Files d'attente markoviennes

Dans les systèmes d'attente $M|M|1|\infty$, les clients arrivent selon un processus de Poisson d'intensité λ. Les temps de service sont des variables aléatoires indépendantes, de mêmes lois exponentielles de paramètre μ et indépendantes du processus des arrivées. On note $X(t)$ le nombre de clients dans le système à l'instant t.

2.1. File d'attente $M|M|1|\infty$

Dans ce système, les arrivées se font selon un processus de Poisson d'intensité λ, c'est-à dire que les durées inter-arrivées sont indépendantes et suivent toutes la même loi exponentielle de paramètre μ, la capacité du système est illimitée et il ya un seul serveur. La discipline est FIFO. Cette file d'attente correspond à un processus de naissance et de mort de taux de transition constants :

$$\lambda_i = \lambda, \quad i \geq 0 \quad \text{et} \quad \mu_i = \mu, \quad i \geq 1.$$

2.1.1. Distribution stationnaire

Nous allons démontrer que la distribution stationnaire π est un vecteur ayant une infinité de composantes.

Comme le système M|M|1 est un cas particulier du processus de naissance et de mort, on a

$$\forall n \geqslant 1, \quad \pi_n = \left(\frac{\lambda_0 \lambda_1 \cdots \lambda_{n-1}}{\mu_1 \mu_2 \cdots \mu_n}\right) \pi_0 = \left(\frac{\lambda}{\mu}\right)^n \pi_0$$

compte tenu de la relation $\sum_{n=0}^{\infty} \pi_n = 1$, la valeur de π_0 est donnée par la formule

$$\pi_0 = \left(1 + \sum_{n=1}^{\infty} \frac{\lambda_0 \lambda_1 \cdots \lambda_{n-1}}{\mu_1 \mu_2 \cdots \mu_n}\right)^{-1} = \left(1 + \sum_{n=1}^{\infty} \left(\frac{\lambda}{\mu}\right)^n\right)^{-1}$$

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

par suite, $\quad \pi_n = \left(1 - \frac{\lambda}{\mu}\right)\left(\frac{\lambda}{\mu}\right)^n, \quad n \geqslant 0 \; \cdots \; (1)$

Dans le cas où $\frac{\lambda}{\mu} > 1$, il arrive en moyenne λ clients par unité de temps, alors que pendant ce même temps, le serveur à plein régime ne peut traiter que μ clients. Donc le nombre de clients en attente va tendre vers l'infini.

Dans le cas où $\frac{\lambda}{\mu} = 1$, les probabilités π_n sont nulles, ce qui est impossible. La distribution stationnaire n'existe pas.

Lorsque $\frac{\lambda}{\mu} < 1$, la formule (1) définit un vecteur π ayant toutes ses composantes (en nombre infini) strictement positives, tout se passe bien. C'est la condition de stabilité que nous ferons dans la suite.

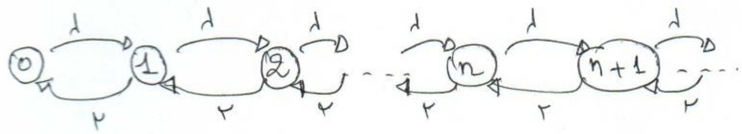
Théorème 1

Soit un système M|M|1 de taux $\lambda > 0$ et $\mu > 0$ tels que $\lambda < \mu$. Alors, la probabilité qu'il y ait n clients dans le système en régime stationnaire est donnée par :

$$\pi_n = (1 - \rho)\,\rho^n \quad , n \geq 0$$

L'expression $\rho = \lambda/\mu$ est souvent appelée coefficient d'utilisation du système ou encore intensité du trafic. En particulier, la probabilité que le serveur soit inoccupé est $\pi_0 = 1 - \frac{\lambda}{\mu}$.

Graphe des transitions.



2-1-2- Paramètres de performance.

A partir de la distribution stationnaire du processus $(X(t))_{t \geq 0}$ On peut calculer les paramètres suivants :

- Le nombre moyen de clients dans le système : $L = \frac{\rho}{1-\rho}$
- Le temps moyen de séjour : $W = \frac{\rho}{\lambda(1-\rho)}$
- Le temps moyen d'attente : $W_q = \frac{\rho^2}{\lambda(1-\rho)}$
- Le nombre moyen de clients dans la file d'attente : $L_q = \frac{\rho^2}{1-\rho}$

Avec $\rho = \lambda/\mu$.

Exemple 1 :

On considère un système M|M|1, un client arrive en moyenne toutes les 12 minutes et la durée moyenne de service est 8 minutes.

1- calculer la probabilité que 3 clients au moins attendent d'être servis.

2- Si le taux d'arrivée λ augmente de 20%, calculer l'augmentation de ρ.

3- Calculer le nombre moyen de clients dans le système et le temps de séjour moyen d'un client dans le système.

Solution

1) prenons l'heure pour unité de temps, on a $\lambda = 5$, $\mu = 7{,}5$ et $\rho = \lambda/\mu < 1$

$\rho = 2/3$. Donc $\pi_n = (1 - 2/3)\left(\frac{2}{3}\right)^n$, $n \geq 0$.

La probabilité p que 3 clients au moins attendent d'être servis est $p = P(X \geq 3) = 1 - P(X < 3) = 1 - \pi_0 - \pi_1 - \pi_2 = (2/3)^3$

2/ λ augmente de 20% donc $\lambda^* = 6/h$ et $\rho^* = \frac{\lambda^*}{\mu} = \frac{4}{5}$, d'où $\rho^* = \left(\frac{4}{5}\right)^3$

ce qui correspond à une augmentation de 73%.

3/ Calculons les paramètres L et W.

- Lorsque $\lambda = 5$ et $\mu = 7{,}5$, on trouve $L = 2$ et $W = 0{,}4h = 24mn$

pour $\lambda^* = 6$ et $\mu = 7{,}5$, on trouve $L^* = 4$ et $W^* = 40mn$.

Une augmentation de 20% du taux d'arrivée provoque donc une augmentation de 100% du nombre moyen de clients dans le système et de 67% du temps moyen de séjour.

2.1.3. Distribution du temps de séjour

Soit un système $M|M|1|\infty$ de taux λ et μ avec $\rho = \lambda/\mu < 1$.
Soit T le temps de séjour d'un client dans le système.
En régime stationnaire, la variable aléatoire T suit une loi exponentielle de paramètre $(\mu - \lambda)$.

$$P(T \leq t) = 1 - e^{-(\mu-\lambda)t} \quad , \ t \geq 0 .$$

2.1.4. Distribution du temps d'attente

Soit T_q le temps d'attente d'un individu, celui-ci est fonction du nombre de clients déjà présents dans la file lorsqu'il arrive. En régime stationnaire, la loi de la variable T_q est donnée par

$$P(T_q \leq t) = 1 - \rho e^{-\mu t(1-\rho)} .$$

2.2. Système d'attente $M|M|1|K$.

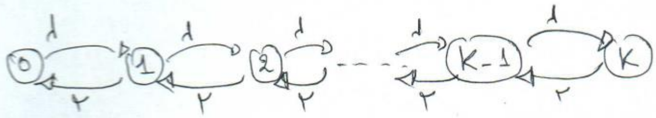
On considère un système à un serveur identique à la file $M|M|1$ excepté que la capacité de la file est finie. On a donc toujours les hypothèses suivantes: Le processus d'arrivée des clients est un processus de Poisson d'intensité λ et le temps de service d'un client est une variable aléatoire exponentielle de taux μ. Soit K la capacité de la file d'attente, quand un client arrive alors qu'il y a déjà K clients présents dans le système, il est perdu. Le système est connu sous le nom $M|M|1|K$. L'espace d'état E est fini, $E=\{0,1,\ldots,K\}$

Le processus de naissance et de mort modélisant ce type de file d'attente est alors défini de la façon suivante.

$$\lambda_n = \begin{cases} \lambda & \text{si } n < K \\ 0 & \text{si } n = K \end{cases} \qquad \mu_n = \begin{cases} \mu & \text{si } n \neq 0,\ n = 1, \dots, K \\ 0 & \text{si } n = 0 \end{cases}$$

La condition de stabilité est : $\rho = \lambda/\mu < 1$.

Graphe :



2.2.1. Distribution stationnaire

L'équation de la distribution π_n est donnée par :

$$\pi_n = \pi_0 \rho^n \qquad \text{si } n \leq K$$

$$\pi_n = 0 \qquad \text{si } n > K$$

$$\pi_0 = \frac{1}{\sum_{n=0}^{K} \rho^n} = \frac{1-\rho}{1-\rho^{K+1}} \qquad \text{si } \lambda \neq \mu \quad \left(\text{et } \frac{1}{K+1} \text{ si } \lambda = \mu\right)$$

Remarque 1 :

La probabilité qu'un client ne soit pas accepté dans le système car la file d'attente est remplie est

$$\pi_K = \rho^K \cdot \frac{1-\rho}{1-\rho^{K+1}}$$

2.2.2. Paramètres de Performance

- $L = \sum_{n=0}^{K} n \pi_n = \frac{\rho}{1-\rho} \left(\frac{1-(K+1)\rho^K + K\rho^{K+1}}{1-\rho^{K+1}} \right)$

Lorsque K tend vers l'infini et $\rho < 1$, on retrouve les résultats du système M|M|1.

A partir des formules de Little, on peut déduire :

- $W = \frac{L}{\lambda_e}$, $\lambda_e = \lambda(1 - \pi_K)$

- $W_q = W - \frac{1}{\mu}$

- $L_q = \lambda_e W_q$.

Exemple 2.

Un cabinet médical ne dispose que d'un seul médecin qui peut traiter 3 patients. Les patients arrivent selon un processus de Poisson de taux $\lambda = 2/h$. Le temps de traitement d'un patient est d'un quart d'heure en moyenne et est distribué selon une loi exponentielle.

1- De quel système d'attente s'agit-il ?

2- Calculer les paramètres de performance.

Solution

1- Il s'agit d'un système $M|M|1|3$.

2- Paramètres de Performance :

- $L = \frac{11}{15} = \frac{\rho}{1-\rho}\left(\frac{1-(K+1)\rho^K + K\rho^{K+1}}{1-\rho^{K+1}}\right)$

- $W = \frac{L}{\lambda_e}$, $\lambda_e = \lambda(1-\pi_3)$ et $\pi_3 = \rho^3 \frac{1-\rho}{1-\rho^4}$

  AN : $\pi_3 = \frac{1}{15}$, $\lambda_e = \frac{28}{15}$ d'où $W = \frac{11}{28}$

- $W_q = W - \frac{1}{\mu} = 1/7$.

- $L_q = \lambda_e W_q = \frac{4}{15}$.

2-3- Système d'attente M/M/S.

Considérons un système identique à la file M/M/1 sauf qu'il comporte S serveurs identiques et indépendants les uns des autres.
On conserve les hypothèses : processus d'arrivée des clients Poissonien de taux λ et temps de service exponentiel de taux μ (pour chacun des serveurs), capacité illimitée. Comme pour la file M/M/1 l'espace des états E est ~~in~~fini : $E = \{0, 1, \ldots\}$. On a un processus de naissance et de mort de taux :

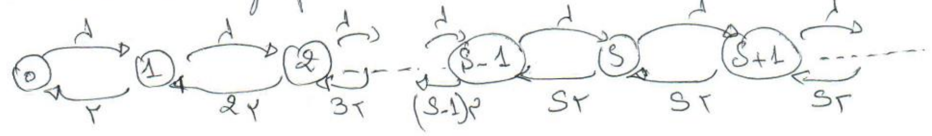
$$\lambda_n = \lambda \qquad n = 0, 1, 2, \ldots$$

$$\mu_n = \begin{cases} 0 & \text{si } n = 0 \\ n\mu & \text{si } 1 \leq n \leq S \\ S\mu & \text{si } n \geq S \end{cases}$$

On appelle $S\mu$ le taux de service global du système, et $\rho = \frac{\lambda}{S\mu}$ l'intensité de trafic globale.

2-3-1- Distribution stationnaire

En utilisant le graphe des transitions suivant :



On obtient la distribution stationnaire du système M/M/S

$$\pi_n = \begin{cases} \dfrac{(\lambda/\mu)^n}{n!}\pi_0 & \text{si } n \leq S \\[2ex] \dfrac{(\lambda/\mu)^n}{S!\, S^{n-S}}\pi_0 = \rho^{n-S}\pi_0 & \text{si } n \geq S. \end{cases}$$

avec $\pi_0 = \left(\sum_{n=0}^{S} \dfrac{(\lambda/\mu)^n}{n!} + \dfrac{(\lambda/\mu)^{S+1}}{S!\,(S-\lambda/\mu)} \right)^{-1}$

Ces relations n'ont un sens que lorsque la série $\pi_0 + \pi_1 + \dots$ converge, c'est-à dire dès que $\rho = \frac{\lambda}{S\mu} < 1$.

Si $S = 1$, on retrouve le résultat correspondant au système M/M/1.

La probabilité qu'un client qui entre dans le système doive attendre est alors:

$$P(X \geq S) = \sum_{n=S}^{\infty} \pi_n = \frac{\pi_S}{1-\rho}$$

2.3.2. Paramètres de Performance.

La distribution stationnaire nous permet de calculer les caractéristiques du système.

- $L_q = \sum_{n=S+1}^{\infty} (n-S)\,\pi_n = \frac{\rho \pi_S}{(1-\rho)^2}$
- $L = L_q + \frac{\lambda}{\mu} = S\rho + \frac{\rho \pi_S}{(1-\rho)^2}$

Les expressions de W_q et W peuvent se déduire à l'aide des formules de Little.

Exemple 3

L'arrivée des clients dans une banque suit un processus de Poisson dont le taux moyen est de 9 clients par heure. La durée de service par client suit une loi exponentielle de moyenne 10 min.

1- Calculer le nombre minimal a de guichets necessaires pour assurer un régime stationnaire de ce processus.

2. Calculer le temps d'attente moyen $S = a$.

Solution

1) Nous avons $\lambda = 9$ clients/h et $\mu = 6$ clients/h et il faut que $\rho = \frac{\lambda}{S\mu} < 1$. D'où $\frac{9}{6S} < 1 \Rightarrow S = 1,5$ donc $a = 2$.

2) Si $S = 2$, On a $\rho = 3/4$ et $\pi_0 = \frac{1}{7}$ d'où $\pi_2 = 9/56$

et $W_q = \frac{\pi_2}{2\mu(1-\rho)^2} = 0,21h = 12,8 \text{ min}$.

Proposition 1.

En régime stationnaire, la distribution du temps d'attente T_q^f d'un nouveau client arrivant dans le système est:

$$P(T_q^f \leq t) = 1 - \frac{\pi_S}{1-\rho} e^{-S\mu t(1-\rho)}$$

2.4. Système M/M/S/S.

Considérons maintenant un système excluant toute possibilité d'attente. Cela signifie qu'un client qui arrive ne peut entrer dans le système que si les S serveurs ne sont pas tous occupés. Il existe donc, en plus du flux de sortie comprenant les clients servis, un flux de demandes refusées ou perdues. On a donc.

$$\lambda_n = \begin{cases} \lambda & \text{si } 0 \leq n \leq S \\ 0 & \text{si } n \geq S \end{cases} \qquad \mu_n = \begin{cases} n\mu & \text{si } 1 \leq n \leq S \\ 0 & \text{si } n = 0 \text{ ou } n > S \end{cases}$$

D'où $\pi_n = \frac{(\lambda/\mu)^n}{n!} \pi_0 \qquad n \leq S$

avec $\pi_0 = \left(\sum_{n=0}^{S} \frac{(\lambda/\mu)^n}{n!} \right)^{-1}$

La probabilité de pertes du système est la probabilité qu'un client qui arrive ne puisse entrer est égale à la probabilité de se trouver dans l'état S.

$$\pi_S = (\lambda/\mu)^S \frac{1}{S!} \pi_0$$

On trouve les paramètres de performance suivants:

- $W = \frac{1}{\mu}$
- $L = \lambda_e W = \lambda(1-\pi_S) W = \frac{\lambda}{\mu}(1-\pi_S)$
- $W_q = L_q = 0$.

Exemple:

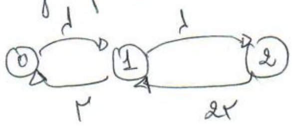
Un central téléphonique d'une petite entreprise comprend deux lignes d'entrée. Les appels arrivent selon un processus de Poisson de taux λ et le temps de service suit une loi exponentielle de paramètre μ. Il n'y a pas de files d'attente.

1- Donner le graphe de transition
2- Donner les équations de balance et déterminer les distributions stationnaires.
3- Calculer les paramètres de Performance.
4- Si $\lambda = \mu$, quel est le pourcentage des appels perdus?

Solution

On a un système $M|M|2|2$.

1/ Le graphe de transition est :



2/ Equations de balance : $0 : \lambda \pi_0 = \mu \pi_1$.

$1 : \lambda \pi_1 + \mu \pi_1 = \lambda \pi_0 + 2\mu \pi_2$.

$2 : 2\mu \pi_2 = \lambda \pi_1$.

$\pi_0 + \pi_1 + \pi_2 = 1$.

$$\Rightarrow \pi_1 = \frac{\lambda}{\mu} \pi_0$$

$$\pi_2 = \frac{\lambda}{2\mu} \pi_1 = \frac{\lambda^2}{2\mu^2} \pi_0 .$$

avec $\pi_0 + \frac{\lambda}{\mu}\pi_0 + \frac{(\lambda/\mu)^2}{2}\pi_0 = \pi_0 \left(1 + \lambda/\mu + \frac{(\lambda/\mu)^2}{2}\right) = 1.$

D'où $\pi_0 = \left(1 + \lambda/\mu + \frac{(\lambda/\mu)^2}{2}\right)^{-1}.$

4/ si $\lambda = \mu$, On trouve

$\pi_0 = 2/5$, $\pi_1 = 2/5$, $\pi_2 = 1/5$ ce qui signifie que 20% des appels sont perdus.

2.5. Système d'attente $M/M/\infty/\infty$

Considérons un système comprenant une infinité de serveurs identiques. Dans ce cas, chaque client est servi dès son entré. Le processus de naissance et de mort associé a pour taux.

$\lambda_n = \lambda \quad n \geq 0$

$\mu_n = n\mu \quad n \geq 1$.

La distribution stationnaire est donnée par:

$$\Pi_n = \left(\frac{\lambda}{\mu}\right)^n \frac{e^{-\lambda/\mu}}{n!} \quad n = 0, 1, \ldots$$

En régime stationnaire le nombre de clients X dans le système $M/M/\infty$ suit une distribution de Poisson de paramètre λ/μ.

Les paramètres de performance sont:

$L = E(X) = \lambda/\mu$, $W = 1/\mu$ et $L_q = W_q = 0$.