CHAPITRE 5

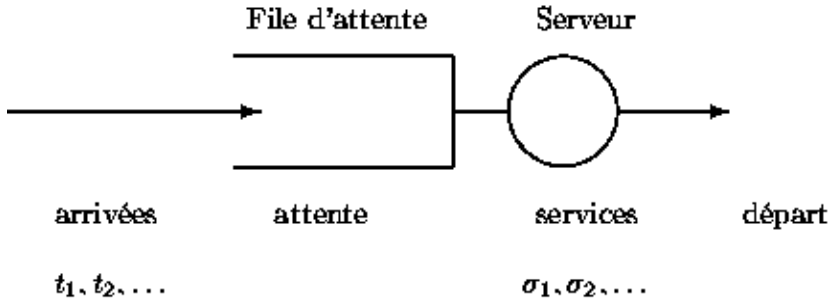
SYSTEMES D'ATTENTE
PROCESSUS DE NAISSANCE ET DE MORT

1 Présentation et notations de Kendall

1.1 Généralités

Un système d'attente est constitué de :

1. Un **flux d'arrivées** qui représentent les instants où arrivent les clients (terme générique représentant aussi bien des véhicules, des coups de téléphone, des appels à la mémoire d'un ordinateur ou les sollicitations de l'organe central d'un réseau téléinformatique) En première approximation, on considère souvent que les délais entre les arrivées sont des variables aléatoires indépendantes de même loi. Le cas le plus simple est celui où la loi commune est une loi exponentielle, le flux d'arrivée est alors poissonien.
2. Un **organe de service** qui est caractérisé par :
   - un temps de service : un client qui commence à être servi sera immobilisé pendant un temps aléatoire dont on supposera connu la loi
   - le nombre de guichets
3. Une **règle ou discipline de services** qui indique comment fonctionne le système :
   - système avec attente ou sans attente (dans un système sans attente, il n'y a pas de queue. Un client qui ne peut être servi à son arrivée est perdu)
   - système de capacité limitée (au-delà d'une certaine longueur de queue, les nouveaux clients sont perdus)
   - ordre dans lequel on est servi (par exemple FIFO "first in first out" c'est-à-dire "premier arrivé premier servi", LIFO "last in first out" c'est-à-dire "dernier arrivé premier servi", cas usuel lorsque entre deux poste d'usinage, les pièces sont empilées les unes sur les autres puis saisies en commençant par celle qui est sur le dessus).
   - plusieurs classes de clients, chacune prioritaire par rapport aux classes suivantes.



1.2 Les notations de Kendall

Pour leur étude mathématique, on classe les systèmes d'attente selon notation standard A/S/P/K/D, où :

- A désigne la loi des interarrivées (par défaut A sera désignée par **M** (comme *Markov*) pour un flux poissonien et une distribution des arrivées exponentielle, ou bien **D** (comme *déterministe*) pour un flux d'arrivées à intervalles réguliers (processus d'arrivées périodique), ou encore **G** pour une distribution *générale*)
- S désigne la loi de service (par défaut S vaut aussi **M, D** ou **G**)
- P désigne le nombre de serveurs (par défaut P vaut **1**)

- K désigne la capacité du système, c'est-à-dire le nombre maximal de clients pouvant être présents simultanément dans le serveur ou la file d'attente) (par défaut K vaut l'**infini**)

- D désigne la discipline de service (par défaut ce sera **FIFO**)

K et D admettent des valeurs par défaut. Ainsi, M/M/1/K est équivalent à M/M/1/K/FIFO et M/M/1 est l'abréviation de M/M/1/∞/FIFO.

Exemple 1.1 *Le système d'attente M/D/2/5 correspond à des arrivées poissonniennes vers deux serveurs de temps de service constant. De plus, la capacité de la file d'attente est de* $5 - 2 = 3$.

Exercice 1 *Expliciter chacun des systèmes M/M/1/1 et M/G/5/1.*

Définition 1.2 ***(Nombre de clients présents dans la file d'attente à l'instant t)*** *C'est une fonction N à valeurs dans l'ensemble* $\{0, 1, ..., K\}$ *où K est la capacité du système. C'est donc une fonction en escaliers.*

Exemple 1.3 *Considérons une file d'attente à un serveur de capacité* $K = \infty$ *et de discipline de service FIFO, dont les dates d'arrivées et les durées de service sont données par le tableau suivant :*

Numéro	*Date d'arrivée*	*Temps de service*
1	*0*	*4*
2	*2*	*8*
3	*6*	*4*
4	*8*	*4*
5	*15*	*2*

Le nombre $N(t)$ *de clients à l'instant t suit la courbe de la figure 1.*

Exercice 2 *On note M/M/1 le système M/M/1/*∞*. Le système est supposé vide au temps* $t = 0$.

1. *On suppose que l'intensité du flux d'arrivée vaut* $0,5$ *et que le temps moyen de service est égal à 1.*

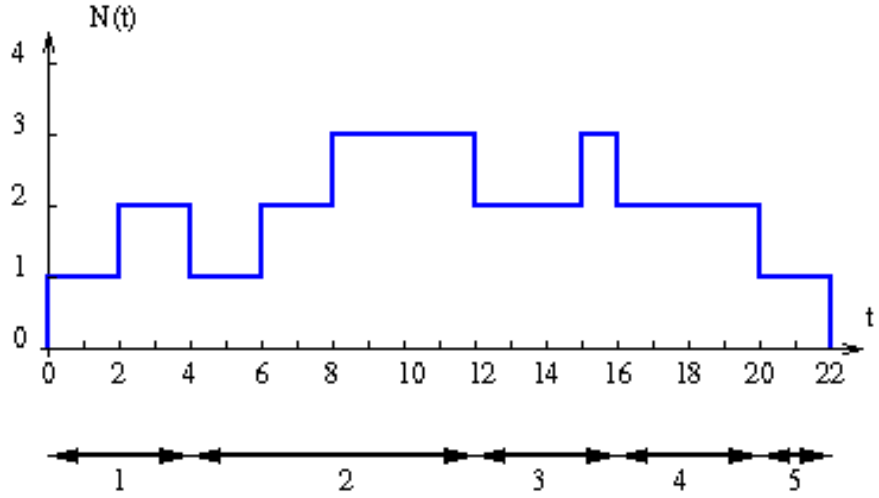


FIG. 1 – exemple

   (a) *Simuler les 5 premiers changements dans la file (on arrondira à l'unité les temps d'arrivées et de services).*

   (b) *Représenter graphiquement la taille du système.*

2. *Reprendre la question 1 lorsque l'intensité du flux d'arrivée vaut 1 et le temps moyen de service vaut 2.*

2 Processus de naissance et de mort

2.1 Construction dynamique d'un processus de naissance ou de mort

Remarque 2.1 *La définition proposée ici n'est pas la plus large. On se restreindra en effet au cas réellement utile dans l'étude des files M/M/1 et M/M/2.*

Définition 2.2 *On suppose qu'une population a* ξ *éléments à l'instant* $t = 0$*, où* ξ *est une variable aléatoire.*

Soit $(\lambda_i)_{i\geq 0}$ *et* $(\mu_i)_{i\geq 1}$ *des réels strictement positifs.*

Le processus $(N_t)_{t\geq 0}$ donnant la taille de la population à l'instant t est un processus de naissance et de mort s'il respecte les points suivants :

1. *Si la population est à un instant nulle, alors il y aura une naissance dans un délai exponentiel de paramètre λ_0*
2. *Si après une naissance ou une mort, la population n'est pas nulle et est égale à i, alors :*
   - *il y a une naissance avec une probabilité de $\dfrac{\lambda_i}{\lambda_i+\mu_i}$*
   - *il y a une mort avec une probabilité de $\dfrac{\mu_i}{\lambda_i+\mu_i}$*
   - *ces événements ont lieu suivant des délais exponentiels de paramètre $\lambda_i+\mu_i$.*

Le processus $(N_t)_{t\geq 0}$ donnant la taille de la population à l'instant t est un processus de naissance et de mort (ou PNM).

Remarque 2.3 *Les processus de naissance et de mort sont des exemples simples de processus Markoviens de sauts.*

2.2 Exercice

Exercice 3 *On considère un PNM tel que $\xi = 2$, $\lambda_0 = 1$, $\lambda_1 = 2$, $\mu_0 = 0,5$, $\mu_1 = 1$, $\mu_2 = 2$. Tous les autres paramètres sont nuls.*

1. *Simuler les 8 premiers événements (naissance ou mort) d'une trajectoire de ce PNM.*
2. *Faire une représentation graphique.*

3 Le système M/M/1

3.1 Description comme PNM

- les arrivées sont poissoniennes d'intensité λ et les temps de service sont exponentiels de paramètre μ
- le système dispose d'un seul serveur
- ξ est la taille du système à l'instant $t = 0$ et N_t est la taille du système à l'instant t

Remarque 3.1 *Quelques pistes intéressantes à étudier en pratique :*

- *Quand est-ce que le système (File + Service) sera débordé ? Existe-t-il des conditions sur les paramètres nous permettant de dire si le système peut servir la clientèle sans être surchargé ?*
- *Une période est appelée « **occupée** »tant qu'il y a des clients présents soit dans la file d'attente, soit dans le service. Une période est appelée« **morte** »si le système est vide. Peut-on envisager de réduire les périodes mortes ? (du point de vue de l'entreprise qui propose le service, c'est intéressant) On peut se demander quelle est la longueur moyenne d'une période morte.*
- *Le **temps d'attente** est le temps passé par le client dans la file d'attente. Bien que le temps d'attente moyen puisse être petit, si son écart-type est grand il se peut que des clients aient à attendre très longtemps avant d'être servis. On peut se demander la probabilité qu'un client ait à attendre plus d'un certain temps fixé à l'avance.*
- *Pour aller plus loin : certains clients se découragent et partent si la file d'attente est trop longue. On peut se demander le nombre de comptoirs de services à mettre en place pour minimiser la perte de clients (pour que la probabilité qu'il y ait plus de k clients en file soit inférieure à un nombre fixé).*

Propriété 3.2 ***(Lemme des réveils)*** *Si S et T sont des variables aléatoires indépendantes de loi exponentielle de paramètres λ et μ, alors la variable aléatoire $\min(S,T)$ est de loi exponentielle de paramètre $\lambda+\mu$. De plus :*

$$P(S<T) = \frac{\lambda}{\lambda+\mu} \text{ et } P(S>T) = \frac{\mu}{\lambda+\mu}$$

On peut alors écrire, en n'oubliant pas que la loi exponentielle est sans mémoire, en supposant que S représente les délais d'arrivées et T les temps de services :

- si le système est vide, alors il y a une arrivée dans un temps exponentiel de paramètre λ

- sinon, il y a une **arrivée** si le temps de service est supérieur au délai d'arrivée (la loi exponentielle est sans mémoire) et la probabilité d'un tel événement est $\frac{\lambda}{\lambda+\mu}$.
- il y a un **départ** si le temps de service est inférieur au délai d'arrivée et la probabilité d'un tel événement est $\frac{\mu}{\lambda+\mu}$

Propriété 3.3 *Le système M/M/1 est alors un PNM avec $\lambda_i = \lambda$ et $\mu_i = \mu$.*

Définition 3.4 *La loi π de la variable aléatoire ξ est dite **stationnaire** si toutes les variables aléatoires $(N_t)_{t\geq 0}$ ont pour loi π.*

Dans ce cas, la loi de N_t ne dépend pas du temps t : si k est un entier quelconque alors $P(N_t = k) = \pi_k$.

Dans toute la suite, on pose $\rho = \frac{\lambda}{\mu}$. C'est l'**intensité du trafic**.

Propriété 3.5

- *si $\rho < 1$, alors il existe une (unique) loi stationnaire π définie pour tout entier positif k par $\pi_k = (1-\rho)\rho^k$*
- *si $\rho > 1$, alors il n'existe pas de loi stationnaire et la taille N_t du système tend vers $+\infty$ quand t tend vers $+\infty$.*

Remarque 3.6

- *On est dans le cas où $\rho > 1$ lorsque l'intensité λ (nombre moyen théorique d'arrivées par unité de temps) est supérieur à μ (nombre moyen théorique de clients servis par unité de temps). A ce moment là la longueur de la file d'attente ne cessera de s'accroitre.*

- *Dans le cas où $\rho < 1$, on peut montrer que quelque soit la population initiale ξ, le système se stabilise toujours : pour tout entier k, $\lim\limits_{t\to+\infty} P(N_t = k) = \pi_k$.*

*Ceci explique que l'on utilise couramment la loi stationnaire explicitée ci-dessus comme loi de ξ (population initiale) et on parlera de **régime stationnaire**. Observons en particulier qu'en régime stationnaire, $\pi_0 = 1-\rho$ est la probabilité que le système soit vide et P(le système est occupé) $= \rho$.*

Propriété 3.7 *En régime stationnaire, on a donc $\lambda < \mu$ et :*

1. ***Taille et temps d'attente pour le système (File d'attente + Service) :***
   - *L'espérance de N_t ne dépend pas de t et vaut $E(N_t) = \frac{\rho}{1-\rho} = \frac{\lambda}{\mu-\lambda}$ (taille moyenne théorique du système)*
   - *Si on note T le temps passé par un client dans le système, alors T suit une loi exponentielle de paramètre $\mu-\lambda$, autrement dit $P(T \leq t) = 1 - e^{-(\mu-\lambda)t}$*

   En particulier, $E(T) = \frac{1}{\mu-\lambda}$ (temps moyen théorique passé dans le système).
2. ***Taille et temps d'attente pour la file d'attente seule :***
   - *Si on note $\widehat{N_t}$ le nombre de clients en attente dans la file à l'instant t, on a $\widehat{N_t} = \max(N_t - 1; 0)$ et $E(\widehat{N_t}) = \frac{\rho^2}{1-\rho} = \rho E(N_t)$ (nombre moyen théorique de clients dans la file d'attente) (indépendant de t)*

   En effet, la file d'attente est vide s'il y a 1 client ou moins dans le système, et elle contient $N_t - 1$ clients s'il y a plus d'un client dans le système.
   - *Si on note $\widehat{T}$ le temps d'attente de clients dans la file, alors on peut montrer que $E(\widehat{T}) = \frac{\rho}{\mu-\lambda} = \rho E(T)$ (temps d'attente moyen théorique dans la file d'attente)*

3.2 Exercices

Dans tous ces exercices, on utilise le système M/M/1 que l'on suppose en régime stationnaire.

Exercice 4 *Un client arrive en moyenne toutes les 12 minutes et la durée moyenne du service est de 8 minutes.*

1. *Calculer la probabilité :*
   (a) *que le système soit occupé*
   (b) *qu'il y ait 5 personnes dans le système*

   (c) *qu'il y ait 2 clients dans la file d'attente*

2. *Calculer la probabilité p qu'il y ait au moins deux clients dans la file d'attente.*
3. *Que se passe-t-il si l'intensité des entrées est doublée ?*

Exercice 5 *Chaque heure, 20 clients arrivent et la probabilité d'être servi sans attendre est 0,5.*

1. *Calculer le temps de service moyen.*
2. *Calculer la probabilité de clients qui arrive trouve devant lui une file de cinq personnes.*
3. *Même question avec n personnes.*

Exercice 6 *Une clinique dispose d'un service d'urgence tenu par un seul médecin. Les malades se présentent selon un processus de Poisson de taux égal à 96 clients par jour (24 heures). Les durées de soins sont indépendantes et suivent une loi exponentielle de moyenne égale à 12 minutes pour chaque malade. Les malades sont soignés dans le cabinet du médecin suivant l'ordre d'arrivée et il n'y a pas de limitation de place dans le service d'urgence. On considère le système associé au nombre de malades présents à l'instant t.*

1. *Donner la notation de Kendall de ce système en précisant les paramètres.*
2. *Montrer l'existence de la loi stationnaire et calculer la probabilité qu'il y ait 10 malades dans le système en régime stationnaire.*
3. *En supposant le régime stationnaire, déterminer les paramètres suivants :*
   (a) *le nombre moyen de malades dans le système*
   (b) *le nombre moyen de malades en attente*
   (c) *le temps moyen de présence dans le système*
   (d) *le temps moyen d'attente.*

Exercice 7 *Un organisme public est ouvert chaque jour ouvrable de 9 h à 17 h sans interruption. Il accueille en moyenne 64 usagers par jour. Un guichet unique sert à traiter le dossier de chaque usager, ceci en un temps moyen de 2,5 minutes. Les usagers si nécessaire font la queue dans l'ordre de leur arrivée, même si la queue est importante on ne refuse aucun usager. Une étude statistique a permis de conclure que la durée aléatoire des services suit une loi exponentielle et que le régime des arrivées des usagers forme un processus de Poisson. On étudie ici le système formé du guichet et de la file d'attente.*

1. *Donner l'expression de la probabilité invariante π, ainsi que la justification de son existence.*
2. *Quel est le temps moyen passé à attendre dans l'organisme par chaque usager ?*
3. *Quelles sont les probabilités qu'il n'arrive aucun client entre 15 h et 16 h ? Que 6 clients arrivent entre 16 h et 17 h ?*
4. *Quelle est, en moyenne et par heure, la durée pendant laquelle l'employé du guichet ne s'occupe pas des usagers ?*
5. *Quelle est la probabilité d'observer une file d'attente de 4 usagers, derrière celui en cours de service ?*

4 Le système M/M/2

4.1 Description comme PNM

- les arrivées sont poissoniennes d'intensité λ et les temps de service sont exponentiels de paramètre μ
- le système dispose de deux serveurs
- ξ est la taille du système à l'instant $t = 0$ et N_t est la taille du système à l'instant t

4.2 Analyse du système M/M/2

- si le système est vide, alors il y a une arrivée dans un temps exponentiel de paramètre λ
- si le système comporte un client, alors ce client peut partir avant une nouvelle arrivée avec la probabilité $\frac{\mu}{\lambda+\mu}$, ou encore être là quand arrive un nouveau client avec la probabilité $\frac{\lambda}{\lambda+\mu}$.
- si le système comporte au moins 2 clients, alors les deux serveurs sont occupés. Le premier client le sera donc suivant un temps exponentiel de paramètre 2μ.

- ce client peut partir avant une nouvelle arrivée avec la probabilité $\dfrac{2\mu}{\lambda+2\mu}$
- ou encore être là quand arrive un nouveau client avec la probabilité $\dfrac{\lambda}{\lambda+2\mu}$

Propriété 4.1 *Le système M/M/2 est un PNM avec :*

- $\lambda_i = \lambda$ *pour tout entier naturel* i
- $\mu_1 = \mu$ *et* $\mu_i = 2\mu$ *pour* $i \geq 2$.

Propriété 4.2 *On rappelle que* $\rho = \dfrac{\lambda}{\mu}$

1. *Si* $\rho < 2$ *alors il existe une (unique) loi stationnaire* π *définie par les* ***formules d'Erlang*** *:*
   $\pi_0 = \dfrac{2-\rho}{2+\rho}$, $\pi_1 = \dfrac{\rho(2-\rho)}{2+\rho}$ *et pour tout entier* $k \geq 2$ $\pi_k = \dfrac{2(2-\rho)}{2+\rho}\left(\dfrac{\rho}{2}\right)^k$
2. *Si* $\rho > 2$ *alors il n'existe pas de loi stationnaire et la taille* N_t *du système tend vers* $+\infty$ *quand* t *tend vers* $+\infty$.

Remarque 4.3 *Avec le même principe, on généralise ces notations aux systèmes* $M/M/n$, $n \geq 2$.

Exercice 8

1. *Avec les notations du cours, déterminer la probabilité que les deux serveurs soient occupés.*
2. *En déduire que E(nombre de serveurs occupés)* $= \rho$