TD 5: Modélisation et chaînes de Markov discrètes

lionel.rieg@ens-lyon.fr

1 Schéma (plus simple) de Matthes et modélisation

Exercice 1

On rappelle qu'un schéma de Matthes est la donnée de :
- un ensemble $\mathcal{E}$ d'*états* ι ;
- un ensemble $\mathcal{S}$ de *sources* α ;
- une fonction d'activation $A : \mathcal{E} \to \mathcal{P}(\mathcal{S})$;
- des vitesses $c(\alpha, \iota)$;
- des probabilités de transition $p(\alpha, \iota, \iota')$;
- des distributions de probabilités données (en général) par leurs fonctions de répartition F_α.

1. Donner la signification des différents éléments d'un schéma de Matthes.

Une file d'attente est composée de deux aspects distincts : *le processus des entrées* et *le mécanisme de service*.
Le processus des entrées est défini par une suite de couples aléatoires (T_n, Z_n) où T_n est le temps d'arrivée du n^e client (on a donc $T_0 = 0 < T_1 < T_2 < \dots$) et Z_n représente les attributs du client n. Les attributs des clients peuvent avoir des formes très variées, parmi lesquelles on trouve la *charge apportée* qui dénote la quantité de service dont à besoin le client et la *charge apportée typée* qui comporte en outre le type de service nécessaire.

Le mécanisme de service est défini par le nombre de serveurs, leurs vitesses, la capacité de la salle d'attente et la *discipline de service* qui explique comment sont traités les clients (notamment la priorité et la préemption). Voici quelques exemples de discipline :

PAPS premier arrivé, premier servi ;

DAPS dernier arrivé, premier servi ;

ALÉA le client est choisi uniformément dans la salle d'attente ;

PP (processeur partagé) tous les clients sont servis en parallèle à la même vitesse.

On se limite au cas où :
- Z_n représente la charge apportée (à valeurs dans $\mathbb{R}^+$) et est i.i.d. ;
- T_n et Z_n sont indépendants ;
- T_n est un *processus de renouvellement*, *i.e.* $s_{n+1} = T_{n+1} - T_n$ est une suite de v.a.i.i.d. ;
- il n'y a qu'un serveur qui travaille à vitesse 1
- la discipline de service est l'une des trois premières présentées ci-dessus.

2. Donner le schéma de Matthes de la file d'attente décrite ci-dessus.
3. Que se passe-t-il si on ajoute des serveurs ?
4. Si on voulait ajouter des types aux clients, comment faudrait-il modifier le schéma ?

Exercice 2

On considère un système composé de n stations qui émettent des messages. Le temps est discret et un entier $W > 0$ est fixé.

Si à un instant une station i à un message à émettre, elle tire indépendamment un entier aléatoire W_i dans $\{0, 1, \dots, W-1\}$. A chaque nouvel instant, elle décrémente de 1 son entier W_i. A l'instant où W_i vaut 0, elle émet son message.

Si une seule station émet à un instant donné, son message est bien transmis. Si elle a un autre message à émettre, elle tire à l'instant suivant un nouvel entier aléatoire et recommence la procédure.

Si deux stations ou plus émettent en même temps, les messages se brouillent, on parle de *collision*. L'émission de tous ces messages échoue. Chaque station concernée conserve son message et tire à l'instant suivant et indépendamment un nouvel entier aléatoire. La procédure continue ainsi de suite.

1. Que se passe-t-il dans le système si on choisit $W = 1$?
2. On suppose que à l'instant $t = 0$, chaque station a un nombre fixé, fini ou infini, de messages qu'elle doit transmettre les uns après les autres. Montrer que l'on peut décrire la dynamique du système sous forme d'une chaîne de Markov $(X_t)_{t\in\mathbb{N}}$ à valeurs dans un ensemble d'états bien choisi. Illustrer en dessinant le graphe de transition dans le cas $n = 2$, $W = 2$ et 1 message à transmettre par station, ainsi que dans le cas $n = 2$, $W = 3$ et une infinité de messages par station.

2 Chaînes de Markov

Exercice 3 (Chapman-Kolmogorov)

On se donne une suite (X_n) de v.a.i.i.d. à valeurs dans $\{-1, 1\}$ telle que $\mathbb{P}(X_0 = -1) = \mathbb{P}(X_0 = 1) = \frac{1}{2}$. On définit la suite (Y_n) par : $Y_{2n} = X_n$ et $Y_{2n+1} = X_n X_{n+1}$. On admet que la suite (Y_n) est i.i.d.

1. Vérifier que la suite (Y_n) satisfait les équations de Chapman-Kolmogorov.
2. Montrer que ce n'est pourtant pas une chaîne de Markov.
3. On considère la suite (Y_n, Y_{n+1}). Montrez qu'il s'agit d'une chaîne de Markov (non homogène).

Exercice 4 (Distribution limite)

Calculer, si elles existent, les distributions limites des chaînes suivantes (étudiées au TD précédents) en utilisant des relations de récurrences :

1. Doudou, le hamster paresseux, ne connaît que trois endroits dans sa cage : les copeaux où il dort, la mangeoire où il mange et la roue où il fait de l'exercice. Ses journées sont assez semblables les unes aux autres, et son activité se représente aisément par une chaîne de Markov. Toutes les minutes, il peut soit changer d'activité, soit continuer celle qu'il était en train de faire. L'appellation processus sans mémoire n'est pas du tout exagérée pour parler de Doudou.
   – Quand il dort, il a 9 chances sur 10 de ne pas se réveiller la minute suivante.
   – Quand il se réveille, il y a 1 chance sur 2 qu'il aille manger et 1 chance sur 2 qu'il parte faire de l'exercice.
   – Le repas ne dure qu'une minute, après il fait autre chose.
   – Après avoir mangé, il y a 3 chances sur 10 qu'il parte courir dans sa roue, mais surtout 7 chances sur 10 qu'il retourne dormir.
   – Courir est fatigant ; il y a 8 chances sur 10 qu'il retourne dormir au bout d'une minute. Sinon il continue en oubliant qu'il est déjà un peu fatigué.
2. Ruine du joueur : un joueur commence avec un pactole de s €. Il lance une pièce : si elle tombe sur pile il gagne 1 €, sinon il perd 1 €. Il joue jusqu'à atteindre une fortune de M € ou être ruiné.
3. Marche aléatoire isotrope sur $\mathbb{Z}$
4. Diffusion de Ehrenfest : N particules sont enfermées dans deux boîtes reliées par un tube. À chaque instant (discret), une particule est choisie de façon uniforme et passe d'une boîte à l'autre. On s'intéresse au nombre de particules dans chaque boîte.

Exercice 5 (Temps d'arrêt)

On rappelle qu'une variable aléatoire T est un temps d'arrêt pour une suite de v.a. (X_n) lorsque l'événement $\{T = m\}$ ne dépend que de $X_1, \ldots, X_m$. On fixe (X_n) et on se donne T_1 et T_2 deux temps d'arrêts pour X. Indiquer si les v.a. suivantes sont des temps d'arrêt :

1. une variable aléatoire constante
2. le premier retour en un point
3. le dernier retour en un point
4. $T_1 + k$, où $k \in \mathbb{N}$
5. $T_1 - k$, où $k \in \mathbb{N}$
6. $\max(T_1, T_2)$ et $\min(T_1, T_2)$
7. $T_1 + T_2$ et $T_1 - T_2$
8. $N(t) = \max\{n \in \mathbb{N} \mid X_1 + \cdots + X_n \leqslant t\}$ (on suppose les X_n positifs)
9. $N(t) + 1$

3 Solutions

▸ **Exercice 1**

1. – $\mathcal{E}$ représente les différentes configurations intéressantes du système
   – $\mathcal{S}$ représente les transitions du système, les sources de changement
   – A représente les sources actives dans un état donné du système
   (et ne dépend pas de la transition précédente, contrairement à ce que j'ai dit en TD)
   – $c(\alpha,\iota)$ représente la vitesse de la source α dans l'état ι, c'est à dire que le temps d'attente entre deux transitions α (donné par F_α) est multiplié par un facteur $\frac{1}{c(\alpha,\iota)}$
   – $p(\alpha,\iota,\iota')$ est le probabilité de passer de l'état ι à l'état ι' au moment d'une transition α
   – F_α est la fonction de répartition de la loi du temps qui sépare deux transitions α successives.
2. Les clients sont caractérisés par leurs charges apportées qui n'est utilisée que par les serveurs pour définir le temps entre deux fins de service. Ils sont donc indistinguables dans la salle d'attente et on peut se contenter de ne conserver que leur nombre. On prend donc $\mathcal{E} = \mathbb{N}$ qui dénote le nombre de clients en attente ou en train d'être servis. Les deux événements de la file sont l'arrivée d'un nouveau client et le départ d'un client servi. On a donc $\mathcal{S} = \{in,\ out\}$. Les arrivées sont permanentes et les départs ne sont possible que lorsqu'il y a un client qui est servi d'où $A(0) = \{in\}$ et $A(n+1) = \{in,\ out\}$. Par hypothèse, $c(out,\_) \equiv 1$ et comme les clients arrivent l'un après l'autre, $c(in,\_) \equiv 1$. Enfin, $F_{in} = F_{s_0}$ et $F_{out} = F_{Z_0}$.
3. Pour ajouter des serveurs, il va falloir ajouter des sources car chaque serveur doit avoir son propre décompte du temps de service. Pour identifier quels serveurs sont actifs (et donc lesquels sont disponibles), on code dans $\mathcal{E}$ le statut des serveurs par un booléen. On obtient donc $\mathcal{E} = \mathbb{N}\times\{0,1\}^K$ où K est le nombre de serveurs. Les sources sont $\mathcal{S} = \{in,\ out_i \mid 1 \leqslant i \leqslant K\}$. Les vitesses restent toutes égales à 1. Les sources actives sont $A(k,\overrightarrow{a}) = \{in\}\cup\{out_i \mid a_i = 1\}$. Les probabilités $p(\alpha,\iota,\iota')$ dépendent de la discipline de service choisie. Par exemple, pour un choix aléatoire du serveur, on prendra $p(in,\iota,\iota') = \frac{1}{J}$ avec $\iota' = \iota + e_j$ où j parcourt les J composantes nulles de ι. Dans tous les cas, pour out_i, cela donne $p(out_i,(n+K+1,\overrightarrow{a}),(n+K,\overrightarrow{a})) = 1$ (s'il reste des clients à servir, le serveur i se remet au travail) et $p(out_i,(n+1,\overrightarrow{a}),(n,\overrightarrow{a}-e_i)) = 1$ (s'il n'y a pas de client qui attend, *i.e.* $n < K$, alors le serveur s'arrête). Enfin, les loi ne sont pas modifiées : $F_{in} = F_{s_0}$ et $F_{out_i} = F_{Z_0}$.
4. Pour ce dernier cas, il faut distinguer les clients selon leur type. Néanmoins, au sein d'un même type, ils restent indistinguables donc on peut se contenter de ne conserver que le nombre de clients de chaque type. Chaque serveur devra également indiquer le type du client qu'il est en train de servir car le type influe sur le service (sinon quel intérêt y a-t-il à introduire des types ?). On obtient alors, en notant T le nombre (supposé fini) de types différents :
   – $\mathcal{E} = \mathbb{N}^T\times\{0,\dots,T\}^K$ où le type 0 pour un serveur dénote l'inactivité (les éléments de $\mathcal{E}$ seront notés $(\overrightarrow{w},\overrightarrow{a})$)
   – $\mathcal{S} = \{in\}\cup\left\{out_{i,t}\ \middle|\ 1\leqslant i\leqslant K, 1\leqslant t\leqslant T\right\}$: chaque service par un serveur comporte en plus son type
   – $A(\overrightarrow{w},\overrightarrow{a}) = \{in\}\cup\left\{out_{i,t}\ \middle|\ a_i = t\right\}$
   – $c(out_{i,t},\iota)$ peut varier en fonction du type du client servi
   – $p(\alpha,\iota,\iota')$ va dépendre de la discipline choisie.
   – $F_{in} = F_{s_0}$ et $F_{out_{i,t}} = F_{Z_0(t)}$ en général (la différence entre serveurs se faisant plutôt au niveau de la vitesse)

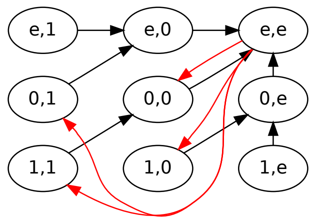
▸ **Exercice 2**

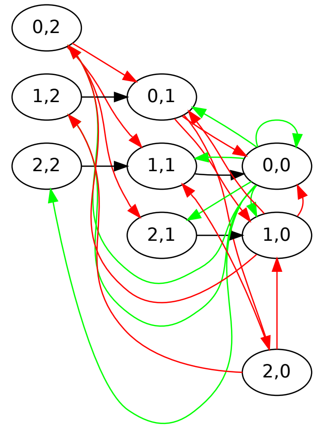
1. Lorsque $W = 1$, en cas de collision, les stations ré-émettent à l'instant d'après (W_i est toujours égal à 0) donc sont à nouveau en collision. Ainsi, la moindre collision bloque tout le système.
2. L'ensemble générique d'états à considérer est : $((\mathbb{N}\cup\{+\infty\}\times W)^S$ où $\mathbb{N}\cup\{+\infty\}$ représente le nombre de messages à transmettre et S est le nombre de stations. Lorsque le nombre de message à transmettre est nul, la station est au repos et il n'est plus nécessaire de conserver des informations pour W_i. En tenant compte de cette observation, l'espace des états devient $(\{\varnothing\}\cup((\mathbb{N}^*\cup\{+\infty\})\times W))^S$.

   Pour $n = 2$, $W = 2$ et 1 message à transmettre, il n'est pas nécessaire de stocker le nombre de messages à envoyer donc les états intéressants sont pour chaque station : e qui indique le repos, $W_i = 0$ avec un message à envoyer et $W_i = 1$ avec un message à envoyer. Cela fait donc neuf états au total.

   Pour $n = 2$, $W = 3$ et une infinité de messages à envoyer, il n'est pas besoin de stocker le nombres de messages à envoyer et il n'y a pas d'état de repos. Ainsi la seule grandeur intéressante est W_i. On a donc à nouveau 9 états possibles.

   Dans les deux figures ci-après, les flèches noires correspondent aux transitions de probabilité 1 alors que les rouges sont uniformes en fonction de leur nombre (quatre ou neuf).





▸ Exercice 3

1. Par indépendance, $\mathbb{P}(Y_m = i \mid Y_n = j) = \mathbb{P}(Y_m = i)$. Or comme Y_i est i.i.d., elle suit la même loi que X_0 donc $\forall i \in \{-1, 1\}$, $\mathbb{P}(Y_m = i) = \mathbb{P}(X_0 = i) = \frac{1}{2}$. On en déduit alors

$$
\begin{aligned}
p_{ij}(m, m+n+r) = \mathbb{P}(Y_{m+n+r} = j \mid Y_m = i) = \mathbb{P}(Y_{m+n+r} = j) = \frac{1}{2} \\
&= \frac{1}{2}\frac{1}{2} + \frac{1}{2}\frac{1}{2} \\
&= \mathbb{P}(Y_{m+n+r} = j \mid Y_{m+n} = 1)\,\mathbb{P}(Y_{m+n} = 1 \mid Y_n = i) + \mathbb{P}(Y_{m+n+r} = j \mid Y_{m+n} = -1)\,\mathbb{P}(Y_{m+n} = -1 \mid Y_n = i) \\
&= \sum_{k \in \{-1,1\}} \mathbb{P}(Y_{m+n+r} = j \mid Y_{m+n} = k)\,\mathbb{P}(Y_{m+n} = k \mid Y_n = i) \\
&= \sum_k p_{ik}(n, n+m)\,p_{kj}(m+n, m+n+r)
\end{aligned}
$$

2. $$\mathbb{P}(Y_{2n+2} = 1 \mid Y_{2n+1} = -1) = \frac{1}{2} \qquad\qquad \mathbb{P}(Y_{2n+2} = 1 \mid Y_{2n+1} = -1,\ Y_{2n} = 1) = 0$$

3. Posons $Z_n = (Y_n, Y_{n+1})$. On peut déjà remarquer que la seconde composante de Z_n impose la première de Z_{n+1}.

$$
\begin{aligned}
\mathbb{P}(Z_{2n+1} = (1,1) \mid Z_{2n} = (1,1)) &= 1 & \mathbb{P}(Z_{2n+2} = (1,1) \mid Z_{2n+1} = (1,1)) &= 0,5 \\
\mathbb{P}(Z_{2n+1} = (1,-1) \mid Z_{2n} = (1,1)) &= 0 & \mathbb{P}(Z_{2n+2} = (1,-1) \mid Z_{2n+1} = (1,1)) &= 0,5 \\
\mathbb{P}(Z_{2n+1} = (-1,1) \mid Z_{2n} = (1,-1)) &= 0 & \mathbb{P}(Z_{2n+2} = (-1,1) \mid Z_{2n+1} = (1,-1)) &= 0,5 \\
\mathbb{P}(Z_{2n+1} = (-1,-1) \mid Z_{2n} = (1,-1)) &= 1 & \mathbb{P}(Z_{2n+2} = (-1,-1) \mid Z_{2n+1} = (1,-1)) &= 0,5
\end{aligned}
$$

▸ Exercice 4

La méthode générale consiste à résoudre le système suivant : $\pi = P\pi$ où P est la matrice de transition de la chaîne de Markov, ce qui correspond exactement à un système de relations de récurrence. On rappelle que cette dernière n'existe que si la chaîne est homogène. On peut réécrire cette équation sous la forme $\pi(P - I) = 0$.

1. La distribution stationnaire est $\frac{1}{181}(160, 8, 13)$, qui vaut approximativement $(0,884,\ 0,044,\ 0,72)$. Doudou passe plus de 88% de son temps à dormir ! Il n'y a ici qu'une unique distribution stationnaire.
2. Il y a deux distributions stationnaires : la ruine ou la fortune (et toute combinaison convexe des deux). Cela peut se voir facilement en remarquant que ces deux états sont absorbants donc s'il y a une probabilité non nulle d'y arriver, ils absorbent les probabilités des états qui y mènent. Comme la chaîne (privée de ces extrémités) est irréductible, aucun autre état ne peut avoir de probabilité non nulle. Il y a dans ce cas une infinité de distributions stationnaires.

3. Par symétrie, tous les états doivent avoir même probabilité. Et de fait, $c(\dots, 1, \dots)$ pour $c \in \mathbb{R}^+$ est une mesure stationnaire mais n'est pas une distribution car la condition de somme égale à un est irréalisable. Il n'y a pour cet exemple aucune distribution stationnaire.
4. Il est légitime d'un point de vue physique de cherche une solution qui soit réversible temporellement. On utilise donc la condition de *réversibilité à l'équilibre* et on résout $\pi_i p_{ij} = \pi_j p_{ji}$. Cette équation n'a d'intérêt que pour $j \in \{i-1, i+1\}$ et par symétrie on peut se limiter à un seul de ces cas. On cherche donc à résoudre $\pi_i \frac{m-i}{m} = \pi_{i+1} \frac{i+1}{m}$. La solution est de la forme $\binom{m}{i}$ à un facteur de normalisation près. En effet :

$$\frac{m-i}{m}\pi_i = \frac{m-i}{m}\binom{m}{i} = \frac{m-i}{m}\frac{m!}{(m-i)!\,i!} = \frac{(m-1)!}{(m-i-1)!\,i!} = \frac{i+1}{m}\frac{m!}{(m-i-1)!\,(i+1)!} = \frac{i+1}{m}\binom{m}{i+1} = \frac{i+1}{m}\pi_{i+1}$$

On peut vérifier que c'est bien une mesure stationnaire de la chaîne de Markov :

$$\sum_k \pi_k p_{ki} = \pi_{i-1}\frac{m-i+1}{m} + \pi_{i+1}\frac{i+1}{m} = \binom{m}{i-1}\frac{m-i+1}{m} + \binom{m}{i+1}\frac{i+1}{m} = \frac{(m-1)!}{(m-i)!\,(i-1)!} + \frac{(m-1)!}{(m-i-1)!\,i!}$$

$$= (m-1)!\frac{i+(m-i)}{(m-i)!\,i!} = \frac{m!}{(m-i)!\,i!} = \binom{m}{i} = \pi_i$$

Le coefficient de normalisation est

$$\left(\sum_i \pi_i\right)^{-1} = \left(\sum_i \binom{m}{i}\right)^{-1} = (2^m)^{-1} = 2^{-m}$$

▸ **Exercice 5** On ne donne ici qu'une intuition et non une démonstration.

1. oui, elle ne dépend de rien !
2. oui, premier exemple vu en cours
3. non, premier contre-exemple vu en cours
4. oui, c'est un cas particulier de la somme avec une v.a. constante
5. non, car il faut savoir que l'événement va se produire dans k pas
6. oui, on attend le second (resp. premier) des deux temps d'arrêts
7. oui pour la somme car on connaît sa valeur au temps $\max(T_1, T_2) \leqslant T_1 + T_2$ mais non pour la différence (c'est déjà faux lorsque T_2 est une constante)
8. non, il faudrait savoir que la valeur du prochain X_n va faire dépasser la somme.
9. oui, cela revient à au temps d'arrêt $N'(t) = \min\{n \in \mathbb{N} \mid X_1 + \cdots + X_n > t\}$