MA 411 : Modélisation et analyse des processus stochastiques

Chaînes de Markov à temps continu (CMTC)
Séance de TD du 13 mai 2020

Vous trouverez ci-après l'énoncé et le corrigé de l'exercice 23 de la séance de TD consacrée aux chaînes de Markov à temps contnu. La correction a été rédigée dans le but de vous aider si vous êtes bloqué ou pour vérifier votre propre travail. Il se peut qu'elle contienne elle-même des erreurs. Si tel est le cas, elles seront corrigées au fur et à mesure qu'elles sont détectées. La version en ligne sur *https://chamilo.grenoble-inp.fr/courses/MA332* sera mise à jour de manière à intégrer ces corrections. Dans de nombreux exercices, il existe plusieurs méthodes pour aboutir au résultat. Si vous avez des doutes sur la méthode que vous avez vous-même employée, n'hésitez pas à m'en faire part (*laurent.lefevre@lcis.grenoble-inp.fr*).

Exercice 23 (la file simple)
On considère un processeur auquel arrivent des requêtes, ces arrivées étant modélisées par un processus de Poisson de paramètre λ. Le temps que met le processeur à traiter chaque requête est modélisé par une variable aléatoire de distribution exponentielle $Exp(\mu)$, de paramètre μ. Le processeur traite la première requête qui lui arrive. Pendant ce temps, les autres requêtes qui arrivent sont rangées dans *une file d'attente*. Lorsque le processeur a terminé la traitement de la requête en cours, il commence le traitement de la requête la plus ancienne dans la file d'attente (discipline de file "*first in, first out*" (FIFO), ou *premier arrivé, premier servi*).

1. Modéliser ce problème par une chaîne de Markov à temps continu (CMTC) : donner le graphe associé avec les taux de transitions, écrire le générateur infinitésimal **A** et donner la caractérisation terme de temps de séjour et de probabiltés de transition.

2. Determiner le régime stationnaire.

3. Quel sont - en régime stationnaire - le nombre moyen de requêtes dans le système (buffer et processeur ensemble), la taille moyenne de la

file (i.e. le nombre de requêtes en attente dans le buffer) et le taux d'utilisation du processeur (en pourcentage du temps total)?

Correction de l'Exercice 23

1. On pose $X(t) \in E := \mathbb{N}$, le nombre de requêtes dans le système (i.e. traitées par le processeur ou en attente dans la file) au temps $t > 0$. Le processus stochastique considéré est appelé *file simple* ou encore *file M/M/1*. Il est caractérisé par des arrivées poissoniennes, des temps de service exponentiels, la présence d'un seul serveur et une gestion FIFO de la file d'attente. Il est représenté schématiquement à la figure 1. Nous allons montrer qu'il s'agit d'un cas particulier de processus de naissances et de morts (avec des taux de naissance et de mort constants), c'est-à-dire que les taux de transition à partir d'une état n ne sont non nuls que pour les transitions vers les états $n-1$ (mort) ou $n+1$ (naissance).
Montrons dans un premier temps que les taux de transition d'un état n vers un état p sont nuls pour $p \geq n+2$.
Une telle transition sur un intervalle de temps $[t, t+dt[$ correspond à deux arrivées au moins pendant l'intervalle de temps dt[1]. La probabilité qu'il y ait $k \geq 2$ arrivées pendant un intervalle de temps de longueur dt, dans un processus de Poisson de paramètre λ, vaut :

$$P\left(N(dt) - N(0)\right) = k) = \frac{(\lambda\, dt)^k}{k!} e^{-\lambda\, dt}$$

Ainsi, la probabilité de transition, pendant un intervalle de temps dt, entre l'état n et l'état p (avec $p \geq n+2$) s'écrit :

$$p_{n,p}\left(dt\right) = \frac{(\lambda\, dt)^2}{2!} e^{-\lambda\, dt} \tilde{P}$$

[1]Une transition de n à $n+2$, par exemple, pourrait également correspondre à 3 arrivées et un départ, quatre arrivées et deux départs, etc.

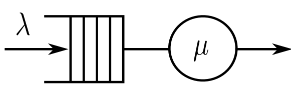

Figure 1: Représentation schématique d'une file simple (M/M/1) : les arrivées se font selon un processus de Poisson de paramètre λ, le temps de service d'un client est distribué selon une exponentielle de paramètre μ, la file est sans limite et la discipline de service est *premier arrivé, premier servi.*

où $\tilde{P}$ représente la probabilité des autres évènements (en plus des deux arrivées nécessaires) pour une transition de n à p pendant l'intervalle de temps dt. On a donc pour le taux de transition correspondant :

$$\lambda_{n,p} := \lim_{dt \to 0} \frac{p_{n,p}(dt)}{dt} = 0$$

De la même manière, une transition pendant un intervalle de temps de longueur dt d'un état n ($n \geq 2$) à un autre état p avec $p \leq n-2$ correspond à deux départs au moins pendant cet intervalle de temps. Comme les temps de service sont des exponentielles indépendantes, la probabilité d'avoir deux départs pendant l'intervalle de temps dt vaut[2] :

$$\begin{aligned}
P\left(T_1^\mu + T_2^\mu \leq dt\right) &= \int_0^{dt} \mu^2 \xi e^{-\mu\xi}\, d\xi \\
&= -\mu dt\, e^{-\mu dt} + 1 - e^{-\mu dt} \\
&= -\mu dt \left(1 - (\mu dt) + \frac{(\mu dt)^2}{2!} - \ldots\right) \\
&\quad +1 - \left(1 - (\mu dt) + \frac{(\mu dt)^2}{2!} - \ldots\right) \\
&= \frac{(\mu dt)^2}{2} + o(dt^2), \text{ pour } dt \to 0
\end{aligned}$$

Ainsi, la probabilité de transition, pendant un intervalle de temps dt, entre l'état n et l'état p (avec $p \leq n-2$) s'écrit :

$$p_{n,p}(dt) = \left(\frac{(\mu dt)^2}{2} + o(dt^2)\right) \cdot \hat{P}$$

où $\hat{P}$ représente la probabilité des autres évènements (en plus des deux départs nécessaires) pour une transition de n à p pendant l'intervalle de temps dt. On a donc pour le taux de transition correspondant :

$$\lambda_{n,p} := \lim_{dt \to 0} \frac{p_{n,p}(dt)}{dt} = 0$$

Les taux de transition d'un état n vers un état p sont donc nuls pour $p \leq n-2$. La *file simple* que nous considérons dans cet exercice est donc bien un exemple de *processus de naissances et de morts*. Il nous reste, pour spécifier complètement son modèle par CMTC, à calculer les taux de transition d'un état n ($n \geq 1$) vers ces deux voisins $n-1$ et $n+1$.

[2]La somme de deux exponentielles indépendantes de même paramètre μ est une loi Gamma $\Gamma(\mu, 2)$.

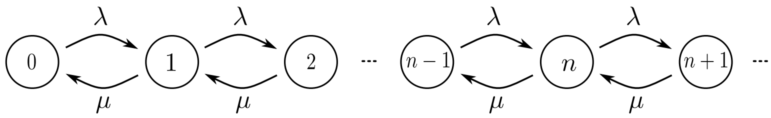

Figure 2: La chaîne de Markov à temps continu assocée à la file simple

Pour une transition de n vers $n+1$, il faut qu'il y ait 1 arrivée de requête et 0 départs, 2 arrivées et 1 départ, 3 arrivées et 2 départ, etc. Nous avons déjà montré que la probabilité associée à l'arrivée de deux requêtes (ou plus) pendant un intervalle de temps dt correspond à un taux de transition nul. Le seul cas à envisager, dans le calcul du taux de transition $\lambda_{n,n+1}$, est donc celui où il y a une arrivée de requête et aucun départ pendant un intervalle de temps dt. La probabilité correspondante est :

$$P\left(T^{\mu} > dt; T^{\lambda} \leq dt\right) = P\left(T^{\mu} > dt\right) \cdot P\left(T^{\lambda} \leq dt\right) = e^{-\mu dt}\left(1 - e^{-\lambda dt}\right)$$

et le taux de transition $\lambda_{n,n+1}$ vaut donc :

$$\lambda_{n,n+1} := \lim_{dt \to 0} \frac{e^{-\mu dt}\left(1 - e^{-\lambda dt}\right)}{dt} = \lambda$$

On trouvera de même

$$P\left(T^{\mu} \leq dt; T^{\lambda} > dt\right) = P\left(T^{\mu} \leq dt\right) \cdot P\left(T^{\lambda} > dt\right) = \left(1 - e^{-\mu dt}\right) e^{-\lambda dt}$$

et

$$\lambda_{n,n-1} := \lim_{dt \to 0} \frac{\left(1 - e^{-\mu dt}\right) e^{-\lambda dt}}{dt} = \mu$$

Nous venons d'établir que la file simple est un processus de naissances et de morts où les taux de transition sont constants (indépendants de l'état). La CMTC correspondante est représentée à la figure 2. Le *générateur infinitésimal* (infini) de la file simple s'écrit donc :

$$\mathbf{A} = \begin{pmatrix} -\lambda & \lambda & 0 & \dots \\ \mu & -(\lambda+\mu) & \lambda & \dots \\ 0 & \mu & -(\lambda+\mu) & \ddots \\ \vdots & \vdots & \ddots & \ddots \end{pmatrix}$$

Le temps de séjour T_n dans un état $n \in \mathbb{N}$ est une variable exponentielle

$$T_n \sim \begin{cases} \mathrm{Exp}\left(\lambda\right) & \text{si } n = 1 \\ \mathrm{Exp}\left(\lambda + \mu\right) & \text{si } n \geq 1 \end{cases}$$

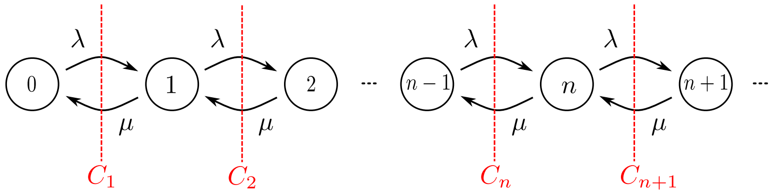

Figure 3: Les coupes qui permettent d'obtenir la distribution stationnaire

Les probabilités de transition s'écrivent

$$p_{n,k} = \begin{cases} \frac{\mu}{\lambda+\mu} & \text{si } k = n-1 \\ \frac{\lambda}{\lambda+\mu} & \text{si } k = n+1 \\ 0 & \text{sinon} \end{cases}$$

pour $n \geq 2$ et

$$p_{1,2} = 1$$

dans le cas où la chaîne est dans l'état $n = 1$.

2. Pour obtenir la distribution stationnaire de probabilité, le plus simple est d'appliquer la méthode des coupes. Les équations qui correspondent aux coupes de la figure 3. sont :

$$\lambda \boldsymbol{\pi}_{n-1} = \mu \boldsymbol{\pi}_n \, , \; \forall n \geq 1$$

On obtient directement :

$$\boldsymbol{\pi}_n = \boldsymbol{\pi}_0 \left(\frac{\lambda}{\mu} \right)^n \, , \; \forall n \geq 1$$

et la condition de normalisation donne :

$$\boldsymbol{\pi}_0 = \frac{1}{\sum_{n=0}^{\infty} \left(\frac{\lambda}{\mu} \right)^n} = 1 - \frac{\lambda}{\mu}$$

La convergence de la série au dénominateur est donc une condition nécessaire et suffisante d'existence de la distribution stationnaire de probabilités (aussi appelée *condition d'ergodicité* de la CMTC). Elle est garantie si et seulement si

$$\lambda < \mu$$

c'est-à-dire si le temps moyen entre deux arrivées est supérieur au temps moyen mis par le processeur pour satisfaire une requête.

3. Le nombre moyen de requête dans le systèmes est

$$\begin{aligned} Q &:= \sum_{n\geq 0} n\boldsymbol{\pi}_n \\ &= \boldsymbol{\pi}_0 \sum_{n\geq 0} n\left(\frac{\lambda}{\mu}\right)^n \\ &= (1-\rho)\sum_{n\geq 0} n\rho^n \end{aligned}$$

avec $\rho := \frac{\lambda}{\mu} < 1$. Or :

$$\begin{aligned} \sum_{n\geq 0} n\rho^n &:= \rho\sum_{n\geq 1} n\rho^{n-1} \\ &= \rho\sum_{n\geq 1} \left.\frac{dx^n}{dx}\right|_{x=\rho} \\ &= \rho\,\frac{d}{dx}\left.\left[\sum_{n\geq 1} x^n\right]\right|_{x=\rho} \\ &= \rho\,\frac{d}{dx}\left.\left[\frac{1}{1-x}\right]\right|_{x=\rho} \\ &= \frac{\rho}{(1-\rho)^2} \end{aligned}$$

Donc

$$Q = \frac{\rho}{(1-\rho)}$$

Ce nombre moyen de requêtes dans le système augmente quand $\rho \to 1$ (avec $\rho < 1$), c'est-à-dire quand le temps moyen entre deux arrivées $\frac{1}{\lambda}$ tend vers la durée moyenne de service $\frac{1}{\mu}$. La taille moyenne du buffer s'écrit

$$\begin{aligned} Q_f &:= \sum_{n\geq 2} (n-1)\boldsymbol{\pi}_n \\ &= \sum_{n\geq 2} n\boldsymbol{\pi}_n - \sum_{n\geq 2}\boldsymbol{\pi}_n \\ &= \sum_{n\geq 0} n\boldsymbol{\pi}_n - \boldsymbol{\pi}_1 - (1-\boldsymbol{\pi}_0-\boldsymbol{\pi}_1) \\ &= Q-(1-\boldsymbol{\pi}_0) \\ &= Q-\rho \\ &= \frac{\rho^2}{1-\rho} \end{aligned}$$

Le taux moyen d'utilisation du processeur est

$$U = 1-\boldsymbol{\pi}_0 = \rho$$