ISFA
GRANDE ÉCOLE D'ACTUARIAT ET DE GESTION DES RISQUES
INSTITUT DE SCIENCE FINANCIÈRE ET D'ASSURANCES

MODÈLES DE DURÉE

Support de cours 2024-2025

Processus poissoniens et files d'attente

Frédéric PLANCHET

Version 4.7

Août 2024

SOMMAIRE

1. Le processus de Poisson

Le processus de Poisson est un outil extrêmement utilisé dans les phénomènes de comptage. Il apparaît en effet naturellement dans ces situations, comme on va le voir ci-après.

1.1. Définition et premières propriétés

On considère des évènements qui se produisent à des dates aléatoires, et on s'intéresse, pour tout $t > 0$ au nombre d'évènements qui se sont produits au cours de l'intervalle $]0, t]$, que l'on note $N(t)$.

On définit alors le *processus de Poisson d'intensité* $\lambda > 0$ comme un processus $N(t)$ satisfaisant les conditions suivantes :

- ✓ P1 : le processus est à accroissements indépendants, i.e. $\forall\, 0 \le t_1 < t_2 .. < t_n$, les variables aléatoires $N(t_i) - N(t_{i-1})$, $i = 1,..,n$ sont globalement indépendantes ;
- ✓ P2 : le processus est à accroissements stationnaires, i.e. $\forall\, t, h > 0$, la loi de $N(t+h) - N(t)$ ne dépend que de h ;
- ✓ P3 : $P\big(N(t+h) - N(t) \ge 1\big) = \lambda h + o(h)$ et $P\big(N(t+h) - N(t) \ge 2\big) = o(h)$

La condition P3 signifie que sur un petit intervalle de temps, on observe 0 ou 1 événement, et que la probabilité d'observer un événement est proportionnelle au temps écoulé. On peut remarquer que $N(0) = 0$.

Les trajectoires sont par définition croissantes, continues à droite avec une limite à gauche. Elles croissent par sauts d'une unité :

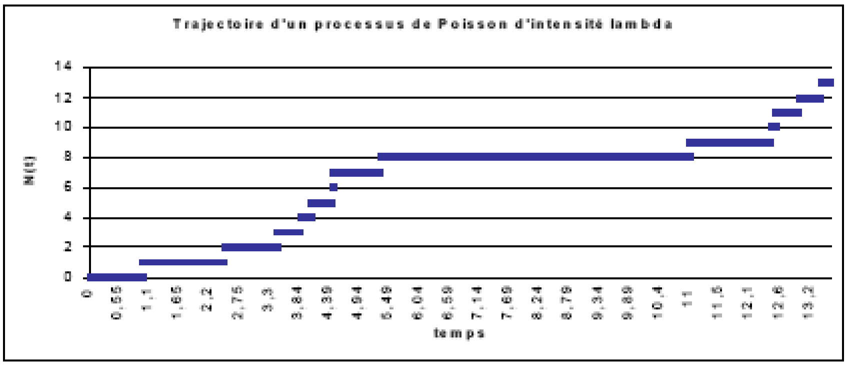


Fig. 1 : Trajectoire d'un processus de Poisson

On vérifie par ailleurs sans difficulté que si (N_t^1) et (N_t^2) sont deux processus de Poisson indépendants d'intensités respectives λ_1 et λ_2, alors $N_t = N_t^1 + N_t^2$ est un processus de

Poisson d'intensité $\lambda_1 + \lambda_2$. Pour le vérifier on remarque que N_t est un processus à accroissements indépendants et stationnaire, et qu'il vérifie également P3.

Ce résultat admet la réciproque suivante : si N_t est un processus de Poisson et que T_n désigne l'instant du $n^{\text{ième}}$ saut, et si on se donne des variables aléatoires de Bernoulli X_n indépendantes des instants de saut et de paramètre *p*, on peut construire deux processus en affectant les sauts de N_t à l'un si $X_n = 1$, à l'autre sinon ; les 2 processus ainsi obtenus sont des processus de Poisson indépendants, de paramètres respectifs $p\lambda$ et $(1-p)\lambda$.

1.2. Lien avec la loi de Poisson

Le processus de Poisson vérifie :

<u>Propriété</u> : pour tout $t > 0$ la variable $N(t)$ suit une loi de Poisson de paramètre λt, i.e. :

$$P(N_t = n) = e^{-\lambda t} \frac{(\lambda t)^n}{n!}$$

<u>Démonstration</u> : On considère tout d'abord le cas de $n = 0$; par croissance de $N(t)$ et P1 on a

$$P(N_{t+h} = 0) = P(N_t - N_0 = 0 \cap N_{t+h} - N_t = 0) = P(N_t - N_0 = 0)P(N_{t+h} - N_t = 0)$$

En utilisant P3 puis en divisant chaque terme de l'égalité ci-dessus par *h* on obtient :

$$\frac{1}{h}\left[P(N_{t+h} = 0) - P(N_t = 0)\right] = P(N_t = 0)\frac{-\lambda h + o(h)}{h}$$

En passant à la limite lorsque $h \to 0$, on obtient, en notant $p_n(t) = P(N_t = n)$ l'équation :

$$\frac{dp_0(t)}{dt} = -\lambda p_0(t)$$

avec la condition initiale $p_0(0) = 1$; on en déduit immédiatement que $p_0(t) = e^{-\lambda t}$.

On peut maintenant traiter le cas général ; on observe alors que :

$$P(N_{t+h} = n) = \sum_{k=0}^{n} P\left[N_t - N_0 = n-k \cap N_{t+h} - N_t = k\right] = \sum_{k=0}^{n} P\left[N_t = n-k\right] P\left[N_{t+h} - N_t = k\right]$$

Cette égalité devient en exploitant P2 et P3 :

$$p_n(t+h) = p_n(t)\left(1 - \lambda h + o(h)\right) + p_{n-1}(t)\left(\lambda h + o(h)\right) + \left[\sum_{k=2}^{n} p_{n-k}(t)\right] o(h) \ ;$$

comme précédemment on trouve en passant à la limite lorsque $h \to 0$:

$$\frac{dp_n(t)}{dt} = \lambda\left[p_{n-1}(t) - p_n(t)\right]$$

avec là encore la condition aux limites $p_n(0) = 0$. On pose alors $f_n(t) = e^{\lambda t} p_n(t)$, et par différenciation on trouve :

$$\frac{df_n(t)}{dt} = \lambda f_{n-1}(t) \text{ et } f_n(0) = 0$$

On achève alors la démonstration par récurrence.

Remarque : on peut obtenir le résultat de manière heuristique en remarquant que si on découpe l'intervalle $]0,t]$ en n sous-intervalles suffisamment petits pour que chacun contienne 0 ou 1 saut, alors la loi de N_t est la loi binomiale $B\left(n, \frac{\lambda t}{n}\right)$; or cette loi converge vers $P(\lambda t)$ lorsque $n \to \infty$.

Ce résultat justifie l'apparition naturelle de la loi de Poisson pour modéliser des évènements « rares ».

1.3. L'apparition de la loi exponentielle

Lorsque l'on observe un processus de Poisson, il est naturel de s'intéresser au temps d'attente entre les sauts ; on a alors le résultat fondamental suivant :

Propriété : Si T_n désigne l'instant du $n^{\text{ième}}$ saut, alors les variables $T_n - T_{n-1}$ sont iid de loi exponentielle de paramètre $\lambda > 0$.

Ce résultat justifie la place particulière de la loi exponentielle dans l'étude des modèles de durée. On en déduit que la loi de T_n est la loi d'Erlang de paramètres (n, λ).

Démonstration : On a l'identité entre évènements $\{N_t < n\} = \{T_n > t\}$, qui implique que T_n est un temps d'arrêt pour le processus $N(t)$; appliquée à $n = 1$ elle conduit à $P(T_1 > t) = P(N_t = 0) = e^{-\lambda t}$; de même, en utilisant la propriété de Markov forte :

$$P(T_n - T_{n-1} > t) = P\left(N_{T_{n-1}+t} - N_{T_{n-1}} = 0\right) = e^{-\lambda t},$$

ce qui établit le résultat pour la loi ; l'indépendance découle également directement de l'application de la propriété de Markov forte.

Par ailleurs on peut remarquer que $N_t = \max\{n\,; T_n \le t\}$, ce qui met le processus de Poisson sous la forme d'un processus de comptage associé à un processus de renouvellement. L'application de ce formalisme au cas du processus de Poisson permet de proposer un algorithme simple pour simuler des trajectoires d'un processus de Poisson, comme on va le voir ci-après.

1.4. Distribution des sauts d'un processus de Poisson

On s'intéresse à la distribution jointe de la statistique d'ordre associée à un *n*-échantillon iid ; on a le résultat suivant :

Proposition : si les variables $(X_1, \ldots, X_n)$ sont iid avec la densité *f*, alors la densité du vecteur $(X_{(1)}, \ldots, X_{(n)})$ s'écrit $f^*(x_1, \ldots, x_n) = n! \prod_{i=1}^{n} f(x_i) 1_{\{x_1 < \ldots < x_n\}}$.

Démonstration : pour une permutation σ de $\{1,\ldots,n\}$, on note φ_σ la transformation de R^n qui permute les coordonnées de $(x_1,\ldots,x_n)$. On a pour tout borélien $A \subset \{x_1 < \ldots < x_n\}$:

$$P\left[\left(X_{(1)},\ldots,X_{(n)}\right)\in A\right] = \sum_{\sigma} P\left[\left(X_1,\ldots,X_n\right)\in \phi_\sigma(A)\right] = \sum_{\sigma} \int_{\phi_\sigma(A)} f(x_1)\ldots f(x_n)dx_1\ldots dx_n$$

On conclut par le changement de variable $(y_1,\ldots,y_n) = \phi_\sigma^{-1}(x_1,\ldots,x_n)$ et en remarquant qu'il y a *n*! permutations de $\{1,\ldots,n\}$.

On utilise maintenant ce résultat dans le cas du processus de Poisson pour démontrer le résultat suivant :

Proposition : La loi conditionnelle des instants de saut $(T_1,\ldots,T_n)$ sachant $\{N_t = n\}$ est celle d'un échantillon uniforme sur $[0,t]$ classé par ordre croissant ; la densité de ce vecteur est donnée par $f_n(t_1,\ldots,t_n) = \frac{n!}{t^n} I_{\{0<t_t\ldots<t_n\le t\}}$.

Ce résultat est utile en pratique car il fournit un moyen simple de simuler des trajectoires d'un processus de Poisson ; en effet, il suffit de simuler une réalisation *n* d'une loi de Poisson de paramètre λt, puis de simuler *n* réalisations d'une loi uniforme sur $[0,t]$; on ordonne alors ces variables dans l'ordre croissant, ce qui donne les instants de saut du processus.

Démonstration : on remarque que l'événement $\{T_i = t_i\ , 1 \le i \le n\ ; N_t = n\}$ est égal à l'événement $\{T_i - T_{i-1} = t_i - t_{i-1}\ , 1 \le i \le n\ ; T_{n+1} - T_n > t - t_n\}$; mais les temps d'attente entre les sauts sont des variables aléatoires iid de loi exponentielle de paramètre λ, ce qui implique facilement le résultat en écrivant que[1] :

$$P\left(T_i = t_i\ , 1 \le i \le n \,\middle|\, N_t = n\right) = \frac{P\left(T_i = t_i\ , 1 \le i \le n \cap N_t = n\right)}{P\left(N_t = n\right)}$$

et en se souvenant que $P\left(N_t = n\right) = e^{-\lambda t}\frac{(\lambda t)^n}{n!}$.

Outre son intérêt en termes de simulation, ce résultat peut être utile pour calculer des intégrales associées aux instants de saut du processus ; on a ainsi :

Proposition : si *g* est intégrable sur $[0,t]$, alors $E\left(g(T_1)\times\ldots\times g(T_n)\middle|N_t = n\right) = E\left[g(U)\right]^n$, où *U* est une variable uniforme sur $[0,t]$.

Démonstration : le résultat est une conséquence directe de la proposition précédente.

[1] Par abus de notation on note $f_T(t) = P(T = t)$ au lieu de $f_T(t)dt = P(t \le T \le t + dt)$.

Si l'on fixe un découpage $0 = t_0 < t_1 < t_2 \ldots < t_k = t$ en k sous-intervalles de $[0,t]$, conditionnellement à $\{N_t = n\}$ on déduit de ce qui précède que, comme la probabilité que U_i tombe dans l'intervalle $[t_{i-1},t_i[$ est $\frac{t_i - t_{i-1}}{t}$, la répartition des sauts dans les sous-intervalles suit une loi multinomiale définie par :

$$P\left(N_{t_i} - N_{t_{i-1}} = n_i \,; 1 \le i \le k \,|N_t = n\right) = \frac{n!}{n_1! \ldots n_k!} \prod_{i=1}^{k} \left(\frac{t_i - t_{i-1}}{t}\right)^{n_i}$$

avec $n_1 + \ldots + n_k = n$. En particulier, si $s < t$, la loi de N_s sachant $\{N_t = n\}$ est une loi binomiale de paramètres *n* et $\frac{s}{t}$.

On peut également obtenir la loi (libre) des *n* premiers instants de saut $(T_1,\ldots,T_n)$; on a ainsi le résultat :

<u>Proposition</u> : La loi des *n* premiers instants de saut $(T_1,\ldots,T_n)$ est donnée par $f_n(t_1,\ldots,t_n) = \lambda^n e^{-\lambda t_n} 1_{\{0<t_i..<t_n\}}$.

<u>Démonstration</u> : D'après 1.3 ci-dessus, les variables $X_i = T_i - T_{i-1}$ sont iid de loi exponentielle de paramètre λ. De plus la loi de $(T_1,\ldots,T_n)$ se déduit de celle de $(X_1,\ldots,X_n)$ par le changement de variable $(x_1,..,x_n) \to (t_1,..,t_n)$ défini par $t_k = x_1 + \ldots + x_k$, soit $x_k = t_k - t_{k-1}$; le jacobien de cette transformation vaut 1 (transformation linéaire triangulaire) et on a donc pour toute fonction borélienne intégrable :

$$E\left[h(T_1,\ldots,T_n)\right] = \int h(x_1, x_1 + x_2, \ldots, x_1 + \ldots + x_x) \prod_{k=1}^{n} \lambda e^{-\lambda x_k} dx_1 \ldots dx_n$$

$$= \int h(t_1,\ldots,t_x) \lambda^n e^{-\lambda t_n} 1_{\{0<t_i \ldots <t_n\}} dt_1 \ldots dt_n$$

ce qui achève la démonstration.

1.5. Martingales associées au processus de Poisson

<u>Proposition</u> : les processus $M_t = N_t - \lambda t$ et $M_t^2 - \lambda t$ sont des martingales relativement à la filtration $\sigma(N_s, s \le t)$.

Pour $\theta > 0$, le processus $\chi_t = \mathbf{exp}\left(-\theta N_t + \lambda t\left(1 - e^{-\theta}\right)\right)$ est également une martingale relativement à la filtration $\sigma(N_s, s \le t)$.

<u>Démonstration</u> : le résultat pour la martingale exponentielle s'obtient en écrivant :

$$E_s(\chi_t) = E_s\left[\mathbf{exp}\left(-\theta N_t + \lambda t\left(1 - e^{-\theta}\right)\right)\right] = E_s\left[\mathbf{exp}\left(-\theta(N_t - N_s) + \lambda(t-s)\left(1 - e^{-\theta}\right)\right)\right] \times \chi_s$$

Mais $N_t - N_s$ est indépendant de $\sigma(N_u, u \le s)$ et de loi la loi de Poisson de paramètre $\lambda(t-s)$; l'espérance conditionnelle de l'expression ci-dessus est donc égale à un en se souvenant que la transformée de Laplace d'une loi de Poisson de paramètre λ s'écrit $Ee^{-uX} = \mathbf{exp}(\lambda[e^{-y} - 1])$, ce qui achève la démonstration.

Pour le processus M_t, on écrit que $E_s[(N_t - N_s)] = \lambda(t-s)$ par la propriété d'accroissements indépendants ; mais on a aussi $E_s[(N_t - N_s)] = E_s[N_t] - N_s$; en égalant les 2 formes de l'expression on obtient $E_s[N_t] - \lambda t = N_s - \lambda s$, ce qui prouve le résultat.

Le raisonnement est analogue pour $M_t^2 - \lambda t$.

Il découle aisément de cette proposition que $\lim\limits_{t \to +\infty} \frac{N_t}{t} = \lambda$, dans L^2. En fait la convergence est même presque sûre. La convergence presque sûre est une conséquence de la loi des grands nombres appliquée aux accroissements ; plus précisément, si t est entier naturel alors on écrit :

$$\frac{N_t}{t} = \frac{(N_t - N_{t-1}) + .. + (N_1 - N_0)}{t}$$

Le résultat est donc prouvé dans ce cas ; pour t réel, il suffit d'écrire que $\frac{[t]}{t}\frac{N_{[t]}}{[t]} \leq \frac{N_t}{t} \leq \frac{[t]+1}{t}\frac{N_{[t]+1}}{[t]+1}$.

Le comportement asymptotique en loi d'un processus de Poisson est décrit par le résultat suivant :

Proposition : $\frac{N_t - \lambda t}{\sqrt{\lambda t}} \underset{t \to \infty}{\to} \mathrm{N}(0,1)$.

Démonstration : posons $Z_t = \frac{N_t - \lambda t}{\sqrt{\lambda t}}$, on a

$$E\left[\mathbf{exp}(-\theta Z_t)\right] = \mathbf{exp}\left[\lambda t\left(e^{-\frac{\theta}{\sqrt{\lambda t}}} - 1\right) + \theta\sqrt{\lambda t}\right]$$

en utilisant l'expression de la transformée de Laplace d'une loi de Poisson rappelée ci-dessus ; on conclut en effectuant un développement limité à l'ordre 2 en $\frac{\theta}{\sqrt{\lambda t}}$ que $\lim\limits_{t \to +\infty} E\,\mathbf{exp}(-\theta Z_t) = \mathbf{exp}\left(\frac{\theta^2}{2}\right)$, ce qui prouve le résultat.

1.6. Le processus de Poisson non homogène

Dans la propriété P3 caractérisant le processus de Poisson, on peut introduire une intensité fonction du temps, $\lambda(t)$. On obtient un processus qui n'est plus stationnaire, la loi de l'accroissement $N(t+h)-N(t)$ est alors une loi de Poisson de paramètre $\int_t^{t+h} \lambda(u)\,du$.

Exemple : Pour modéliser l'arrivée des clients dans une agence bancaire, on peut imaginer que les arrivées sont d'abord espacées, puis qu'en milieu de journée le rythme augmente, pour décroître ensuite. On pourra par exemple prendre[2] :

$$\lambda(t) = \begin{cases} t(6-t) \; 0 \le t \le 6 \\ 0 \; t > 6 \end{cases}$$

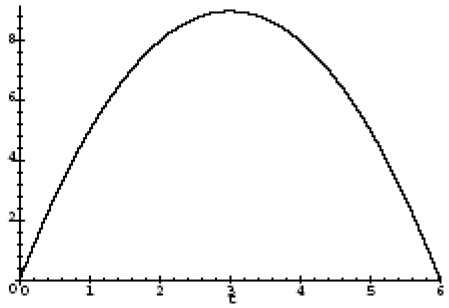


Fig. 2 : *Intensité parabolique*

Le temps d'attente entre deux clients n'est plus en général une loi exponentielle.

2. Rappels sur les processus markoviens

Dans la plupart des situations qui mettent en jeu l'analyse d'une durée, il s'agit d'une durée dans un état donné (vivant, au chômage, en incapacité, en fonctionnement, *etc.*) ; les chaînes de Markov fournissent un cadre d'analyse naturel de ces phénomènes et des transitions d'un état vers un autre.

Les paragraphes qui suivent ne constituent pas un cours sur les chaînes de Markov, mais un simple rappel des résultats utiles dans le cadre de l'analyse des durées (temps de passage dans un état). Pour un cours complet sur le sujet, le lecteur peut se reporter à KARLIN et TAYLOR [1975].

2.1. Chaînes de Markov en temps discret[3]

2.1.1. Définition et propriétés de base

On considère ici un processus avec un nombre d'états fini[4], égal à *n*, et on note $P = (p_{i,j})_{1 \le i,j \le n}$ la matrice de transition du système ; $p_{i,j}$ est la probabilité de transition de

[2] L'unité de temps est l'heure.

[3] On ne s'intéresse ici qu'aux chaînes homogènes dans lesquelles la matrice de transition est indépendante de t.

[4] La plupart des résultats énoncés ici s'adaptent au cas d'un ensemble d'états dénombrable.

l'état *i* à l'état *j*, ce qui implique que la somme des éléments de chaque ligne de *P* est égale à 1. Si $\pi_i(t) = P(X_t = i)$ désigne la probabilité pour que le système soit dans l'état *i* à la date *t*, alors on obtient les par récurrence à l'aide du produit matriciel :

$$\pi(t+1) = \pi(t)P$$

avec $\pi = (\pi_i)_{1 \le i \le n}$ le vecteur (ligne) des probabilités d'état du système.

La distribution des états du système est donc entièrement déterminée par la donnée de la matrice de transition et de la condition initiale, et on obtient par récurrence à partir de l'équation de transition :

$$\pi(t) = \pi(0)P^t$$

En d'autres termes la matrice de transition du processus entre 0 et t, P_t, est égale à P^t (et vérifie l'équation de Kolmogorov $P_{t+s} = P_t P_s$).

Une manière commode de résoudre cette équation en évitant le calcul de puissances de la matrice *P* est de calculer la fonction génératrice des composantes de $\pi = (\pi_i)_{1 \le i \le n}$ vues comme des fonctions de *t* ; on obtient ainsi un vecteur $\pi = (\pi_i)_{1 \le i \le n}$. On considère alors :

$$\hat{\pi}_i(z) = \sum_{t=0}^{+\infty} \pi_i(t) z^t$$

On obtient en appliquant cette transformation à l'équation de récurrence $\pi(t+1) = \pi(t)P$, on trouve $\frac{1}{z}(\hat{\pi}(z) - \pi(0)) = \hat{\pi}(z)P$, et donc :

$$\hat{\pi}(z) = \pi(0)(Id - zP)^{-1}$$

<u>Exemple</u> : soit un dispositif composé de 2 machines fonctionnant en parallèle, de manière indépendante, la probabilité de panne au cours d'une journée étant pour chaque machine de $1-p$. À la date initiale les 2 machines fonctionnent, et on suppose qu'il n'y a pas de possibilité de réparation ; l'évolution du système peut être décrite par le « graphe de transition » ci-après :

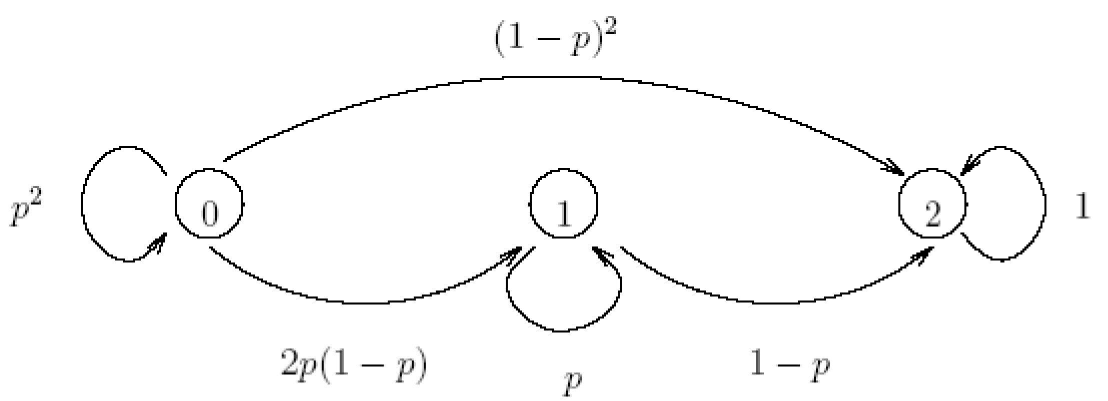


Fig. 3 : *Diagramme d'état*

Il s'agit donc un système à 3 états, un état étant caractérisé par le nombre de machines en fonctionnement. La matrice de transition s'écrit (en se souvenant que la ligne n°*i* représente la distribution conditionnelle sachant qu'on part de l'état *i*) :

$$P = \begin{bmatrix} p^2 & 2p(1-p) & (1-p)^2 \\ 0 & p & 1-p \\ 0 & & 1 \end{bmatrix}$$

Dans le cas où il y a possibilité de réparation d'une machine à la fois, le système se réduit à un système à deux états (il n'est pas possible d'avoir les 2 machines en panne en même temps).

2.1.2. Distribution stationnaire

On s'intéresse ici aux distributions constantes au cours du temps, *i.e.* telles que :

$$\pi = \pi P$$

Ces distributions sont les « distributions candidates » pour être une distribution limite de la chaîne, si celle-ci existe. En pratique on a le résultat suivant :

Proposition : Une chaîne de Markov à espace d'états fini admet au moins une distribution stationnaire ; si de plus la chaîne est apériodique[5] et irréductible[6] alors la distribution limite (lorsque $t \to +\infty$) existe et $\pi(\infty)$ est l'unique distribution stationnaire de la chaîne.

On a dans ce cas un lien fort avec le temps moyen passé dans un état : le système est dit *ergodique*, au sens où le pourcentage de temps passé par le processus dans un état est égal à la probabilité de cet état dans la distribution limite (stationnaire). L'ergodicité s'exprime formellement par :

$$\lim_{t \to +\infty} \frac{S_j(t)}{t} = \pi_j(\infty)$$

où $S_j(t)$ désigne le temps passé dans l'état *j* avant t : la moyenne temporelle et la moyenne spatiale coïncident.

2.1.3. Le temps de séjour dans un état

On traite d'abord le cas simple d'un processus à 2 états, 0 et 1, avec une probabilité de transition *p* :

[5] Une chaîne de Markov est apériodique si le PGCD de la longueur de tous les cycles de probabilité non nulle est égal à 1, un cycle étant l'évolution de la file entre deux instants où elle est vide.
[6] On dit qu'une chaîne de Markov est irréductible si tout état est atteignable à partir de tout autre état par des transitions de probabilités strictement positives.

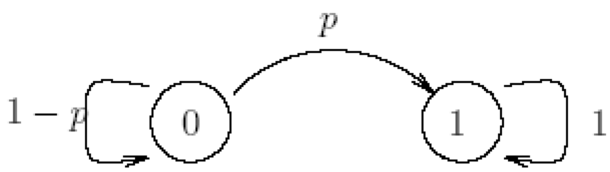


Fig. 4 : *Diagramme d'état*

On note *T* la durée de séjour dans l'état initial ; on a évidemment $P(T=k)=(1-p)^{k-1}p$, ce qui montre que *T* est distribué selon une loi géométrique de paramètre *p*. Le temps moyen de séjour dans l'état initial est donc $E(T)=\frac{1}{p}$.

Dans le cas général, le calcul du temps de séjour dans l'état *i* se ramène au cas précédent, en considérant $p=1-p_{i,i}$ la somme des probabilités d'aller vers un autre état que *i*. On obtient alors que le temps de séjour dans un état est distribué selon une loi géométrique de paramètre $p=1-p_{i,i}$, et en particulier $E(T)=\frac{1}{1-p_{i,i}}$.

2.2. Chaînes de Markov en temps continu

Formellement, tout se passe comme dans le cas précédent, mais le temps est maintenant indicé par $t>0$. On continue à ne considérer que le cas des chaînes homogènes, pour lesquelles les probabilités de transition $P\left(X_{t+u}=j \mid X_t=i\right)$ sont indépendantes de *t*.

2.2.1. Générateur infinitésimal

On considère un système dynamique dont l'évolution est gouvernée par une chaîne de Markov en temps discret, le pas de temps étant supposé arbitrairement petit, $\varepsilon>0$. La matrice de transition $P(\varepsilon)$ est proche de l'identité car on suppose que le système « évolue peu » pendant le temps $\varepsilon>0$. On suppose donc que l'on a au premier ordre :

$$P(\varepsilon)=Id+\varepsilon A+o(\varepsilon)$$

L'équation d'état du système $\pi(t+\varepsilon)=\pi(t)P(\varepsilon)$ devient

$$\frac{1}{\varepsilon}\left(\pi(t+\varepsilon)-\pi(t)\right)=\pi(t)A+o(\varepsilon),$$

et en passant à la limite on obtient l'équation de Chapman-Kolmogorov « avant » :

$$\frac{d\pi(t)}{dt}=\pi(t)A$$

La matrice *A* est le *générateur infinitésimal* de la chaîne. Par construction, il s'agit d'une matrice carrée dont la somme des termes de chaque ligne est nulle, avec des termes positifs ou nuls en dehors de la diagonale.

On vérifie que la solution de l'équation d'état est de la forme $\pi(t) = \pi(0)e^{tA}$, avec par définition $e^{tA} = \sum_{n=0}^{+\infty} \frac{t^n}{n!} A^n$.

Réciproquement, lorsque l'on veut discrétiser une chaîne de Markov initialement définie en temps continu, on déduit de la formule $\pi(t) = \pi(0)e^{tA}$ que $\pi(t+1) = \pi(t)e^{A}$, ce qui prouve que e^A est la matrice de transition de la chaîne discrétisée.

2.2.2. Distribution stationnaire

En régime stationnaire, l'équation d'état devient $\pi A = 0$.

2.2.3. Temps de séjour dans un état

Le résultat de la section 2.1.3 ci-dessus se généralise en temps continu de la manière suivante :

Proposition : soit (X_t) un processus de sauts Markovien et T_x le temps d'arrêt représentant le temps écoulé avant de quitter le point *x*, alors il existe pour tout *x* une constante $\lambda(x)$ telle que T_x suive une loi exponentielle de paramètre $\lambda(x)$.

La fonction $\lambda(x)$ s'appelle l'intensité du processus ; on peut noter que ce résultat est la transposition directe en temps continu du résultat sur les temps de séjour en temps discret, la loi exponentielle étant la version continue de la loi géométrique.

Démonstration : Le résultat découle directement de la propriété de Markov forte. On note pour simplifier $T_x = T$, et $T_s = \mathbf{inf}\{u > 0\,, X_{s+u} \neq x\}$ le temps de première sortie[7] de *x* après *s*. On a alors sur l'ensemble $\{s < T\}$, $T = s + T_s$. En utilisant le caractère markovien du processus, on en déduit que :

$$P_x[T > t+s] = P_x[T_s > t\,, T > s] = E_x\left[I_{\{T>s\}} P_x\left[T_s > t \,|\sigma(X_u\,, u \leq s)\right]\right]$$

Mais $P_x[T_s > t\,|\sigma(X_u\,, u \leq s)] = P_{X_s}\{T > t\}$ par la propriété de Markov et comme $X_s = x$ lorsque $s < T$ on a montré finalement que :

$$P_x\{T > t+s\} = P_x\{T > s\} P_x\{T > t\}$$

Cela signifie que la fonction $t \rightarrow P_x\{T > t\}$ est multiplicative ; comme elle est par ailleurs comprise entre 0 et 1 et continue à droite, elle est de la forme $\mathbf{exp}(-\lambda(x) \times t)$; il reste à montrer que $\lambda(x) < \infty$; mais si on avait pour un x $\lambda(x) = \infty$, on aurait pour tout $\varepsilon > 0$ $P_x(T > \varepsilon) = 0$, ce qui implique que $T = 0$ P_x- presque sûrement, ce qui est impossible, puisque les trajectoires sont en escalier continues à droites.

[7] T_s n'est pas un temps d'arrêt

Lorsque l'on « part » de x, le processus saute à un point X_{T_x}, et on peut montrer que T_x et X_{T_x} sont indépendants, i.e. l'endroit où on saute est indépendant du moment où le saut se produit (voir par exemple DACUNHA-CASTELLE et DUFLO [1983]).

2.3. Processus de naissance et de mort

On considère une population de taille maximum[8] n dont l'évolution entre les instants t et $t+\varepsilon$, conditionnellement au fait que l'effectif de la population est k, est définie par :

- la probabilité pour qu'une naissance ait lieu est $\lambda_k \varepsilon + o(\varepsilon)$, $k = 0,\ldots,n-1$;
- la probabilité pour qu'un décès ait lieu est $\mu_k \varepsilon + o(\varepsilon)$, $k = 1,\ldots,n$;
- les probabilités d'autres évènements sont en $o(\varepsilon)$.

Le processus de Poisson constitue un cas particulier de cette dynamique, lorsque $\lambda_k = \lambda$, $\mu = 0$ et $n = \infty$. Plus généralement, de nombreuses situations classiques de files d'attente ou de dynamiques de populations peuvent être modélisées de cette manière.

2.3.1. Le processus de naissance et de mort comme chaîne de Markov

La définition précédente caractérise un processus de Markov en temps continu dont le diagramme d'états est :

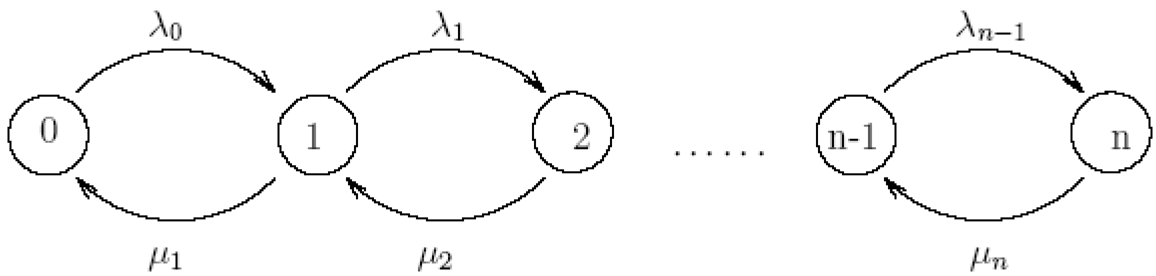


Fig. 5 : *Diagramme d'états*

La matrice de transition $P(\varepsilon)$ s'écrit donc, en posant par convention $\mu_0 = 0$ et $\lambda_n = 0$:

- $p_{i,i}(\varepsilon) = 1 - (\lambda_{i-1} + \mu_{i-1})\varepsilon + o(\varepsilon)$
- $p_{i+1,i}(\varepsilon) = \mu_i \varepsilon + o(\varepsilon)$
- $p_{i,i+1}(\varepsilon) = \lambda_i \varepsilon + o(\varepsilon)$

Ainsi dans le cas n=4 on a :

[8] On peut également considérer le cas où n est infini.

$$P(\varepsilon) = \begin{pmatrix} 1-\lambda_0\varepsilon & \lambda_0\varepsilon & 0 & 0 & 0 \\ \mu_1\varepsilon & 1-(\lambda_1+\mu_1)\varepsilon & \lambda_1\varepsilon & 0 & 0 \\ 0 & \mu_2\varepsilon & 1-(\lambda_2+\mu_2)\varepsilon & \lambda_2\varepsilon & 0 \\ 0 & 0 & \mu_3\varepsilon & 1-(\lambda_3+\mu_3)\varepsilon & \lambda_3\varepsilon \\ 0 & 0 & 0 & \mu_4\varepsilon & 1-\mu_4\varepsilon \end{pmatrix} + o(\varepsilon)$$

Le générateur infinitésimal s'écrit alors :

- $a_{i,i} = 1 - (\lambda_{i-1} + \mu_{i-1})$
- $a_{i+1,i} = \mu_i$
- $a_{i,i+1} = \lambda_i$

$$A = \begin{pmatrix} -\lambda_0 & \lambda_0 & 0 & 0 & 0 \\ \mu_1 & -(\lambda_1+\mu_1) & \lambda_1 & 0 & 0 \\ 0 & \mu_2 & -(\lambda_2+\mu_2) & \lambda_2 & 0 \\ 0 & 0 & \mu_3 & -(\lambda_3+\mu_3) & \lambda_3 \\ 0 & 0 & 0 & \mu_4 & -\mu_4 \end{pmatrix}$$

La distribution stationnaire de cette chaîne est donnée par la résolution du système linéaire $\pi A = 0$. La résolution de ce système linéaire conduit aux équations $\pi_i = \frac{\lambda_{i-1}}{\mu_i}\pi_{i-1}$, et donc, en posant $\varphi_i = \frac{\lambda_0 \cdots \lambda_{i-1}}{\mu_1 \cdots \mu_i}$ on obtient $\pi_i = \varphi_i \pi_0$; la condition $\sum_{i=0}^{n} \pi_i = 1$ conduit finalement à :

$$\pi_i = \frac{\varphi_i}{1 + \sum_{k=1}^{n} \varphi_k}$$

2.3.2. Le raisonnement différentiel

En notant comme à la section 1.2 ci-dessus $p_n(t) = P(N_t = n)$, et en raisonnant de la même manière on peut obtenir l'équation différentielle suivante :

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

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

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

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

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

3. Introduction aux files d'attente

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

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

3.1. Classification des files d'attente

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

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

La classification des systèmes d'attente fait appel à une notation symbolique comprenant quatre symboles dans l'ordre suivant[9] : A/B/s/N, avec :

- A = distribution du temps entre deux arrivées successives,
- B = distribution des durées de service,
- s = nombre de stations de service montées en parallèle,
- N = capacité du système (serveurs + file d'attente) ; lorsque $N = \infty$, on omet cet élément dans la notation.

Pour spécifier les distributions A et B, on adopte de plus la convention suivante :

- M = distribution qui vérifie la propriété de Markov, i.e. processus de Poisson pour les arrivées, loi exponentielle pour le temps de service,
- E(k) = distribution d'Erlang d'ordre k,
- G = distribution générale (on ne sait rien sur ses caractéristiques),
- D = cas déterministe (la variable ne prend qu'une seule valeur admise).

[9] Cette notation s'appelle « la notation de Kendall ».

Exemple : La notation M/D/1/4 définit un système d'attente comprenant une station-service et pour lequel la longueur maximale de la file d'attente vaut 4-1 = 3. Le processus d'arrivée est un processus de Poisson (voir la section 1 ci-dessus) et la durée du service est constante.

3.2. La file M/M/1

3.2.1. Propriétés de base

On suppose :

- que les clients arrivent dans la file selon un processus de Poisson de paramètre $\lambda > 0$;
- le temps de service est une loi exponentielle de paramètre $\mu > 0$;
- il y a un seul serveur et la file a une capacité illimitée.

Cette file est un cas particulier du processus de naissance et de mort vu à la section 2.3 ci-dessus, dans lequel les probabilités de naissance et de décès sont constantes.

En appliquant les résultats précédemment obtenus on trouve que $\varphi_i = \left(\frac{\lambda}{\mu}\right)^i$, ce qui implique que la distribution stationnaire existe si et seulement si $\lambda < \mu$ et qu'alors le nombre de clients dans le système est distribué selon une loi géométrique de raison $\varphi = \frac{\lambda}{\mu}$:

$$\pi_k = \left(1 - \frac{\lambda}{\mu}\right)\left(\frac{\lambda}{\mu}\right)^k$$

Le théorème ergodique permet alors de calculer la fraction de temps pendant laquelle le serveur est, en moyenne, occupé, $1 - \pi_0 = \frac{\lambda}{\mu}$.

On peut également calculer la loi du temps passé par un client dans la file :

Proposition : en régime stationnaire, le temps passé par un client dans la file (attente + service) est distribué selon une loi exponentielle de paramètre $\mu - \lambda$.

Démonstration : lorsqu'un client arrive dans la file et que *n* clients sont présents, il devra attendre pour sortir de la file un temps de loi d'Erlang de paramètres $(n+1, \mu)$: en effet, les n personnes présentes doivent être servies, puis il sera servi ; de plus, du fait de la propriété d'absence de mémoire d'une variable exponentielle, pour le client arrivant, le temps restant pour terminer le service de la personne en cours est une variable exponentielle de paramètre μ (propriété « PASTA », qui sera également utilisée en 3.4 ci-dessous). Si on note *T* le temps passé par le client dans la file, on a donc, en régime stationnaire :

$$P(T \le t) = \sum_{n=0}^{+\infty} \pi_n \int_0^t \frac{\mu^{n+1} x^n}{n!} e^{-\mu x} dx$$

avec $\pi_n = \left(1 - \frac{\lambda}{\mu}\right) \frac{\lambda^n}{\mu^n}$; en appliquant Fubini (qui est immédiat ici puisque les fonctions intégrées sont positives), on trouve après quelques manipulations que $P(T \le t) = \int_0^t (\mu - \lambda) e^{-(\mu-\lambda)x} dx$, ce qui prouve le résultat. □

3.2.2. Prise en compte d'un facteur d'impatience

On fait ici l'hypothèse que le client qui arrive dans la file peut décider de la quitter immédiatement si le nombre de clients avant lui l'incite à le faire ; on suppose que lorsque *k* clients sont déjà présents, un client supplémentaire renonce à attendre avec la probabilité $\frac{k}{k+1}$.

On se trouve donc dans le cas d'un processus de naissance et de mort dans lequel la probabilité de décès reste constante égale à μ, la probabilité de naissance étant elle donnée par $\lambda_k = \lambda \left(1 - \frac{k}{k+1}\right) = \frac{\lambda}{k+1}$.

D'après les résultats de la section 2.3 ci-dessus on trouve que $\pi_k = \pi_0 \frac{\varphi^k}{k!}$, et la condition $\sum \pi_k = 1$ conduit à $\pi_0 = e^{-\varphi}$, ce qui prouve que la distribution stationnaire du nombre de clients dans la file est une loi de Poisson de paramètre $\varphi = \frac{\lambda}{\mu}$.

3.3. La file M/G/1

On étudie ici une file d'attente avec un seul serveur et des temps de service indépendants identiquement distribués de loi générale. On notera *F* la fonction de répartition de cette loi et, lorsqu'elle est bien définie *f* sa fonction de densité. On note *S* la variable aléatoire parente. On supposera que la discipline de service est FIFO.

Le processus d'arrivée de clients, noté A_t, est un processus de Poisson d'intensité λ. On se donne comme variable d'état le couple (N_t, D_t) où N_t est le nombre de clients présents dans la file à l'instant t et D_t le temps de service déjà effectué pour le client en service (âge du client).

3.4. La formule de Pollaczek-Khintchine

Afin de se ramener à un problème discret, on considère le processus $N_n = N(t_n)$, où les instants $0 < t_1 < \ldots < t_n$ désignent les instants de départ des clients $1, \ldots, n$. N_n est donc le nombre de clients dans la file juste après le départ du *n*ème client.

Il s'agit d'un processus de Markov discret dont les équations s'écrivent :

$$\begin{cases} N_{n+1} = N_n + \left(A_{t_{n+1}} - A_{t_n}\right) - 1 \; si \; N_n \geq 1 \\ A_{t_{n+1}} - A_{t_n} \; si \; N_n = 0 \end{cases}$$

Les variables aléatoires $\Delta A_n = A_{t_{n+1}} - A_{t_n}$ qui comptabilisent le nombre d'arrivées pendant le service du (n+1)ème client sont :

- indépendantes car les intervalles $[t_n, t_{n+1}[$ sont disjoints et que le processus d'arrivé est un processus de Poisson ;
- identiquement distribuées car les temps de service sont indépendants et de même loi.

(N_n) est donc une chaîne de Markov homogène, et les probabilités de transition sont entièrement définies par $a_i = P(\Delta A_n = i)$; mais conditionnellement au fait que la durée de service est égale à t, le nombre d'arrivée suit une loi de Poisson de paramètre λt, ce qui conduit en intégrant par rapport à la loi de la durée de service à :

$$a_i = \int_0^{+\infty} \frac{(\lambda t)^i}{i!} e^{-\lambda t} f(t) dt$$

Si $G_a(z) = Ez^{\Delta A_n}$ désigne la fonction génératrice associée à la distribution (a_i), on peut noter que $G_a(z) = \phi(\lambda(1-z))$, avec ϕ la transformée de Laplace de la distribution du temps de service ; la chaîne (N_n) peut être décrite par le graphe ci-dessous :

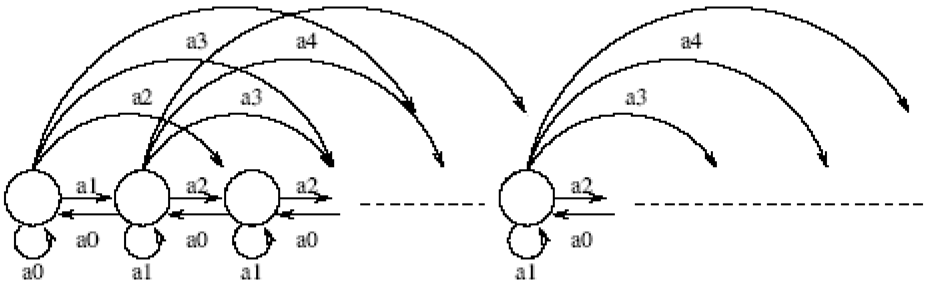


Fig. 6 : *Diagramme d'états*

On en déduit que la matrice de transition est :

$$\begin{pmatrix} a_0 & a_1 & a_2 & \cdots & \cdots & \cdots & \cdots & \cdots \\ a_0 & a_1 & a_2 & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & a_0 & a_1 & a_2 & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & a_0 & a_1 & a_2 & \cdots & \cdots & \cdots \\ \vdots & \ddots & \ddots & \ddots & & & & \\ \vdots & & & & & & & \\ \vdots & & & & & & & \end{pmatrix}$$

La positivité stricte des coefficients a_i assure que la chaîne est apériodique et irréductible, et donc il existe une unique distribution stationnaire. Les équations de Chapman-Kolmogorov (voir 2.1.2 ci-dessus) s'écrivent ici :

$$\pi_k = \pi_0 a_0 + \sum_{j=1}^{k+1} \pi_j a_{k+1-j}$$

On note $G(z) = \sum_{k=0}^{+\infty} \pi_k z^k$ la fonction génératrice de la distribution stationnaire. À l'aide de l'équation ci-dessus on trouve que G satisfait l'équation :

$$G(z) = \pi_0 G_a(z) + \frac{1}{z}\left(G(z) - \pi_0\right) G_a(z)$$

et donc $G(z) = \pi_0 \dfrac{G_a(z)}{1 - \dfrac{G_a(z)-1}{z-1}}$; il reste à déterminer la valeur de π_0 ; pour cela on utilise le fait que $G(1) = 1$ et $G_a'(1) = \lambda E(S) = \rho$ [10].

On en déduit finalement la proposition suivante :

<u>Proposition</u> (formule de Pollaczek-Khintchine) : La série génératrice du nombre de clients à l'état stationnaire, aux instants de départ des clients d'une file d'attente M/G/1 est donnée par la formule :

$$G(z) = (1-\rho) \frac{\varphi(\lambda(1-z))}{1 - \dfrac{\varphi(\lambda(1-z)) - 1}{z-1}}$$

La difficulté est de passer de la distribution stationnaire de $N_n = N(t_n)$ à la distribution stationnaire de $N(t)$. Une propriété importante de la file d'attente M/G/1 est que les clients qui arrivent « voient le système à l'état stationnaire ». Cette propriété, notée

[10] ρ s'appelle la charge du système.

PASTA (*Poisson Arrival see Time Average*), peut être démontrée dans un cadre beaucoup plus général[11].

Comme conséquence de cette propriété, G est la fonction génératrice de $N(t)$ à l'état stationnaire.

3.5. Applications

Cette formulation de la fonction génératrice permet de calculer le nombre moyen de clients dans la file à l'état stationnaire :

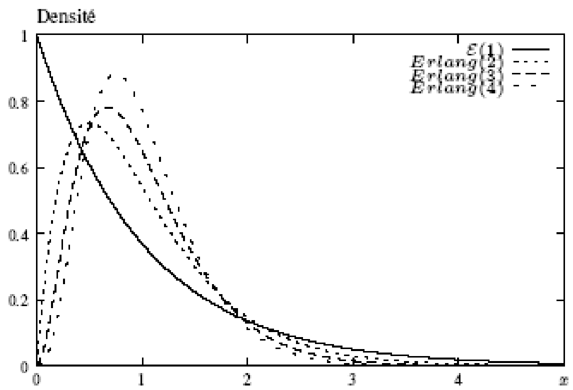
$$E(N) = \rho + \rho^2 \frac{1+c^2}{2(1-\rho)}$$

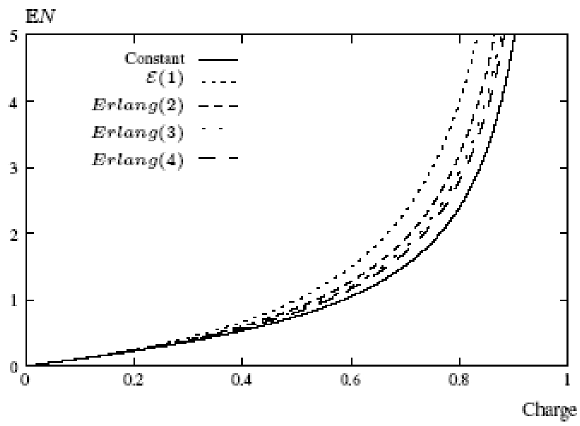
avec $\rho = \lambda E(S)$ et $c = \frac{\sigma(S)}{E(S)}$.

Cas particuliers :

À titre indicatif, on a ainsi dans quelques cas particuliers standards :

- file M/D/1 : $\frac{\rho}{1-\rho}\left(1-\frac{\rho}{2}\right)$ car la variance du temps de service est nulle.
- file M/M/1 : $\frac{\rho}{1-\rho}$ car le coefficient de variation de la loi exponentielle est égal à 1.
- file M/E(k)/1 : $\frac{\rho}{1-\rho}\left(1-\frac{\rho}{2}\left(1-\frac{1}{k}\right)\right)$ car le coefficient de variation de la loi d'Erlang de paramètre *k* est 1/*k*.





[11] Voir WOLFF [1989] p.253.

4. Références

BACCELLI F. and BRÉMAUD P. [1994] « *Elements of Queuing Theory* », New-York : Springer-Verlag

BAYNAT B. [2000] « *Théorie des files d'attente* » Hermes Sciences Publications, Paris.

DACUNHA-CASTELLE D., DUFLO M. [1983] *Probabilités et Statistiques*. Vol. 1 et 2, Paris : Masson.

KARLIN S., TAYLOR H [1975] « *A first course in stochastic processes* », Academic Press.

RIVEST T., LEISERSON C., RIVEST R. [1994] « *Introduction à l'algorithmique* » Paris : Dunod.

SAPORTA G. [1978], « *Théories et méthodes de la statistique* » Institut Français du Pétrole

VILLEMEUR A. « *Sûreté de fonctionnement des systèmes industriels* » volume 67. Paris : Eyrolles.

WOLFF R.W. [1989], « *Stochastic Modelling and the Theory of Queues* » Prentice-Hall, Englewood Cliff