Files d'attente

Recherche opérationnelle
Dimitri Watel - ENSIIE

2024

On va s'intéresser dans ce cours à un modèle simple de files d'attente. Ces modèles permettent d'estimer des informations sur la file telle que sa taille ou le temps d'attente.

Concrètement une file d'attente est constituée d'une ou plusieurs listes de personnes et d'un ou plusieurs guichets pour traiter les personnes à l'avant de la file. Quand une personne est traitée, elle quitte la file. On suppose les personnes bien éduquées: les personnes qui arrivent se mettent à l'arrière et personne ne double personne.

Ce type de file peut, bien entendu, modéliser une file d'attente à un guichet mais également des évènements différents tels que des processus de traitement de paquets dans un routeur ou des temps d'attente de patients pour un don d'organe, des stocks de fruits et légumes au supermarché, ...

Une première difficulté dans une file d'attente est l'incertitude. On ne sait pas exactement quand les personnes vont arriver dans la file et quand elles vont repartir. Mais à l'aide de modèles statistiques, on peut estimer une loi de distribution associée à l'arrivée et au départ des personnes. On verra qu'on peut modéliser ces files avec une chaîne de Markov de taille infinie. Une autre difficulté est l'aspect continu de la file, les personnes pouvant arriver et partir à tout moment.

1 Le modèle M/M/k

Le modèle M/M/k est un modèle simplifié de file qui suppose que les personnes arrivent et repartent de la file en suivant un processus stochastique de poisson. Le mélange de ces deux processus définit une file d'attente. On définit une variable aléatoire $X(t)$ qui correspond au nombre de personnes dans la file à l'instant t. Le premier processus de poisson, appelé *processus de naissance* fait apparaître une personne en fin de file aléatoirement et augmente X. Le second processus de poisson, appelé *processus de mort* fait disparaître une personne en début de file aléatoirement et diminue X.

1.1 Equations du modèle

Puisque le temps est continu, on ne peut définir la probabilité que $X(t) = n$ facilement. A la place, on se fixe un pas de temps dt, les équations qui suivent indiquent la probabilité d'apparition ou de disparition d'une personne entre les instances t et $t + dt$, donc $X(t + dt) - X(t)$. Ces probabilités dépendent du nombre de personnes dans la file à l'instant t, on laisse au modèle la liberté de faire varier les probabilités en fonction de la taille de la file. C'est utile par exemple si on veut modéliser un comportement d'abandon, quand une personne se dit qu'elle reviendra plus tard, ou d'attraction, quand une foule rend les gens curieux et qu'ils s'approchent de la foule pour regarder.

On obtient les équations qui régissent X suivantes:

$$\begin{aligned}
Pr(X(t+dt) - X(t) = 1 | X(t) = n) &= \lambda_n \cdot dt + o(dt) \\
Pr(X(t+dt) - X(t) = -1 | X(t) = n > 0) &= \mu_n \cdot dt + o(dt) \\
Pr(X(t+dt) - X(t) = 0 | X(t) = n > 0) &= \\
& 1 - (\lambda_n + \mu_n) \cdot dt + o(dt) \\
Pr(X(t+dt) - X(t) = 0 | X(t) = 0) &= 1 - \lambda_0 \cdot dt + o(dt) \\
(|Pr(X(t+dt) - X(t)| > 1 | X(t) = n) &= o(dt))
\end{aligned}$$

où $(\lambda_n)_{n\in\mathbb{N}}$ et $(\mu_n)_{n\in\mathbb{N}^*}$ sont des valeurs strictement positives.

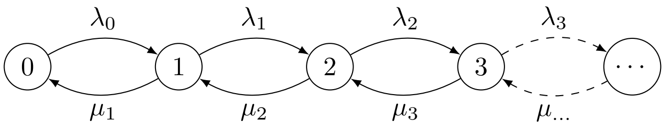
La première équation correspond au premier processus, quand une personne apparait. La seconde correspond au second processus, quand une personne disparait. La troisième et la quatrième indiquent la probabilité que personne n'apparaisse ou ne disparaisse. On peut voir qu'il y a un cas particulier quand $n = 0$ car, dans ce cas, les personnes ne peuvent que apparaître. La dernière est la probabilité que plus d'une personne apparaisse ou disparaisse. Le $o(dt)$ peut être compris comme *une toute petite valeur, qui diminue très vite et tend vers 0 quand dt diminue et tend vers 0.* On peut voir que si dt tend vers 0, la probabilité d'apparition ou de disparition des personnes est très faible, il est plus probable que X ne change pas entre t et $t + dt$. On peut voir aussi que la probabilité d'avoir plus d'un évènement est très faible avec dt. En effet, plus on réduit le pas de temps et moins il y a de chances que deux personnes arrivent ou quittent la file entre t et $t + dt$. En supposant dt assez petit, on obtient donc un modèle assez simple de file où, parfois, une seule personne arrive et parfois une seule personne repart.

Concrètement, à quoi correspondent λ_n et μ_n. Il s'agit de *taux d'arrivée* et *taux de départ*. Ces taux indiquent, en moyenne, le nombre de personnes par seconde qui arrivent et qui repartent de la file s'il y a n personne dans la file. On peut noter que μ_0 n'a pas de sens puisque personne ne peut quitter la file si $n = 0$. C'est pourquoi il y a un cas particulier quand $n = 0$.

Remarque 1. Ces taux peuvent être exprimés dans une autre unité, il faut juste s'assurer qu'il s'agit de la même unité pour les deux.

1.2 Représentation graphique d'une file

On peut représenter graphiquement une file d'attente avec le dessin suivant:



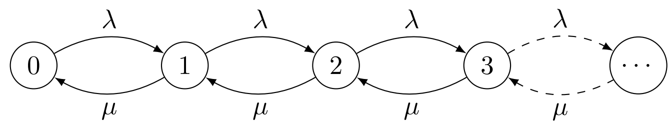
Cette représentation fait écho à celles des chaînes de Markov. **Chaque état indique le nombre de personnes dans la file.** Il ne s'agit pas de moments ou de durées. Chaque arc indique le taux d'arrive et de départ d'une seule personne (donc le passage d'un état n à un état $n+1$ ou $n-1$). Il y a ici plusieurs différences avec les représentations faites dans le modèle des chaînes de Markov. Premièrement la chaîne est infinie, elle ne peut donc être représentée en entier et la plupart des résultats sur les chaînes varient quand on passe d'une chaîne finie à une chaîne infinie. On va donc refaire dans ce cours une étude spécifiques à ce type de chaînes que sont les files d'attente. Deuxièmement, les probabilités ne sont pas indiquées sur les arcs. Seuls les taux sont indiqués. Enfin, pour avoir une chaîne complète, il faudrait les arcs indiquant que le nombre de personnes dans la file est constant (donc de l'état n vers lui-même) et ceux indiquant que le nombre de personnes varie de plus que 1.

Détail important: notez bien l'emplacement des taux en fonction de n. Chaque arc indique le passage d'un état n vers un état m, valant $n+1$ ou $n-1$. Sur cet arc est indiqué le taux λ_n si $m = n+1$ et μ_n si $m = n-1$. Ainsi les flèches de taux λ_n et μ_n ne sont pas alignées. Aussi, on revoit bien sur le dessin que μ_0 n'a pas de sens.

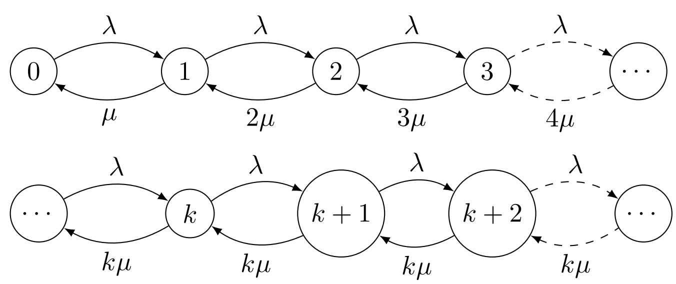
Enfin, attention à un autre détail. Il faut bien prendre du recul sur ces notations. λ_n et μ_n sont des taux d'arrivées, en nombre de personne par seconde. Mais ce n'est pas parce qu'il est écrit à l'état n qu'il y a 30 personnes qui arrivent chaque seconde qu'on aura $n+30$ personnes dans la file à la seconde suivante. Ce taux indique une moyenne et non un départ déterministe. Il indique donc une probabilité de passage de n à $n+1$ entre des instances t et $t+dt$. Sans tenir compte de cet aspect aléatoire, on peut se faire une fausse idée de l'évolution de la file. Par exemple, ce n'est pas parce qu'il y a des taux μ_n très élevés et des taux λ_n faibles que la file ne va pas grossir. C'est improbable mais il est possible que personne ne quitte la file pendant un certain temps.

1.3 Files particulières

On peut mentionner deux files particulières. La première est la file où les taux ne dépendent pas de n. On a donc $\lambda_n = \lambda$ et $\mu_n = \mu$ pour tout $n \in \mathbb{N}$ et où $\lambda, \mu > 0$.



Un second cas reconnaissable est celui des files avec plusieurs guichets. Chaque guichet fait que le nombre de personne qui quittent la file par seconde est multiplié par le nombre de guichets ouverts. Cependant, un guichet ne peut traiter qu'une personne dans la file, un guichet n'ouvre que s'il peut traiter quelqu'un. Donc au départ, quand il n'y a personne dans la file, aucun guichet n'est ouvert. Puis une personne arrive, on ouvre un guichet, puis une autre personne arrive, on ouvre un autre guichet, et ainsi de suite, jusqu'à ce qu'on ait ouvert tous les guichets. Le taux augmente au fur et à mesure des ouvertures jusqu'à ce moment où le taux devient constant. S'il y a k guichets on a la représentation suivante:



1.4 Nom du modèle

Dans le modèle, le nom M/M/k fait référence respectivement:

- au processus d'arrivée, M signifie Poisson et vient du fait qu'on modélise ainsi un processus Markovien.
- au processus de départ, M signifie ecore Poisson
- au nombre de guichets.

D'autres lettres existent comme le D pour des arrivées et départs déterministes, prévus à l'avance, G pour général indiquant qu'on ne connait pas la distribution (plus exactement elle n'est pas précisée par le modèle). Ces notations sont les notations dites de Kendall.

2 Régime permanent

On souhaite pouvoir obtenir des informations sur la file. On va s'intéresser dans la fin de ce cours à la loi de distribution de X en fonction de t: on note $P_n(t)$ la probabilité qu'il y ait n personne dans la file à l'instant t. Autrement dit $P_n(t) = Pr(X(t) = n)$.

On fait deux hypothèses dans cette partie.

- On suppose la file vide à l'instant 0.
- On suppose, que, dans toutes les équations régissant $X(t)$, les $o(dt)$ sont la même fonction qu'on notera $g(dt)$ dans la suite. On vérifie donc $\frac{g(dt)}{dt} \to 0$ si $dt \to 0$.

Théorème 2.1. *P_n est solution du système d'équations différentielles suivant:*

$$\begin{aligned} P_n'(t) &= \lambda_{n-1} \cdot P_{n-1}(t) + \mu_{n+1} \cdot P_{n+1}(t) \\ &\quad - (\mu_n + \lambda_n) \cdot P_n(t) \text{ si } n > 0 \qquad (1) \\ P_0'(t) &= \mu_1 \cdot P_1(t) - \lambda_0 \cdot P_0(t) \qquad (2) \\ P_0(0) &= 1 \qquad (3) \\ P_n(0) &= 0, n > 0 \qquad (4) \end{aligned}$$

Proof. Les deux cas initiaux correspondent au fait qu'on fait l'hypothèse que la file est vide au début du processus. Pour démontrer la première équation, on va utiliser la définition de la dérivée comme taux d'accroissement.

Attention, la preuve ici est faite pour $n > 0$. Elle doit être adaptée pour le cas $n = 0$.

$$P_n'(t) = \lim_{dt \to 0} \frac{P_n(t+dt) - P_n(t)}{dt}$$

Il faut décrire $P_n(t+dt)$. La probabilité qu'il y ait n personnes à l'instance $t+dt$ dépend du nombre de personnes à l'instant t. Par la formule des probabilités totales:

$$P_n(t+dt) = \sum_{i=0}^{+\infty} Pr(X(t+dt) = n | X(t) = i) P_i(t)$$

En utilisant les équations qui régissent X

$$\begin{aligned} P_n(t+dt) &= (\lambda_{n-1} \cdot dt + g(dt)) \cdot P_{n-1}(t) \\ &\quad + (\mu_{n+1} \cdot dt + g(dt)) \cdot P_{n+1}(t) \\ &\quad + (1 - (\lambda_n - \mu_n) \cdot dt + g(dt)) \cdot P_n(t) \\ &\quad + \sum_{\substack{i \neq n-1 \\ i \neq n \\ i \neq n+1}} g(dt) P_i(t) \\ P_n(t+dt) &= \lambda_{n-1} \cdot dt \cdot P_{n-1}(t) \\ &\quad + \mu_{n+1} \cdot dt \cdot P_{n+1}(t) \\ &\quad + (1 - (\lambda_n - \mu_n) \cdot dt) \cdot P_n(t) \\ &\quad + g(dt) \sum_{i=0}^{+\infty} P_i(t) \end{aligned}$$

On rappelle que $\sum_{i=0}^{+\infty} P_i(t) = 1$.

$$\begin{aligned} \frac{P_n(t+dt) - P_n(t)}{dt} &= \lambda_{n-1} \cdot P_{n-1}(t) + \mu_{n+1} \cdot P_{n+1}(t) \\ &\quad + (-(\lambda_n - \mu_n)) \cdot P_n(t) \\ &\quad + \frac{g(dt)}{dt} \\ P_n'(t) &= \lambda_{n-1} \cdot P_{n-1}(t) + \mu_{n+1} \cdot P_{n+1}(t) \\ &\quad + (-(\lambda_n - \mu_n)) \cdot P_n(t) \end{aligned}$$

□

On va résoudre ce système dans un cas simple, nommé régime permanent. Le régime permanent est un état de la file où les probabilités $P_n(t)$ sont constantes. Il peut exister ou non et être atteint plus ou moins rapidement. On se pose ici la question de l'existence de ce régime.

Définition 1. Il existe un régime permanent si et seulement s'il existe un ensemble de fonctions de probabilité $P_n(t)$ solutions des equations 1 et 2 du système du théorème 2.1 telles que $P_n'(t) = 0$. On note alors $P_n(t) = P_n$.

Remarque 2. Ce n'est pas parce que $P_n(t)$ devient constante après un certain rang t que le nombre de personnes dans la file devient constant. Seules les probabilités d'avoir n personnes dans la file deviennent constantes. Observer la file montrera que sa taille varie aléatoirement avec le temps, par contre la distribution gérant le nombre de personnes dans la file ne dépendra pas du temps.

Théorème 2.2. *La série $\sum_{n=1}^{+\infty} \frac{\lambda_0 \cdot \lambda_1 \cdots \lambda_{n-1}}{\mu_1 \cdot \mu_2 \cdots \mu_n}$ converge si et seulement s'il existe un régime permanent. Dans ce cas, on a*

$$\begin{aligned} P_n &= \frac{\lambda_0 \cdot \lambda_1 \cdot \lambda_2 \cdots \lambda_{n-1}}{\mu_1 \cdot \mu_2 \cdot \mu_3 \cdots \mu_n} \cdot P_0 \\ P_0 &= \frac{1}{1 + \sum_{n=1}^{+\infty} \frac{\lambda_0 \cdot \lambda_1 \cdots \lambda_{n-1}}{\mu_1 \cdot \mu_2 \cdots \mu_n}} \end{aligned}$$

Proof. $(P_n)_{n \in \mathbb{N}}$ est régime permanent si et seulement si

$$\begin{aligned} 0 &= \lambda_{n-1} \cdot P_{n-1} + \mu_{n+1} \cdot P_{n+1} \\ &\quad - (\mu_n + \lambda_n) \cdot P_n \text{ si } n > 0 \\ 0 &= \mu_1 \cdot P_1 - \lambda_0 \cdot P_0 \\ \sum_{i=1}^{+\infty} P_i &= 1 \end{aligned}$$

On peut alors montrer par récurrence que $(P_n)_{n \in \mathbb{N}}$ est régime permanent si et seulement si

$$\begin{aligned} P_n &= \frac{\lambda_0 \cdot \lambda_1 \cdot \lambda_2 \cdots \lambda_{n-1}}{\mu_1 \cdot \mu_2 \cdot \mu_3 \cdots \mu_n} \cdot P_0 \\ \sum_{i=1}^{+\infty} P_i &= 1 \end{aligned}$$

Puisque $\sum_{i=1}^{+\infty} P_i = 1$ on peut déduire P_0 à partir de P_n pour tout $n > 1$.

$$P_n = \frac{\lambda_0 \cdot \lambda_1 \cdot \lambda_2 \cdots \lambda_{n-1}}{\mu_1 \cdot \mu_2 \cdot \mu_3 \cdots \mu_n} \cdot P_0$$
$$P_0 = \frac{1}{1 + \sum_{n=1}^{+\infty} \frac{\lambda_0 \cdot \lambda_1 \cdots \lambda_{n-1}}{\mu_1 \cdot \mu_2 \cdots \mu_n}}$$

La série $\sum_{n=1}^{+\infty} \frac{\lambda_0 \cdot \lambda_1 \cdots \lambda_{n-1}}{\mu_1 \cdot \mu_2 \cdots \mu_n}$ converge si et seulement si $P_0 \neq 0$. Si la série diverge alors $P_0 = 0$. On en déduit alors que $P_n = 0$ pour tout n, on a donc une contradiction et il n'y a pas de régime permanent. Si la série converge, $P_0 \neq 0$ et les valeurs sont bien définies et le régime permanent existe. □

Maintenant qu'on dispose d'une caractérisation de l'existence du régime permanent dans le cas général, on peut l'appliquer sur des cas plus simple. Si par exemple les taux d'arrivée et de départ sont constants égaux respectivement à λ et μ alors il existe un régime permanent si et seulement si la série géométrique de coefficient λ/μ converge, donc si et seulement si $\lambda < \mu$. Ce constat est assez logique puisqu'une file avec $\lambda > \mu$ signifie qu'il y a plus de chances qu'une personne arrive que de chances qu'une personne reparte. Et donc, pour tout i, la probabilité $P_i(t)$ augmente petit à petit, jusqu'à ce qu'on ait environ i personnes dans la file, puis diminue et converge vers 0 au fur et à mesure que d'autres personnes arrivent dans la file au cours du temps. Ainsi en cas de régime permanent, la seule solution possible serait $P_i = 0$ pour tout i, ce qui est exclus. A l'inverse, si $\lambda < \mu$, alors dès qu'une personne arrive dans la file, il y a des chances qu'elle reparte avant qu'une autre personne arrive. Plus de gens entrent dans la file, plus on a de chances de ramener le nombre de personnes à 0. On fini par atteindre une sorte d'équilibre et les probabilités convergent.

3 Espérance de la taille de la file

Définition 2. L'espérance du nombre de personnes dans la file est définie par

$$L(t) = E(X(t)) = \sum_{n=0}^{+\infty} nP_n(t)$$

Dans une file avec k guichets, l'espérance du nombre de personnes en attente est définie par

$$L'(t) = E(\max(0, X(t) - k)) = \sum_{n=k}^{+\infty} (n-k)P_n(t)$$

On donne ci-après un exemple de calcul de L et L' dans un cas particulier.

Théorème 3.1. *Dans une file $M/M/1$ où les taux d'arrivée et de départ sont constants égaux respectivement à λ et μ avec $\lambda < \mu$, alors en régime permanent, $L = \frac{\lambda}{\mu - \lambda}$ et $L' = \frac{\lambda^2}{\mu(\mu-\lambda)}$.*

Proof. Puisque $\lambda < \mu$ alors

$$P_0 = \frac{1}{1 + \sum_{n=1}^{+\infty} \frac{\lambda^n}{\mu^n}}$$

Or

$$\sum_{n=0}^{+\infty} \frac{\lambda^n}{\mu^n} = \frac{1}{1 - \frac{\lambda}{\mu}}$$
$$P_0 = 1 - \frac{\lambda}{\mu}$$
$$P_n = \frac{\lambda^n}{\mu^n}\left(1 - \frac{\lambda}{\mu}\right)$$

$$L = \sum_{n=0}^{+\infty} nP_n = \sum_{n=1}^{+\infty} nP_n = \sum_{n=1}^{+\infty} n\frac{\lambda}{\mu}^n \left(1 - \frac{\lambda}{\mu}\right)$$

On effectue un changement de variable, on reconnait alors une série usuelle.

$$L = \left(\sum_{n=0}^{+\infty}(n+1)\frac{\lambda}{\mu}^n\right)\frac{\lambda}{\mu}\left(1 - \frac{\lambda}{\mu}\right)$$
$$L = \frac{1}{\left(1 - \frac{\lambda}{\mu}\right)^2}\frac{\lambda}{\mu}\left(1 - \frac{\lambda}{\mu}\right)$$
$$L = \frac{\lambda}{\mu - \lambda}$$

Pour L', l'idée est similaire.

$$L' = \sum_{n=1}^{+\infty}(n-1)P_n = \sum_{n=2}^{+\infty}(n-1)P_n$$
$$L' = \sum_{n=2}^{+\infty}(n-1)\frac{\lambda}{\mu}^n\left(1 - \frac{\lambda}{\mu}\right)$$
$$L' = \sum_{n=0}^{+\infty}(n+1)\frac{\lambda}{\mu}^n\frac{\lambda}{\mu}^2\left(1 - \frac{\lambda}{\mu}\right)$$
$$L' = \frac{\lambda}{\mu}L$$
$$L' = \frac{\lambda^2}{\mu(\mu - \lambda)}$$

□