pour $m \leq n$, on a :

$$\mathbf{E} [|X_m| Y_m] \leq \mathbf{E} [|X_n| Y_m]$$

pour toute variable aléatoire Y_m \mathcal{F}_m mesurable positive. Par conséquent :

$$\begin{aligned} \mathbf{E} [|X_{\tau \wedge n}|] &\leq \sum_{m=0}^{n-1} \mathbf{E} [|X_m| \mathbf{1}_{\{\tau=m\}}] + \mathbf{E} [|X_n| \mathbf{1}_{\{\tau \geq n\}}] \\ &\leq \sum_{m=0}^{n-1} \mathbf{E} [|X_n| \mathbf{1}_{\{\tau=m\}}] + \mathbf{E} [|X_n| \mathbf{1}_{\{\tau \geq n\}}] \leq \mathbf{E} [|X_n|] < +\infty. \end{aligned}$$

De plus :

$$\begin{aligned} X_{\tau \wedge (n+1)} - X_{\tau \wedge n} &= X_{n+1} \mathbf{1}_{\{\tau \geq n+1\}} - X_n \mathbf{1}_{\{\tau > n\}} \\ &= \mathbf{1}_{\{\tau > n\}} (X_{n+1} - X_n). \end{aligned}$$

Donc :

$$\mathbf{E} [X_{\tau \wedge (n+1)} - X_{\tau \wedge n} \mid \mathcal{F}_n] = \mathbf{1}_{\{\tau > n\}} \mathbf{E} [X_{n+1} - X_n \mid \mathcal{F}_n] = 0.$$

Le résultat s'ensuit. □

DÉFINITION A.28.– Si τ est un temps d'arrêt, on note par \mathcal{F}_τ la tribu définie par :

$$A \in \mathcal{F}_\tau \iff A \cap (\tau \leq n) \in \mathcal{F}_n,$$

pour tout $n \geq 0$.

LEMME A.25.– Si τ un temps d'arrêt et X une variable aléatoire \mathcal{F}_τ -mesurable alors la variable aléatoire :

$$X \mathbf{1}_{\{\tau \leq n\}}$$

est \mathcal{F}_n -mesurable pour tout $n \in \mathbf{N}$.

Démonstration. Soit $A \in \mathcal{F}_\tau$ et $X = \mathbf{1}_A$. Alors la conclusion est vraie trivialement. Par linéarité, elle est aussi vraie pour des variables aléatoires étagées. Si X est quelconque, alors il existe une suite de variables aléatoires étagées $(X_k, k \geq 1)$ \mathcal{F}_τ -mesurables, qui converge vers X presque sûrement. Par conséquent, on a :

$$X \mathbf{1}_{\{\tau \leq n\}} = \lim_{k \rightarrow \infty} X_k \mathbf{1}_{\{\tau \leq n\}}$$

appartient à \mathcal{F}_n . □

THÉORÈME A.26.– Soit $(X_n, n \geq 0)$ une martingale (respectivement une sous-martingale) et $\tau_1 \leq \tau_2$ deux temps d'arrêt tels que pour $i = 1, 2$, on ait :

$$\mathbf{E}[|X_{\tau_i}|] < \infty \text{ et } \liminf_{n \rightarrow \infty} \int_{\{\tau_i > n\}} |X_n| d\mathbf{P} = 0.$$

L'identité suivante est alors vérifiée :

$$\mathbf{E}[X_{\tau_2} | \mathcal{F}_{\tau_1}] = X_{\tau_1}, \mathbf{P} \text{ presque sûrement.}$$

Démonstration. Supposons d'abord que $\tau_2 \leq k$ où $k \in \mathbf{N}$ est une constante. Alors nous avons $|X_{\tau_2}| \leq |X_1| + \dots + |X_k|$, par conséquent X_{τ_2} est intégrable. Pour $n \leq k$ et $A \in \mathcal{F}_{\tau_1}$, nous avons :

$$\mathbf{E}[X_k \mathbf{1}_A] = \sum_{j=0}^k \mathbf{E}[X_k \mathbf{1}_{A \cap \{\tau_1=j\}}] = \sum_{j=0}^k \mathbf{E}[X_j \mathbf{1}_{A \cap \{\tau_1=j\}}] = \mathbf{E}[X_{\tau_1} \mathbf{1}_A].$$

Par conséquent, nous avons :

$$\mathbf{E}[X_k | \mathcal{F}_{\tau_1}] = X_{\tau_1}.$$

Définissons la martingale $X_m^{\tau_2}$ par $X_{\tau_2 \wedge m}$. Le raisonnement ci-dessus appliqué à $X_k^{\tau_2} = X_{\tau_2 \wedge k} = X_{\tau_2}$ implique que :

$$\mathbf{E}[X_{\tau_2} | \mathcal{F}_{\tau_1}] = X_{\tau_1}, \mathbf{P} \text{ presque sûrement.}$$

Pour le cas général, d'après ce que nous venons de voir, pour tout $m \geq n$, nous avons :

$$\mathbf{E}[X_{\tau_2 \wedge m} | \mathcal{F}_{\tau_1 \wedge n}] = X_{\tau_1 \wedge n}.$$

Pour $B \in \mathcal{F}_{\tau_1}$, nous avons donc :

$$\mathbf{E}[X_{\tau_2 \wedge m} \mathbf{1}_{B \cap \{\tau_1 \leq m\}}] = \mathbf{E}[X_{\tau_1 \wedge m} \mathbf{1}_{B \cap \{\tau_1 \leq m\}}].$$

Par convergence, le terme de droite de cette égalité converge vers : $\mathbf{E}[X_{\tau_1} \mathbf{1}_B]$. Par conséquent le terme de gauche converge aussi. De plus nous pouvons le décomposer comme :

$$\mathbf{E}[X_{\tau_2 \wedge m} \mathbf{1}_{B \cap \{\tau_1 \leq m\}}] = \mathbf{E}[X_{\tau_2} \mathbf{1}_{B \cap \{\tau_2 \leq m, \tau_1 \leq m\}}] + \mathbf{E}[X_m \mathbf{1}_{B \cap \{\tau_1 \leq m\} \cap \{\tau_2 > m\}}].$$

Quand $m \rightarrow \infty$, le premier terme à droite converge (convergence dominée), donc le second terme doit aussi converger. Dans ce cas nous pouvons remplacer la limite de ce terme avec la \liminf qui est nulle d'après l'hypothèse. \square

DÉFINITION A.29.— Un processus $(A_n, n \geq 0)$ est dit prévisible lorsque A_n est \mathcal{F}_{n-1} mesurable pour tout $n \geq 1$.

THÉORÈME A.27 (Décomposition de Doob).— Soit $(X_n, n \geq 0)$ une sous-martingale. Il existe une martingale $(M_n, n \geq 0)$ et une suite $(A_n, n \geq 0)$ prévisible et croissante telles que :

$$X_n = M_n + A_n, \quad A_0 = 0.$$

De plus, cette décomposition est unique.

Démonstration. Supposons qu'il y ait deux décompositions comme ci-dessus, notées (M, A) et (M', A') . Alors $(A_n - A'_n, n \geq 0)$ est une martingale prévisible, donc elle est constante, par hypothèse nous avons $A_n = A'_n$, d'où l'unicité.

On définit la suite A_n par ses incréments :

$$\Delta A_n = A_{n+1} - A_n = \mathbf{E}[X_{n+1} - X_n | \mathcal{F}_n] \text{ et } A_0 = 0.$$

Comme X est une sous-martingale, ΔA_n est positif et donc A_n est positif croissant. Par définition de l'espérance conditionnelle, ΔA_n est \mathcal{F}_n mesurable donc A_{n+1} aussi et ainsi A est clairement un processus prévisible. Il découle de la chaîne d'égalités :

$$0 < \mathbf{E}[A_n] = \sum_{p=1}^{n-1} \mathbf{E}[\mathbf{E}[\Delta X_p | \mathcal{F}_p]] = \mathbf{E}[X_n - X_0] \leq \mathbf{E}[|X_n| + |X_0|],$$

que A est intégrable. Enfin, on définit M_n par $M_n = X_n - A_n$. L'intégrabilité de M s'ensuit de celles de A et X . De plus, un calcul simple :

$$\mathbf{E}[\Delta M_n | \mathcal{F}_n] = \mathbf{E}[\Delta X_n | \mathcal{F}_n] - \mathbf{E}[\Delta A_n | \mathcal{F}_n] = 0,$$

montre que M_n est une martingale. \square

Soit a et b deux réels tels que $a < b$, on définit les temps d'arrêt suivants :

$$\begin{aligned} \tau_1 &= \inf\{k > 0, X_k \leq a\}, \\ \tau_2 &= \inf\{k > \tau_1, X_k \geq b\}, \\ &\vdots \\ \tau_{2m-1} &= \inf\{k > \tau_{2m-2}, X_k \leq a\}, \\ \tau_{2m} &= \inf\{k > \tau_{2m-1}, X_k \geq b\}, \end{aligned}$$

où l'une quelconque de ces variables est infinie dès que l'ensemble sur lequel est calculé l'indice minimum est vide. Pour un entier $n \in \mathbf{N}$ fixe, on considère :

$$\beta_n([a, b]) = \begin{cases} 0 & \text{si } \tau_2 > n \\ \max\{m, \tau_{2m} \leq n\} & \text{sinon.} \end{cases}$$

C'est-à-dire que β_n est le nombre de traversées en montant de $[a, b]$, avant l'instant n par X , voir la figure A.5.

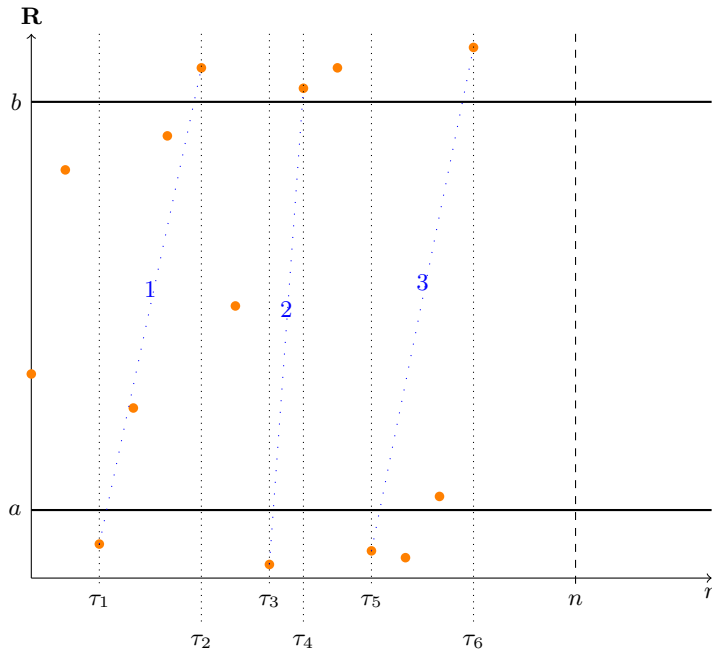


Figure A.5 – Traversées en montant d'une martingale

LEMME A.28.– Soit (X_n) une sous-martingale, pour tout $n \geq 1$, l'identité suivante est satisfaite :

$$\mathbf{E} [\beta_n([a, b])] \leq \frac{1}{b-a} \mathbf{E} [(X_n - a)^+].$$

Démonstration. Le nombre de traversées de X à travers $[a, b]$ est le même que celui de $(X_n - a)^+$ à travers $[0, b - a]$. On suppose donc que $(X_n, n \geq 0)$ est une sous-martingale positive, que $a = 0$ et l'on veut montrer que :

$$\mathbf{E} [\beta_n([0, b])] \leq \frac{\mathbf{E} [X_n]}{b}. \quad (\text{A.15})$$

Posons :

$$\phi_i = \begin{cases} 1 & \text{si } \tau_{2k+1} < i \leq \tau_{2k+2} \\ 0 & \text{si } \tau_{2k+2} < i \leq \tau_{2k+3}. \end{cases}$$

On voit facilement que :

$$\beta_n([0, b]) \leq \sum_{i=1}^n \frac{1}{b} (X_i - X_{i-1}) \phi_i$$

et :

$$\{\phi_i = 1\} = \bigcup_{m \text{ pairs}} [\{\tau_m < i\} \setminus \{\tau_{m+1} < i\}] \in \mathcal{F}_{i-1}.$$

Par conséquent, on a :

$$\begin{aligned} b \mathbf{E} [\beta_n([0, b])] &\leq \sum_{i=1}^n \mathbf{E} [\mathbf{1}_{\{\phi_i=1\}} (X_i - X_{i-1})] \\ &\leq \sum_{i=1}^n \mathbf{E} [\mathbf{1}_{\{\phi_i=1\}} (\mathbf{E} [X_i | \mathcal{F}_{i-1}] - X_{i-1})] \\ &\leq \sum_{i=1}^n \mathbf{E} [\mathbf{E} [X_i - X_{i-1} | \mathcal{F}_{i-1}]] \end{aligned}$$

car $\mathbf{E} [X_i | \mathcal{F}_{i-1}] - X_{i-1} \geq 0$. Puisque $\mathbf{E} [X_0] \geq 0$, on a :

$$b \mathbf{E} [\beta_n([0, b])] \leq \mathbf{E} [X_n] - \mathbf{E} [X_0] \leq \mathbf{E} [X_n].$$

D'où le résultat. □

THÉORÈME A.29.– Soit X une sous-martingale telle que $\sup_n \mathbf{E} [X_n^+] < +\infty$. Le processus $(X_n, n \geq 0)$ converge presque sûrement vers une variable aléatoire X_∞ telle que $\mathbf{E} [|X_\infty|] < +\infty$.

Démonstration. Si X_n ne converge pas presque sûrement alors :

$$\mathbf{P}(\limsup X_n > \liminf X_n) > 0,$$

or :

$$(\limsup X_n > \liminf X_n) = \bigcup_{a,b \in \mathbf{Q}} (\limsup X_n > b > a > \liminf X_n).$$

Il existe donc un couple de rationnels (a, b) tels que :

$$\mathbf{P}(\limsup X_n > b > a > \liminf X_n) > 0.$$

Par conséquent, il existe un mesurable A de probabilité strictement positive tel que :

$$\forall \omega \in A, \limsup X_n > b \text{ et } \liminf X_n < a,$$

ce qui signifie que pour $\omega \in A$, $\lim_{n \rightarrow +\infty} \beta_n([a, b]) = +\infty$ donc, d'après le théorème de convergence monotone :

$$\lim_{n \rightarrow +\infty} \mathbf{E}[\beta_n([a, b])] = \mathbf{E} \left[\lim_{n \rightarrow +\infty} \beta_n([a, b]) \right] = \infty.$$

Mais, on sait par ailleurs que :

$$\begin{aligned} \mathbf{E}[\beta_n([a, b])] &\leq \frac{1}{b-a} \mathbf{E}[(X_n - a)^+] \\ &\leq \frac{1}{b-a} (\mathbf{E}[X_n^+] + |a|) \\ &\leq \frac{1}{b-a} (\sup_n \mathbf{E}[X_n^+] + |a|) < \infty. \end{aligned}$$

On aboutit donc à une contradiction en faisant tendre n vers l'infini. En conclusion, $(X_n, n \geq 1)$ converge presque sûrement vers une variable aléatoire notée X_∞ . De plus :

$$\begin{aligned} \mathbf{E}[|X_n|] &= \mathbf{E}[X_n^+] + \mathbf{E}[X_n^-] = \mathbf{E}[X_n^+] + \mathbf{E}[X_n^+ - X_n] \\ &= 2\mathbf{E}[X_n^+] - \mathbf{E}[X_n] \leq 2\mathbf{E}[X_n^+] - \mathbf{E}[X_1], \end{aligned}$$

car $(X_n, n \geq 0)$ est une sous-martingale. On en déduit que $\sup_n \mathbf{E}[|X_n|]$ est fini, ce qui par le lemme de Fatou permet de conclure que :

$$\mathbf{E}[|X_\infty|] = \mathbf{E}[\liminf X_n] \leq \liminf \mathbf{E}[|X_n|] \leq \sup_n \mathbf{E}[|X_n|] < \infty.$$

La variable aléatoire limite X_∞ est donc intégrable. \square

On tire de ce théorème un corollaire fort utile pour les applications,

COROLLAIRE A.30.– Une sur-martingale positive converge presque sûrement.

Démonstration. Si X est une sur-martingale positive alors $-X$ est une sous-martingale de partie positive nulle, c'est-à-dire que $(-X_n)^+ = 0$, donc on peut appliquer le théorème précédent à $-X$. \square

A.5.2. Martingales à temps continu

DÉFINITION A.30.– Une filtration $(\mathcal{F}_t, t \in \mathbf{R}^+)$ est une famille croissante de tribus. Elle est dite continue à droite lorsque :

$$\bigcap_{s>t} \mathcal{F}_s = \mathcal{F}_t, \text{ pour tout } t \in \mathbf{R}.$$

Elle est dite complète lorsque tous les ensembles négligeables appartiennent à \mathcal{F}_0 .

On suppose dans toute la suite que toutes les filtrations rencontrées sont complètes et continues à droite.

DÉFINITION A.31.– Soit $(\Omega, \mathcal{F} = (\mathcal{F}_t, t \geq 0), \mathbf{P})$ un espace filtré. Un processus $M = (M(t), t \geq 0)$ est une \mathcal{F} -martingale (respectivement sous-martingale, sur-martingale) lorsque pour tout $0 \leq s \leq t$, $M(t) \in L^1(\mathbf{P})$ et :

$$\mathbf{E}[M(t) | \mathcal{F}_s] = M(s), \tag{A.16}$$

respectivement $\geq M(s)$ et $\leq M(s)$.

On admettra que toutes les martingales admettent une version à trajectoire cadlag. Il y a deux grands types de martingales : les continues et les purement discontinues. Les martingales continues dont l'archétype est le mouvement brownien ne sont pas à variation finie. Les martingales qui nous intéresseront sont des martingales à variation finie donc discontinues.

EXEMPLE A.3 (Processus de Poisson).– Soit N un processus de Poisson sur \mathbf{R}^+ d'intensité μ . Le processus $M(t) = N(t) - \mu([0, t])$ est une martingale pour la filtration engendrée par les trajectoires de N : $\mathcal{F}_t = \sigma(N(u), u \leq t)$. En effet, pour tout processus de Poisson (voir chapitre 6), $N(t) - N(s)$ est indépendant de \mathcal{F}_s donc :

$$\mathbf{E}[N(t) - N(s) | \mathcal{F}_s] = \mathbf{E}[N(t) - N(s)] = \mu(]s, t]),$$

Le résultat en découle.

La notion de temps d'arrêt nécessite une légère adaptation par rapport à la définition pour les processus à temps discret.

DÉFINITION A.32.– Une variable aléatoire τ à valeurs dans $\mathbf{R}^+ \cup \{\infty\}$ est un \mathcal{F} -temps d'arrêt lorsque pour tout $t \geq 0$, l'événement $(\tau \leq t)$ appartient à \mathcal{F}_t .

La tribu \mathcal{F}_τ est la tribu des événements A de \mathcal{F}_∞ tels que pour tout $t \geq 0$, $A \cap (\tau \leq t)$ appartient à \mathcal{F}_t .

Les théorèmes d'arrêt et de convergence mentionnés dans la partie sur les martingales discrètes se perpétuent sans changement pour les martingales indexées par \mathbf{R}^+ . En particulier, si M est une martingale et T un temps d'arrêt, le processus :

$$M^T = \{M(t \wedge T), t \geq 0\},$$

est une martingale. La propriété de martingale est souvent vérifiée formellement mais il se peut que les variables aléatoires manipulées ne soient pas intégrables. Pour circonvier ce problème on introduit la notion de martingale locale.

DÉFINITION A.33.– Une \mathcal{F} -martingale M est dite fermée s'il existe $M_\infty \in L^1$ telle que pour tout $t \geq 0$, on ait :

$$M(t) = \mathbf{E}[M_\infty | \mathcal{F}_t].$$

DÉFINITION A.34.– Un processus adapté à trajectoire cadlag est une martingale locale s'il existe une suite croissante de temps d'arrêt $(T_n, n \geq 1)$ tendant presque sûrement vers l'infini tels que pour tout n , M^{T_n} soit une martingale fermée. On dit que la suite de temps d'arrêt $(T_n, n \geq 1)$ réduit M .

THÉORÈME A.31.– Soit M une martingale locale. S'il existe $Z \in L^1$ telle que $M(t) \leq Z$ pour tout $t \geq 0$ alors M est une martingale.

Démonstration. Soit $(T_n, n \geq 1)$ une suite de temps d'arrêt qui réduit M . Pour $0 \leq s \leq t$, on a :

$$M(s \wedge T_n) = \mathbf{E}[M(t \wedge T_n) | \mathcal{F}_s].$$

Comme T_n tend vers l'infini, $s \wedge T_n = s$ à partir d'un certain rang (dépendant de ω) donc presque sûrement, $M(s \wedge T_n)$ tend vers $M(s)$. Par convergence dominée, on obtient bien $M(s) = \mathbf{E}[M(t) | \mathcal{F}_s]$. \square

DÉFINITION A.35.– La tribu prévisible est la tribu sur $\Omega \times \mathbf{R}^+$ engendrée par les processus adaptés et continus. Elle est aussi engendrée par les processus adaptés et

continus à gauche ainsi que par les processus de la forme :

$$u(\omega, t) = \mathbf{1}_{]a, b]}(t)\alpha(\omega)$$

avec $\alpha \in \mathcal{F}_a$.

THÉORÈME A.32.– Soit M une martingale à variation finie. Soit u prévisible tel que :

$$\mathbf{E} \left[\int_0^\infty |u(s)| \, d \text{Var}(M)(s) \right] < \infty.$$

Le processus :

$$M^u(t) = \int_0^t u(s) \, dM(s)$$

est une martingale.

L'intégrale par rapport à M est une intégrale de Stieltjes comme définie dans la section A.4.

Démonstration. D'après l'hypothèse sur u et les propriétés de l'intégrale de Stieltjes, l'intégrabilité de $M^u(t)$ est assurée. Il reste à prouver (A.16). Montrons-le d'abord pour u prévisible simple :

$$u(\omega, t) = \mathbf{1}_{]a, b]}(t)\alpha(\omega)$$

avec $\alpha \in \mathcal{F}_a$. Pour $0 \leq s \leq a \leq t \leq b$, on a :

$$M^u(t) = \alpha(M(t) - M(a)) \text{ et } M^u(s) = 0.$$

Par conséquent :

$$\begin{aligned} \mathbf{E} [M^u(t) - M^u(s) \mid \mathcal{F}_s] &= \mathbf{E} [\alpha(M(t) - M(a)) \mid \mathcal{F}_s] \\ &= \mathbf{E} [\mathbf{E} [\alpha(M(t) - M(a)) \mid \mathcal{F}_a] \mid \mathcal{F}_s] \\ &= \mathbf{E} [\alpha \mathbf{E} [(M(t) - M(a)) \mid \mathcal{F}_a] \mid \mathcal{F}_s] \\ &= 0, \end{aligned}$$

puisque M est une martingale. Il suffit ensuite, sur le même principe, de discuter les autres cas suivant les positions relatives de a , s , t , b . Par passage à la limite, le résultat reste pour tous les processus prévisibles satisfaisant la propriété d'intégrabilité énoncée. \square

Parmi toutes les martingales, celles issues des processus ponctuels nous intéressent particulièrement.

DÉFINITION A.36.— Un processus ponctuel indexé par \mathbf{R}^+ est une suite presque sûrement strictement croissante $(T_n, n \geq 1)$ de variables aléatoires positives. On pose :

$$N(t) = \sum_{n \geq 1} \mathbf{1}_{[0,t]}(T_n),$$

le nombre de points avant t . On dira que le processus ponctuel est intégrable si $\mathbf{E}[N(t)] < \infty$ pour tout $t \geq 0$. En particulier, cela implique que T_n tend vers l'infini presque sûrement.

A partir de la connaissance des T_n , on retrouve facilement la trajectoire de N . A partir de la trajectoire de N , les instants T_n ne sont rien d'autre que ses instants de sauts. Il y a donc équivalence entre une mesure purement atomique sur \mathbf{R}^+ et le processus associé N . Nous emploierons le terme de processus ponctuel indifféremment pour l'un ou l'autre de ces objets.

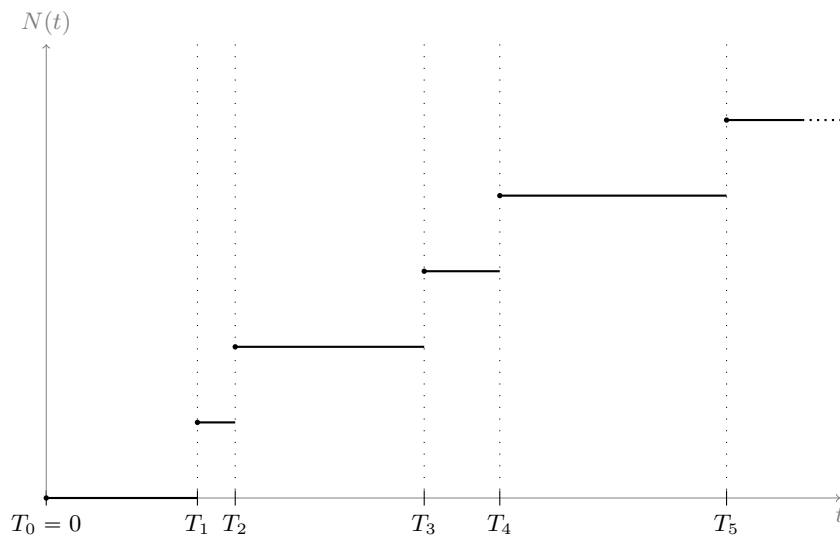


Figure A.6 – Processus ponctuel et mesure associée

DÉFINITION A.37.— Soit N un processus ponctuel intégrable. On appelle compensateur de N un processus croissant prévisible A , nul en 0 tel que $N - A$ soit une martingale locale.

EXEMPLE.— On tire de l'exemple A.3 que le compensateur du processus de Poisson est le processus $A(t) = \mu([0, t])$.

THÉORÈME A.33.— Soit N un processus ponctuel tel que $\sup_t \mathbf{E} [N(t)^2] < \infty$ de compensateur A . Le processus $((N - A)^2(t) - A(t), t \geq 0)$ est une martingale.

Démonstration. D'après le théorème A.22, on sait que :

$$(N(t) - A(t))^2 = 2 \int_0^t N(s^-) (dN(s) - dA(s)) + \sum_{s \leq t} \Delta N(s)^2.$$

Comme un processus ponctuel a des sauts de hauteur 1, il vient :

$$\Delta N(s)^2 = \Delta N(s) \text{ et } \sum_{s \leq t} \Delta N(s) = N(t).$$

Par conséquent :

$$\begin{aligned} (N(t) - A(t))^2 - A(t) &= 2 \int_0^t N(s^-) (dN(s) - dA(s)) + (N(t) - A(t)) \\ &= 2U^1(t) + U^2(t). \end{aligned}$$

D'après le théorème A.32, U^1 est une martingale et d'après la définition de A , U^2 aussi. \square

DÉFINITION A.38.— Un processus ponctuel marqué R à valeurs dans E polonais est une suite de variables aléatoires $((T_n, Z_n), n \geq 1)$ où $0 < T_n \leq T_{n+1}$ et $Z_n \in E$ pour tout n . Il est dit intégrable lorsque $\mathbf{E} [\sum_{n=1}^{\infty} \mathbf{1}_{[0, t]}(T_n)] < \infty$. On pose :

$$\sum_{n \geq 1} \psi(T_n, Z_n) = \iint_{\mathbf{R}^+ \times E} \psi(s, z) dR(s, z),$$

si tant est que le terme gauche soit défini.

REMARQUE.— Un processus ponctuel n'est rien d'autre qu'un processus ponctuel marqué avec E réduit à un singleton.

DÉFINITION A.39.— La filtration canoniquement associée à R processus ponctuel marqué est définie par :

$$\mathcal{F}_t = \sigma\{R([0, s] \times B), s \leq t, B \in \mathcal{B}(E)\}.$$

La tribu prévisible associée à R processus ponctuel marqué est la tribu sur $\Omega \times \mathbf{R}^+ \times E$ engendrée par les processus de la forme :

$$\psi(\omega, s, z) = \alpha(\omega) \mathbf{1}_{]a, b]}(s) g(z),$$

avec g fonction mesurable bornée de $(E, \mathcal{B}(E))$ dans $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$, $\alpha \in \mathcal{F}_a$.

DÉFINITION A.40.– Une mesure aléatoire sur $\mathbf{R}^+ \times E$ est dite prévisible si pour tout $B \in \mathcal{B}(E)$, le processus :

$$t \mapsto R([0, t] \times B)$$

est \mathcal{F} -prévisible.

DÉFINITION A.41.– Soit R processus ponctuel marqué. On note \mathbf{Q}_n la loi conditionnelle de (T_{n+1}, Z_{n+1}) sachant $\mathcal{H}_n = (T_j, Z_j, j = 1, \dots, n)$. Pour ψ prévisible positif, on définit :

$$\begin{aligned} & \int_0^t \int_E \psi(s, z) \, d\nu(s, z) \\ &= \sum_{n \geq 0} \int_0^t \int_E \psi(s, z) \frac{1}{\mathbf{Q}_n([s, \infty[\times E])} \mathbf{1}_{]T_n, T_{n+1}]}(s) \, d\mathbf{Q}_n(s, z). \end{aligned} \quad (\text{A.17})$$

THÉORÈME A.34.– Soit R processus ponctuel marqué. Pour tout processus prévisible ψ tel que :

$$\sup_t \mathbf{E} \left[\int_0^t \int_E \psi^2(s, z) \, d\nu(s, z) \right] < \infty, \quad (\text{A.18})$$

le processus :

$$M^\psi : t \mapsto \int_0^t \int_E \psi(s, z) \, dR(s, z) - \int_0^t \int_E \psi(s, z) \, d\nu(s, z)$$

est une martingale locale. De plus, ν est l'unique mesure prévisible qui satisfait cette propriété. De plus, dans ce cas, on a :

$$\langle M^\psi, M^\psi \rangle(t) = \int_0^t \int_E \psi^2(s, z) \, d\nu(s, z). \quad (\text{A.19})$$

Démonstration. Sur l'intervalle $]T_n, T_{n+1}]$, le processus :

$$t \mapsto \int_0^t \int_E \psi(s, z) \, d\nu(s, z)$$

est \mathcal{H}_n mesurable par conséquent il est prévisible. Montrons maintenant que pour ψ prévisible positif, l'on a $\mathbf{E}[M^\psi(t)] = 0$. Puisque ψ est prévisible, on a l'identité suivante :

$$\mathbf{E}[\psi(t, z) \mathbf{1}_{]T_n, T_{n+1}]}(t) \mid \mathcal{H}_n] = \psi(t, z) \mathbf{E}[\mathbf{1}_{]T_n, T_{n+1}]}(t) \mid \mathcal{H}_n].$$

D'autre part, la loi de T_{n+1} sachant \mathcal{H}_n est par définition la marginale sur \mathbf{R}^+ de \mathbf{Q}_n :

$$\mathbf{E}[\mathbf{1}_{[s, \infty]}(T_{n+1}) \mid \mathcal{H}_n] = \int_s^\infty \int_E d\mathbf{Q}_n(r, \tau) = \mathbf{Q}_n([s, \infty[\times E).$$

Par conséquent :

$$\begin{aligned}
& \mathbf{E} \left[\sum_{n \geq 1} \int_0^t \int_E \psi(s, z) \frac{1}{\mathbf{Q}_n([s, \infty[\times E)} \mathbf{1}_{]T_n, T_{n+1}]}(s) d\mathbf{Q}_n(s, z) \right] \\
&= \sum_{n \geq 1} \mathbf{E} \left[\int_0^t \int_E \psi(s, z) \frac{1}{\mathbf{Q}_n([s, \infty[\times E)} \mathbf{E}[\mathbf{1}_{]T_n, T_{n+1}]}(s) | \mathcal{H}_n] d\mathbf{Q}_n(s, z) \right] \\
&= \sum_{n \geq 1} \mathbf{E} \left[\int_0^t \int_E \psi(s, z) \frac{1}{\mathbf{Q}_n([s, \infty[\times E)} \int_s^\infty \int_E d\mathbf{Q}_n(r, \tau) d\mathbf{Q}_n(s, z) \right] \\
&= \sum_{n \geq 1} \mathbf{E} \left[\int_0^t \int_E \psi(s, z) d\mathbf{Q}_n(s, z) \right] \\
&= \sum_{n \geq 1} \mathbf{E} \left[\int_0^t \int_E \mathbf{E}[\psi(T_{n+1}, Z_{n+1}) | \mathcal{H}_n] \right] \\
&= \mathbf{E} \left[\int_0^t \int_E \psi(s, z) dR(s, z) \right].
\end{aligned} \tag{A.20}$$

Si ψ est prévisible simple positif, pour $t > r > 0$ considérons Y - \mathcal{F}_r mesurable positive, le processus :

$$s \mapsto \psi(s, z) Y \mathbf{1}_{]r, t]}(s)$$

est encore prévisible. Par passage à la limite, cela reste vrai pour ψ prévisible positif. Par conséquent :

$$\begin{aligned}
& \mathbf{E} [M^\psi(t) - M^\psi(r) | \mathcal{F}_r] \\
&= \mathbf{E} \left[\int_0^t \mathbf{1}_{]r, t]}(s) \int_E \psi(s, z) (dR(s, z) - d\nu(s, z)) | \mathcal{F}_r \right]. \tag{A.21}
\end{aligned}$$

Or, pour $Y \in \mathcal{F}_r$, d'après (A.20), on a :

$$\mathbf{E} \left[\int_0^t \int_E Y \mathbf{1}_{]r, t]}(s) \psi(s, z) (dR(s, z) - d\nu(s, z)) \right] = 0,$$

donc $\mathbf{E} [M^\psi(t) - M^\psi(r) | \mathcal{F}_r] = 0$ et M^ψ est bien une martingale. En raisonnant comme dans le théorème A.33, on obtient (A.19). Tout ce qui précède n'a de sens que si les espérances sont finies, ce dont on ne sait rien *a priori* pour un ψ prévisible quelconque. Afin de donner un sens à ces calculs, on considère la suite :

$$\tau_n = \inf \left\{ t, \int_0^t \int_E \psi(s, z) (dR(s, z) + d\nu(s, z)) > n \right\},$$

et on applique le raisonnement de (A.20) sur $[0, t \wedge \tau_k]$ au lieu de $[0, t]$. Toutes les espérances sont finies et on a un calcul parfaitement rigoureux. L'équation (A.21) devient :

$$\begin{aligned} & \mathbf{E} [M^\psi(t \wedge \tau_k) - M^\psi(r \wedge \tau_k) | \mathcal{F}_r] \\ &= \mathbf{E} \left[\int_0^{t \wedge \tau_k} \mathbf{1}_{]r, t \wedge \tau_k]}(s) \int_E \psi(s, z) (dR(s, z) - d\nu(s, z)) | \mathcal{F}_r \right]. \end{aligned}$$

On en déduit que M^ψ est une martingale locale. La condition (A.18) permet de passer à la limite dans les espérances d'où le résultat final. \square

On admet le résultat suivant dont la preuve s'inspire de celle du théorème A.27 mais est beaucoup plus technique.

THÉORÈME A.35.– Une sous-martingale M locale admet une décomposition sous la forme :

$$M(t) = X(t) + A(t),$$

où X est une martingale locale et A un processus prévisible croissant nul en 0. Le processus A est souvent noté $\langle M, M \rangle$.

Pour les propriétés de type PASTA, nous avons besoin du théorème suivant dont la démonstration très technique est omise.

THÉORÈME A.36.– Soit M une martingale et $\langle M, M \rangle$ son crochet. Si $\langle M, M \rangle(t)$ tend vers l'infini quand t tend vers l'infini alors :

$$\frac{M(t)}{\langle M, M \rangle(t)} \xrightarrow{t \rightarrow \infty} 0.$$

COROLLAIRE A.37.– Soit R un processus ponctuel marqué sur E polonais de compensateur ν . On note N le processus ponctuel associé : $N(t) = R([0, t] \times E)$. Soit $\psi : \Omega \times \mathbf{R}^+ \times E \rightarrow \mathbf{R}$ un processus prévisible. Supposons qu'il existe $c > 0$ tel que presque sûrement, pour tout $t \geq 0$, on ait :

$$\int_0^t \int_E (1 + \psi^2(s, z)) d\nu(s, z) \leq c \nu([0, t] \times E), \quad (\text{A.22})$$

alors, presque sûrement, on a :

$$\lim_{t \rightarrow \infty} \left(\frac{1}{N(t)} \int_0^t \psi(s, z) dR(s, z) - \frac{1}{\nu([0, t] \times E)} \int_0^t \psi(s, z) d\nu(s, z) \right) = 0. \quad (\text{A.23})$$

⌚ Cette formule est à la base de toutes les propriétés de type PASTA qui apparaissent dans cet ouvrage. Elle permet de relier les moyennes calculées vues des utilisateurs en fonction des moyennes calculées sur des temps longs.

Démonstration. Pour simplifier les notations, on pose :

$$\nu(t) = \nu([0, t] \times E) \text{ et } \nu^\phi(t) = \int_0^t \phi(s, z) d\nu(s, z).$$

Remarquons d'abord que d'après le théorème A.36, on sait que :

$$\frac{1}{\nu(t)} (N(t) - \nu(t)) \xrightarrow{t \rightarrow \infty} 0. \quad (\text{A.24})$$

Par conséquent :

$$\frac{N(t)}{\nu(t)} \xrightarrow{t \rightarrow \infty} 1. \quad (\text{A.25})$$

On en déduit que :

$$\frac{\nu^\psi(t)}{N(t)} = \frac{\nu(t)}{N(t)} \frac{\nu^\psi(t)}{\nu(t)} \leq \frac{\nu(t)}{N(t)} \frac{1}{\nu(t)} \int_0^t \int_E (1 + \psi^2)(s, z) d\nu(s, z).$$

D'après (A.22) et (A.25), cette quantité est donc bornée uniformément par rapport au temps. En écrivant :

$$\begin{aligned} \frac{1}{N(t)} \int_0^t \int_E \psi(s, z) dR(s, z) &= \frac{\nu(t)}{N(t)} \frac{\nu^{\psi^2}(t)}{\nu(t)} \\ &\times \frac{1}{\nu^{\psi^2}(t)} \int_0^t \int_E \psi(s, z) (dR(s, z) - d\nu(s, z)) + \frac{\nu^\psi(t)}{N(t)}, \end{aligned}$$

on déduit de ce qui précède qu'il existe $r > 0$ tel que :

$$\limsup_{t \rightarrow \infty} \frac{1}{N(t)} \int_0^t \int_E \psi(s, z) dR(s, z) \leq r. \quad (\text{A.26})$$

Munis de ces résultats de domination, nous pouvons maintenant procéder au calcul de la limite qui nous intéresse vraiment :

$$\begin{aligned} &\frac{1}{N(t)} \int_0^t \psi dR - \frac{1}{\nu(t)} \int_0^t \psi d\nu \\ &= \frac{1}{N(t)} \left(\int_0^t \int_E \psi dR \right) \left(\frac{\nu(t) - N(t)}{\nu(t)} \right) + \frac{\nu^{\psi^2}(t)}{\nu(t)} \frac{1}{\nu^{\psi^2}(t)} \int \psi (dR - d\nu). \end{aligned}$$

On déduit (A.23) du théorème A.36 et de (A.24), (A.25) et (A.26). \square

Le théorème suivant est une application directe du théorème A.34. Il stipule que les moyennes calculées du point de vue des clients sont égales aux moyennes temporelles lorsque le processus d'arrivée est un processus de Poisson. On regardera utilement la figure 9.2 et ses commentaires pour voir que ce n'est pas toujours le cas.

THÉORÈME A.38 (Propriété PASTA).— Si N est un processus de Poisson d'intensité λ alors :

$$\lim_{t \rightarrow \infty} \frac{1}{N(t)} \sum_{n: T_n \leq t} \psi(T_n) = \lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t \psi(s) \, ds,$$

dès que l'une des deux limites existe.

A.6. Transformée de Laplace

DÉFINITION A.42.— Soit Y une variable aléatoire à valeurs dans \mathbf{R}^+ . La transformée de Laplace de la loi de Y est la fonction à valeur dans \mathbf{R}^+ et définie pour tout $s \in \mathbf{R}^+$ par :

$$\mathcal{L}_Y(s) = \mathbf{E} [e^{-sY}].$$

LEMME A.39.— La transformée de Laplace a les propriétés suivantes :

1) elle caractérise la loi : si Y et Z sont deux variables aléatoires telles que $\mathcal{L}_Y(s) = \mathcal{L}_Z(s)$ pour tout s , alors Y et Z ont la même loi ;

2) en particulier, elle caractérise tous les moments de la loi : si Y a un moment d'ordre n alors \mathcal{L}_Y est n fois dérivable en 0, et pour tout $k = 1, \dots, n$, on a :

$$\mathbf{E} [Y^k] = (-1)^k \mathcal{L}_Y^{(k)}(0) ;$$

3) si Y_1, Y_2, \dots, Y_n sont n variables aléatoires indépendantes admettant des transformées de Laplace en s , alors la transformée de Laplace de $\sum_{i=1}^n Y_i$ est définie en s et vaut :

$$\mathcal{L}_{\sum_{i=1}^n Y_i}(s) = \prod_{i=1}^n \mathcal{L}_{Y_i}(s) ;$$

4) soit $(Y_n, n \in \mathbf{N}^*)$ et Y , des variables aléatoires admettant des transformées de Laplace sur un ouvert commun, on a alors l'équivalence suivante :

$$Y_n \xrightarrow{n \rightarrow \infty} Y \text{ en loi} \implies \mathcal{L}_{Y_n} \xrightarrow{n \rightarrow \infty} \mathcal{L}_Y \text{ simplement.}$$

Démonstration. Le résultat 1) repose sur un théorème d'analyse complexe, sur lequel nous ne nous étendrons pas : on peut inverser toute transformée de Laplace \mathcal{L} définie sur \mathbf{R} pour retrouver l'unique fonction à valeurs positive f telle que $\mathcal{L}(s) = \int_{\mathbf{R}^+} e^{-sx} f(x) \, dx$. Ainsi, deux variables aléatoires qui ont la même transformée de Laplace ont la même densité.

Si Y a un moment d'ordre n alors par dérivation sous le signe somme, on a :

$$\begin{aligned}\mathcal{L}'_Y(s) &= \int_{\mathbf{R}^+} -ye^{-sy} d\mathbf{P}_Y(y) \\ \mathcal{L}_Y^{(2)}(s) &= \int_{\mathbf{R}^+} y^2 e^{-sy} d\mathbf{P}_Y(y) \\ &\vdots \\ \mathcal{L}_Y^{(n)}(s) &= \int_{\mathbf{R}^+} (-1)^n y^n e^{-sy} d\mathbf{P}_Y(y),\end{aligned}$$

ce qui montre bien 2) annoncé en posant $s = 0$.

On raisonne par récurrence. Si 3) est vrai au rang n , pour tout s tel que les transformées de Laplace sont définies en s , on a :

$$\begin{aligned}\mathcal{L}_{\sum_{i=1}^{n+1} Y_i}(s) &= \mathbf{E} \left[e^{-s \sum_{i=1}^{n+1} Y_i} \right] = \mathbf{E} \left[e^{-s \sum_{i=1}^n Y_i} e^{-s Y_{n+1}} \right] \\ &= \mathbf{E} \left[e^{-s \sum_{i=1}^n Y_i} \right] \mathbf{E} \left[e^{-s Y_{n+1}} \right] \\ &= \prod_{i=1}^n \mathcal{L}_{\sum_{i=1}^n Y_i}(s) \mathcal{L}_{Y_{n+1}}(s) = \prod_{i=1}^{n+1} \mathcal{L}_{Y_i}(s).\end{aligned}$$

Pour 4) le sens direct est automatique en prenant pour tout s , la fonction test $f(x) = e^{-xs}$, qui est mesurable et positive. Pour la réciproque, il suffit d'utiliser 1 et d'inverser la transformée de Laplace limite pour obtenir la convergence en loi. \square

A.7. Notes et commentaires

L'espérance conditionnelle est un grand classique de l'enseignement des probabilités avancées. Il existe de nombreuses références. Nous nous sommes particulièrement inspirés de [CHU 01, KAL 98]. On trouvera une présentation de la théorie de la mesure et des espaces de Hilbert dans [RUD 80, YOS 95]. Tout ce qui concerne les compensateurs de processus ponctuels apparaît de façon complète mais un peu ardue dans [JAC 79] sous le nom de processus ponctuels multivariés. On pourra aussi consulter [LAS 95]. La théorie générale de l'intégration par rapport à une martingale se trouve dans de nombreux ouvrages pour le cas continu, les livres qui traitent du cas général, c'est-à-dire pour les trajectoires cadlag sont beaucoup plus rares et plus difficiles d'accès, citons [BRE 81, DEL 76, JAC 79].

Bibliographie

- [ARC 10] ARCEP, La qualité des services de voix et de données des réseaux mobiles (2G et 3G) en France métropolitaine, ARCEP, 2010.
- [ASM 03] ASMUSSEN S., *Applied probability and queues*, vol. 51, Springer-Verlag, New York, 2003.
- [BAC 84] BACCELLI F., BOYER P., HÉBUTERNE G., « Single-server queues with impatient customers », *Advances in Applied Probability*, vol. 16, n° 4, p. 887-905, 1984.
- [BAC 89] BACCELLI F., MAKOWSKI A., « Multidimensional stochastic ordering and associated random variables », *Operations Research*, vol. 37, n° 3, p. 478-487, 1989.
- [BAC 02] BACCELLI F., BRÉMAUD P., *Elements of Queueing Theory*, Springer-Verlag, Berlin, 2002.
- [BAC 09a] BACCELLI F., BLASZCZYSZYN B., *Stochastic Geometry and Wireless Networks*, vol. I, Now publishers, Hanover, 2009.
- [BAC 09b] BACCELLI F., BLASZCZYSZYN B., *Stochastic Geometry and Wireless Networks*, vol. II, Now publishers, Hanover, 2009.
- [BAL 01] BALDI P., MAZLIAK L., PRIOURET P., *Martingales et chaînes de Markov*, Hermann, Paris, 2001.
- [BON 11] BONALD T., FEUILLET T., *Performances des réseaux et des systèmes informatiques*, Hermès Sciences, Paris, 2011.
- [BOR 84] BOROVKOV A.A., *Asymptotic methods in queuing theory*, John Wiley & Sons, Chichester, 1984.
- [BOR 92] BOROVKOV A.A., FOSS S.G., « Stochastically recursive sequences and their generalizations », *Siberian Advances in Mathematics*, vol. 2, n° 1, p. 16-81, 1992.
- [BOR 94] BOROVKOV A.A., FOSS S.G., « Two ergodicity criteria for stochastically recursive sequences », *Acta Applicandae Mathematicae*, vol. 34, n° 1-2, p. 125-134, 1994.
- [BOR 98] BOROVKOV A.A., *Ergodicity and stability of stochastic processes*, John Wiley & Sons, Chichester, 1998.

- [BRA 90] BRANDT A., FRANKEN P., LISEK B., *Stationary stochastic models*, vol. 78, Akademie-Verlag, Berlin, 1990.
- [BRE 81] BRÉMAUD P., *Point processes and queues, martingale dynamics*, Springer-Verlag, New York, 1981.
- [CHA 99] CHAO X., MIYAZAWA M., PINEDO M., *Queueing networks : customers, signals, and product form solutions*, John Wiley & Sons, Chichester, 1999.
- [CHU 01] CHUNG K., *A course in probability theory*, Academic Press, San Diego, CA, 2001.
- [CIN 75] ÇINLAR E., *Introduction to stochastic processes*, Prentice-Hall, Englewood Cliffs, NJ, 1975.
- [CRO 96] CROVELLA M., BESTAVROS A., « Self-Similarity in World Wide Web Traffic : Evidence and Causes », *ACM Sigmetrics Conference on Measurement and Modeling of Computer Systems*, 1996.
- [DAL 03] DALEY D.J., VERE-JONES D., *An introduction to the theory of point processes.*, vol. I, Springer-Verlag, New York, 2003.
- [DEL 76] DELLACHERIE C., MEYER P., *Probabilités et Potentiel*, vol. 2, Masson, Paris, 1976.
- [DOU 02] DOUKHAN P., OPPENHEIM G., TAQQU M., DIR., *Long range dependence : theory and applications*, Birkhäuser, Boston, 2002.
- [ETH 86] ETHIER S., KURTZ T., *Markov Processes : Characterizations and Convergence*, John Wiley & Sons, New York, 1986.
- [FIS 93] FISCHER W., MEIER-HELLSTERN K., « The Markov-modulated Poisson process (MMPP) cookbook », *Performance Evaluation*, vol. 18, n° 2, p. 149-171, 1993.
- [FLI 81] FLIPO D., « Comparaison des disciplines de service des files d'attente $G/G/1$ », *Annales de l'Institut Henri Poincaré. Section B.*, vol. 17, n° 2, p. 191-212, 1981.
- [FOS 81] FOSS S.G., « Comparison of service disciplines in multichannel systems with waiting », *Sibirskii Matematicheskii Zhurnal*, vol. 22, n° 1, p. 190-197, 1981.
- [GRA 08] GRAHAM C., *Chaînes de Markov*, Dunod, Paris, 2008.
- [HAE 08] HAENGGI M., GANTI R., « Interference in Large Wireless Networks », *Foundations and Trends in Networking*, vol. 3, n° 2, p. 127-248, 2008.
- [HOU 02] HOUDRÉ C., PRIVAULT N., « Concentration and deviation inequalities in infinite dimensions via covariance representations », *Bernoulli*, vol. 8, n° 6, p. 697-720, 2002.
- [IVE 01] IVERSEN V., *Teletraffic engineering and network planning*, Technical University of Denmark, 2001.
- [JAC 79] JACOD J., *Calcul stochastique et problèmes de martingales*, Springer-Verlag, Berlin, 1979.
- [KAL 98] KALLENBERG O., *Foundations of Modern Probability*, Springer-Verlag, New York, 1998.
- [KEL 79] KELLY F., *Reversibility and stochastic networks*, John Wiley & Sons, Chichester, 1979.

- [KLE 76] KLEINROCK L., *Queueing systems*, John Wiley & Sons, Chichester, 1976.
- [LAG 96] LAGRANGE X., GODLEWSKI P., « Performance of a hierarchical cellular network with mobility dependent hand-over strategies », *Vehicular Technology Conference*, p. 1868-1872, 1996.
- [LAS 95] LAST G., BRANDT A., *Marked point processes on the real line*, Springer-Verlag, New York, 1995.
- [LEL 94] LELAND W., TAQQU M., WILLINGER W., WILSON D., « On the Self-similar Nature of Ethernet Traffic », *IEEE/ACM transactions on networking*, vol. 2, n° 1, p. 1-15, 1994.
- [LOY 62] LOYNES R.M., « Stationary waiting-time distributions for single-server queues », *Annals of Mathematical Statistics*, vol. 33, p. 1323-1339, 1962.
- [MEI 89] MEIER-HELLSTERN K., « The analysis of a queue arising in overflow models », *IEEE Transactions on Communications*, vol. 37, n° 4, p. 367-372, 1989.
- [MOY 08a] MOYAL P., « Convex comparison of service disciplines in real time queues », *Operations Research Letters*, vol. 36, n° 4, p. 496-499, 2008.
- [MOY 08b] MOYAL P., « Stability of a processor-sharing queue with varying throughput », *Journal of Applied Probability*, vol. 45, n° 4, p. 953-962, 2008.
- [MOY 10] MOYAL P., « The queue with impatience : construction of the stationary workload under FIFO », *Journal of Applied Probability*, vol. 47, n° 2, p. 498-512, 2010.
- [NEU 94] NEUTS M., *Matrix-geometric solutions in stochastic models*, Dover Publications, New York, 1994.
- [NEV 84] NEVEU J., « Construction de files d'attente stationnaires », *Modelling and performance evaluation methodology (Paris, 1983)*, vol. 60 de *Lecture Notes in Control and Information Science*, p. 31-41, Springer, Berlin, 1984.
- [NOR 94] NORROS I., « A Storage Model with Self-Similar Inputs », *Queueing Systems*, vol. 16, p. 387-396, 1994.
- [PAN 88] PANWAR S., TOWSLEY D., WOLF J., « Optimal scheduling policies for a class of queues with customer deadlines to the beginning of service », *Journal of the Association for Computing Machinery*, vol. 35, n° 4, p. 832-844, 1988.
- [PRI 09] PRIVAULT N., *Stochastic analysis in discrete and continuous settings with normal martingales*, vol. 1982, Springer-Verlag, Berlin, 2009.
- [RIG 98] RIGAULT C., *Principes de commutation numérique*, Hermès, Paris, 1998.
- [ROB 03] ROBERT P., *Stochastic networks and queues*, vol. 52, Springer-Verlag, Berlin, 2003.
- [RUD 80] RUDIN W., *Analyse réelle et complexe*, Masson, Paris, 1980.
- [SAM 94] SAMORODNITSKY G., TAQQU M., *Stable non-Gaussian random processes*, Chapman & Hall, New York, 1994.
- [SHE 97] SHERMAN R., TAQQU M., WILLINGER W., « Proof of a fundamental result in self-similar traffic modeling », *Computer Communication Review*, vol. 27, n° 2, 1997.

- [STO 83] STOYAN D., *Comparison methods for queues and other stochastic models*, John Wiley & sons, Chichester, 1983.
- [THO 00] THORISSON H., *Coupling, stationarity, and regeneration*, Springer-Verlag, New York, 2000.
- [WU 00] WU L., « A new modified logarithmic Sobolev inequality for Poisson point processes and several applications », *Probability Theory and Related Fields*, vol. 118, n° 3, p. 427-438, 2000.
- [YOS 95] YOSIDA K., *Functional analysis*, Springer-Verlag, Berlin, 1995.

Index

- équation
 - de Pollacek-Khinchin, 166
- argument d'échange, 114
- cadlag, 351
- chaîne de Markov
 - incluse, 208
- classe monotone, 357, 359, 360
- compensateur, 383
- couplage, 46
 - arrière fort, 46
 - indice *backward*, 46
 - indice *forward*, 46
- équation
 - de Kiefer et Wolfowitz, 122
- ergodique
 - lemme, 39
 - probabilité, 38
 - quadruplet, 38
 - suite, 38
 - théorème, 44
- état
 - absorbant, 61
 - cimetière, δ , 61
 - conduisant à, 68
 - périodique, 83
 - récurrent, 66
 - récurrent nul, 72
 - récurrent positif, 72
 - transient, 66
- file
 - à plafond de charge, 156
 - à plusieurs serveurs, 131
 - à rejet, 155
 - processor sharing*, 261
 - avec clients impatients, 142
 - Engset, 292
 - G/G/1, 98
 - IPP/S/S, 291
 - M/M/1, 208, 259
 - M/M/ ∞ , 262, 319
 - M/M/1/S+1, 284
 - M/M/S, 259
 - M/M/S/S, 282
 - MMPP/S/S, 299
- files en parallèle, 121
- générateur infinitésimal, 337
- générateur infinitésimal, 213
 - A, 213
 - \mathfrak{L}_Λ , 337
- gradient discret, $D_x F$, 334
- intégration
 - par parties, 371
 - Stieltjes, $\int g d f$, 369
- Kronecker
 - produit, \otimes , 238
 - somme, \oplus , 238
- $l^2(E, \pi)$, $\| \cdot \|_2$, 350
- $l^\infty(E)$, $\| \cdot \|_\infty$, 350
- méthode de Stein, 342
- martingale, 227, 373, 380
 - arrêtée, 374
 - convergence, 378, 381

- crochet, $\langle M, M \rangle$, 387
- décomposition, 376, 387
- intégration, 382
- locale, 227, 381, 383
- sous-martingale, 373
- sur-martingale, 373
- mesure
 - de Radon, 317
 - image, $\psi^* \mathbf{P}$, 37
- MMPP, 237
 - IPP, 290
- opérateur
 - carré du champ, 89
 - décalage, 36
 - de décalage, 59
 - de transition, 60
- ordre
 - fort, \leq_{st} , 52
 - générique, \leq , 42
 - Schur-convexe, $<_c$, 111, 363
 - vectériel, $<$, 52, 363
- PASTA, 291, 387, 389
- politique de service
 - conservative, 21
 - EDF, 21, 115
 - FCFS, FIFO, 20, 105
 - LCFS, LIFO, 21, 106
 - LDF, 116
 - préemptive, 21
 - SRPT, 21, 106
- prévisible
 - mesure, 385
 - processus, 376
 - tribu, 381, 384
- probabilité
 - blocage, 281
 - conditionnelle, 357
 - dépassement de seuil, 281
 - perte, 281
- processus
 - naissance et mort, 235
 - réversible, 234
- processus ponctuel
 - intégrable, 383
- processus de Poisson
 - composé, 330
- processus de Markov
 - régulier, 210
- processus de Poisson, 317, 380
 - amincissement, 190, 323
 - déplacement, 324
 - Glauber, \mathfrak{N}^Λ , 335
 - intensité, 182
 - superposition, 190
- processus ponctuel, 383
 - intégrable, 384
 - marqué, 384
 - superposition, 238, 323
- semi-groupe, 212, 337
- semi-groupe
 - P_t , 212
 - \mathfrak{P}^Λ , 337
- simulation
 - MMPP, 239
 - Poisson sur \mathbf{R}^+ , 196, 197
 - Poisson sur \mathbf{R}^d , 318
 - Poisson-Glauber, 336
 - processus de Markov, 207
- stationnaire
 - probabilité, 37
 - suite, 37
- Strassen (théorème), 53
- TCH, 287
- temps
 - d'attente, 170
 - de séjour, 170
 - d'atteinte, τ_x^1 , 64
 - d'attente, TA, 109, 252
 - de séjour, Ts, 109, 252, 259