TP 8

Chaînes de Markov

Compétences attendues.

- ✓ Savoir écrire la matrice de transition d'une chaîne de Markov.
- ✓ Savoir simuler n termes d'une chaîne de Markov.
- ✓ Savoir déterminer la loi stationnaire d'une chaîne de Markov.

Mathieu Mansuy - Professeur de Mathématiques en ECS2 au Lycée Louis Pergaud (Besançon)
`mansuy.mathieu@hotmail.fr`

1 Un exemple introductif : évolution sociologique d'une société

À des fins d'étude sociologique, une population d'individus est divisée en trois classes sociales :

- *upper class* (classe n°1) ;
- *middle class* (classe n°2) ;
- *lower class* (classe n°3).

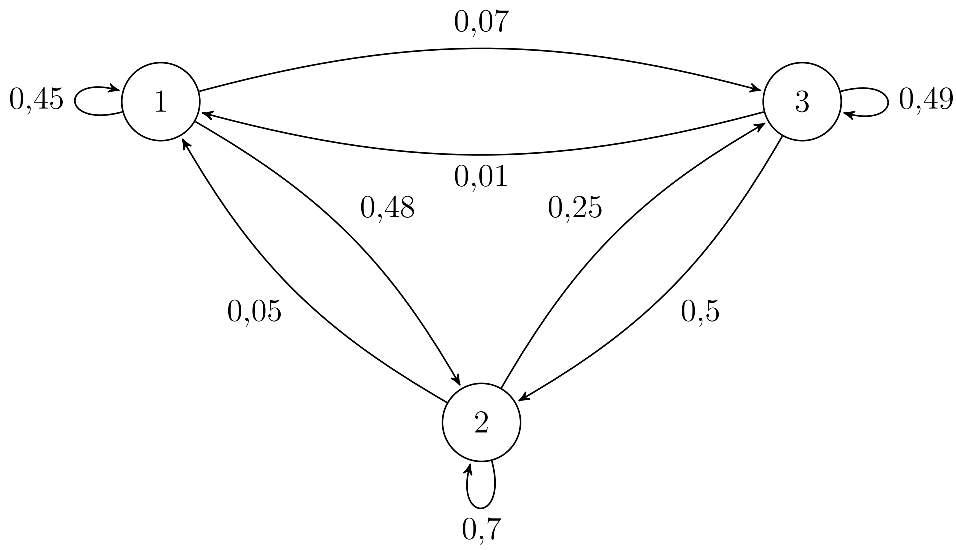
On souhaite étudier l'évolution des classes sociales des individus au fil des générations.

Dans la population étudiée, on considère qu'un enfant :

- né dans la *upper class* a, à l'âge adulte, 45% de chances de faire partie de la *upper class*, 48% de chances de faire partie de la *middle class* et 7% de chances de faire partie de la *lower class*,
- né dans la *middle class* a, à l'âge adulte, 5% de chances de faire partie de la *upper class*, 70% de chances de faire partie de la *middle class* et 25% de chances de faire partie de la *lower class*,
- né dans la *lower class* a, à l'âge adulte, 1% de chances de faire partie de la *upper class*, 50% de chances de faire partie de la *middle class* et 49% de chances de faire partie de la *lower class*.

On suppose que ces probabilités d'évolution sociologique ne changent pas au cours des générations.

On peut représenter ce modèle sociologique par un graphe dont les sommets sont les *états*, c'est-à-dire 1, 2 et 3.



Soit $x_0 \in [\![1, 3]\!]$. Considérons un individu de la classe sociale x_0 à la génération 0, et notons X_0 la variable aléatoire certaine égale à x_0. Pour tout $n \in \mathbb{N}$, on note X_n la variable aléatoire indiquant la classe sociale d'un de ses descendants à la n-ème génération.

La suite $(X_n)_{n\in\mathbb{N}}$ est ce qu'on appelle *une chaîne de Markov* : la loi de X_{n+1} ne dépend que de celle de X_n (elle ne dépend pas des lois des variables précédentes $X_0, \ldots, X_{n-1}$).

1. En supposant que $X_0 = 1$ (c'est-à-dire que l'individu de la génération 0 est de la *upper class*), déterminer les lois de X_1 et X_2.

2. Simulation naïve.

   (a) Compléter la fonction `classe_suivante(i)` afin qu'elle prenne en entrée un paramètre $i \in [\![1,3]\!]$ correspondant à la classe sociale d'un individu à l'âge adulte, et qu'elle simule la classe social de son enfant (c'est-à-dire un élément de $[\![1,3]\!]$ en suivant les règles ci-dessus).

```
function j=classe_suivante(i)
  j=i
  p=rand()
  if i==1 then
    if p<=0.48 then
        j=2
    else if p<=0.55 then
        j=3
        end
    end
  end
  if i==*** then
    if p<=*** then
        j=***
    else if p<=*** then
        j=***
        end
    end
  end
  if i==*** then
    if p<=*** then
        j=***
    else if p<=*** then
        j=***
        end
    end
  end
endfunction
```

   (b) À l'aide de la fonction `classe_suivante`, simuler l'évolution sociologique d'un individu de la *upper class* sur 20 générations et représenter graphiquement cette trajectoire.

   Faire de même avec un individu de la *middle class* puis de la *lower class*.

L'écriture d'une fonction de simulation peut rapidement être laborieuse si les états et les règles sont plus nombreuses. Une manière commode de s'en sortir plus aisément est de représenter notre chaîne de Markov à l'aide de la matrice suivante, appelée *matrice de transition* :

$$A = \begin{pmatrix} P_{[X_n=1]}(X_{n+1}=1) & P_{[X_n=2]}(X_{n+1}=1) & P_{[X_n=3]}(X_{n+1}=1) \\ P_{[X_n=1]}(X_{n+1}=2) & P_{[X_n=2]}(X_{n+1}=2) & P_{[X_n=3]}(X_{n+1}=2) \\ P_{[X_n=1]}(X_{n+1}=3) & P_{[X_n=2]}(X_{n+1}=3) & P_{[X_n=3]}(X_{n+1}=3) \end{pmatrix}.$$

3. (a) Déterminer explicitement la matrice A.

   Calculer la somme de chaque colonne de A. Que remarquez-vous ? Est-ce surprenant ?

   (b) On peut simuler aisément l'évolution sociologique d'un individu à l'aide de la commande suivante.

Définition.

Soit `A` la matrice de transition de la chaîne de Markov.

L'instruction `grand(n,"markov",A',x0)` rend un vecteur ligne contenant `n` simulations correspondant à la trajectoire d'un individu partant de l'état initial `x0`.

Par exemple, si A est la matrice de transition correspondant à l'évolution de la société considérée, alors `grand(20,"markov",A',1)` simule 20 générations d'un individu appartenant initialement à la *upper class*. Attention de bien rentrer la transposée `A'` de la matrice A dans la commande `grand`.

Simuler 20 générations d'un individu de la *upper class* à la génération 0. Quelle est la classe sociale au bout de 20 générations ?

(d) Construire un vecteur u contenant 1000 simulations de la classe sociale au bout de 20 générations, toujours en partant d'un individu de la *upper class*.

Que renvoie la commande `length(find(u==1))` ? Déterminer la proportion de 1, 2 et 3 dans u.

Lancer plusieurs fois le programme et commenter. Recommencer en partant d'un individu de la *middle class* et de la *lower class*.

Pour $n \in \mathbb{N}$, on s'intéresse désormais à la loi de X_n et on pose :

$$a_n = P(X_n = 1),\ b_n = P(X_n = 2),\ c_n = P(X_n = 3) \text{ et } U_n = \begin{pmatrix} a_n \\ b_n \\ c_n \end{pmatrix}.$$

On supposera qu'à la génération 0, la proportion de la population dans chacune des classes *upper class*, *middle class* et *lower class* est respectivement 20%, 70% et 10%, de sorte que $a_0 = 0.2$, $b_0 = 0.7$ et $c_0 = 0.1$.

4. (a) À l'aide de la formule des probabilités totales, montrer que pour tout $n \in \mathbb{N}$, on a $U_{n+1} = AU_n$.

   (b) En déduire que pour tout $n \in \mathbb{N}$, on a $U_n = A^n U_0$.

   (c) À l'aide du résultat précédent, déterminer la probabilité qu'un individu de la 30-ème génération pris au hasard soit de la *upper class* ? et un de la 20-ème génération soit de la *lower class* ?

   (d) On se propose de représenter graphiquement les lois de X_1, X_3, X_5, X_{10} et X_{50}. Pour cela on utilise les lignes de codes suivantes.

```
--> U0 = [0.2 ; 0.7 ; 0.1]
--> bar((1:3)',[A*U0,(A^3)*U0,(A^5)*U0,(A^10)*U0,(A^50)*U0])
```

   Éxécuter ce code, et interpréter le résultat. À quoi correspond chaque couleur ?

La suite $(U_n)_{n\in\mathbb{N}}$ tend visiblement vers un vecteur-limite U.

5. Justifions ce résultat.

   (a) À l'aide de `Scilab`, déterminer une matrice diagonale D et une matrice inversible P telles que $A = PDP^{-1}$.

On pourra pour cela utiliser l'instruction suivante.

Définition.

Soit `A` une matrice.

- `spec(A)` donne les valeurs propres de A.
- Si A est diagonalisable alors `[P,D]=spec(A)` rend une matrice P inversible et une matrice diagonale D telles que $A = PDP^{-1}$.

(b) Justifier que pour tout $n \in \mathbb{N}$, $A^n = PD^nP^{-1}$.

(c) Déterminer la limite de $(D^n)_{n\in\mathbb{N}}$ lorsque n tend vers l'infini puis celle de $(A^n)_{n\in\mathbb{N}}$. En déduire la limite U de $(U_n)_{n\in\mathbb{N}}$.

(d) Montrer que $U = AU$.

La suite (U_n) tend vers un vecteur U satisfaisant $U = AU$: la loi correspondant à U est appelée la loi stationnaire de la chaîne de Markov.

6. On suppose désormais qu'à la génération 0, la proportion de la population dans chacune des classes *upper class*, *middle class* et *lower class* est respectivement 5%, 50% et 45%.

   Déterminer la limite de $(U_n)_{n\in\mathbb{N}}$ dans ce cas. Que constate-t-on ?

   La loi stationnaire ne dépend pas de U_0.

7. Dans ce modèle, la matrice de transition ne dépend pas de n. Cela vous paraît-il plausible ?

2 Théorie des chaînes de Markov

2.1 Définition

Définition.

Soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires définies sur un même espace probabilisé $(\Omega, \mathscr{A}, P)$ à valeurs dans un ensemble E.

On dit que $(X_n)_{n\in\mathbb{N}}$ est une *chaîne de Markov* lorsque pour tout $n \in \mathbb{N}$, la loi de X_{n+1} ne dépend que de la loi de X_n (et non des lois des variables précédentes), soit de manière formelle :

$$\forall (i_0, \ldots, i_n, i_{n+1}) \in E^{n+2},\ P_{[X_0=i_0]\cap\cdots\cap[X_n=i_n]}([X_{n+1} = i_{n+1}]) = P_{[X_n=i_n]}(X_{n+1} = i_{n+1})$$

On parle de *processus sans mémoire* : la loi du future (instant $n+1$) ne dépend que du présent (instant n) et non du passé (instants $0, \ldots, n-1$). Les éléments de E sont alors appelés les *états* de la chaîne de Markov.

Dans toute la suite, $(X_n)_{n\in\mathbb{N}}$ désigne une chaîne de Markov :

- *à espace d'états finis* ce qui signifie que l'ensemble E est fini, noté $E = \{e_1, e_2, \ldots, e_p\}$ avec $p \geq 1$,
- *homogène* ce qui signifie que les probabilités $P_{[X_n=a]}(X_{n+1} = b)$ pour $(a, b) \in E^2$ ne dépendent pas de n,
- *irréductible* ce qui signifie que sur le graphe, il y a toujours un chemin pour passer d'un état à un autre.

2.2 Matrice de transition

Définition.

On appelle *matrice de transition* la matrice $A = (a_{i,j}) \in \mathcal{M}_p(\mathbb{R})$ définie par :

$$A = \begin{pmatrix} P_{[X_n=e_1]}(X_{n+1}=e_1) & P_{[X_n=e_2]}(X_{n+1}=e_1) & \dots & P_{[X_n=e_p]}(X_{n+1}=e_1) \\ P_{[X_n=e_1]}(X_{n+1}=e_2) & P_{[X_n=e_2]}(X_{n+1}=e_2) & \dots & P_{[X_n=e_p]}(X_{n+1}=e_2) \\ \vdots & \vdots & & \vdots \\ P_{[X_n=e_1]}(X_{n+1}=e_p) & P_{[X_n=e_2]}(X_{n+1}=e_p) & \dots & P_{[X_n=e_p]}(X_{n+1}=e_p) \end{pmatrix}$$

Le coefficient $a_{i,j}$ est la *probabilité de transition de l'état e_j à l'état e_i*.

Mise en garde.

Dans la littérature, la matrice de transition désigne fréquemment la transposée de la matrice précédente. Il faut dans ce cas transposer les matrices dans tous les résultats qui suivent, et remplacer *colonne* par *ligne*.

Propriété 1

(1) A est une *matrice stochastique par colonne* : tous ses cœfficients sont positifs et la somme de chaque colonne vaut 1.

(2) 1 est valeur propre de ${}^tA$ et donc de A.

Preuve.

□

Mise en garde.

On a montré que 1 est valeur propre de A, et plus précisément que $\dim E_1(A) = \dim E_1({}^tA)$. Attention cependant, il n'existe pas de lien autre immédiat entre $E_1(A)$ et $E_1({}^tA)$. En particulier, on a $V = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \in E_1({}^tA)$, mais $V \notin E_1(A)$ en général.

Remarques. D'autres résultats sur les matrices stochastiques par ligne ou par colonne ont été obtenus en TD et dans le sujet d'EM Lyon 2010. Rappelons notamment que si A est une matrice stochastique, alors $\mathrm{Spec}(A) \subset [-1, 1]$.

Pour tout $n \in \mathbb{N}$, on s'intéresse à la loi de X_n et on note $U_n = \begin{pmatrix} P(X_n = e_1) \\ P(X_n = e_2) \\ \vdots \\ P(X_n = e_p) \end{pmatrix} \in \mathscr{M}_{p,1}(\mathbb{R})$.

Propriété 2

On a :

$$\forall\, n \in \mathbb{N},\ U_{n+1} = AU_n \quad \text{ce qui donne par récurrence} \quad \forall\, n \in \mathbb{N},\ U_n = A^n U_0.$$

Preuve.

□

2.3 Comportement limite

Vocabulaire. Soit $U \in \mathscr{M}_{p,1}(\mathbb{R})$. On dit que U est un vecteur-colonne *stochastique* (ou *probabiliste*) si tous les cœfficients de U sont positifs et si leur somme est égale à 1.

Rappel. U est un vecteur-colonne stochastique si et seulement si il définit une loi de probabilité, c'est-à-dire qu'il existe une variable aléatoire X telle que $X(\Omega) = E$ et

$$U = \begin{pmatrix} P(X = e_1) \\ P(X = e_2) \\ \vdots \\ P(X = e_p) \end{pmatrix}.$$

Définition.

Soit $U \in \mathscr{M}_{p,1}(\mathbb{R})$ un vecteur-colonne stochastique. La loi correspondant à U est appelée *loi stationnaire* ou *état stable* lorsque :

$$AU = U.$$

Remarque.

- Si la loi de U est stationnaire, alors U est un vecteur propre de A associé à la valeur propre 1.
- Si à un instant $n_0 \in \mathbb{N}$, on a $U_{n_0} = U$ alors pour tout $n \geq n_0$, on a $U_n = A^{n-n_0}U = U$ et la suite $(U_n)_{n\in\mathbb{N}}$ est stationnaire.

Théorème 3 (*Théorème de Perron-Fröbenius*)

Soit A une matrice de transition d'une chaîne de Markov. Alors (sous quelques hypothèses techniques) :

- Il existe un unique vecteur-colonne stochastique U tel que $AU = U$.
- La suite $(U_n)_{n\in\mathbb{N}}$ tend vers U et ceci indépendamment de l'état initial U_0. En particulier, la suite $(X_n)_{n\in\mathbb{N}}$ converge en loi vers l'unique loi stationnaire définie par U.

Remarque. Si on prend $U_0 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$ alors pour tout $n \in \mathbb{N}$, $U_n = A^n U_0$ est la première colonne de A^n. D'après le théorème précédent, on obtient donc que la première colonne de A^n tend vers U. On montre de même que chaque colonne de A^n tend vers U (en prenant U_0 égal à chaque vecteur de la base canonique).

Exemple. Vérifier sur l'exemple précédent de l'évolution sociologique d'une société, qu'il existe bien une unique loi stationnaire pour cette chaîne de Markov et la déterminer. Cela correspond-t-il à ce qu'on attendait ?

3 Exemple du système Bonus - Malus des assurances

Le système des bonus-malus chez les assureurs des conducteurs a été introduit par la loi. Il poursuit essentiellement deux buts :

- responsabiliser les assurés et les inciter à plus de prudence au volant en pénalisant par une augmentation de prime le responsable d'un ou de plusieurs accidents tandis que dans la situation inverse, l'assuré bénéficiera d'une réduction de prime.
- ajuster le montant de la prime au cours du temps afin que celui-ci reflète le risque réel que représente l'assuré.

On s'intéresse ici à un modèle où l'assureur propose trois taux à ses assurés :

- taux A (Bonus 50%) ;
- taux B (tarif de base) ;
- taux C (Malus 100%).

Les nouveaux assurés font leur entrée dans le système au niveau B. Chaque année sans sinistre est récompensée par un passage de C vers B ou de B vers A. Chaque année avec sinistre est pénalisée par un passage de A vers B ou de B vers C.

Un bon conducteur a chaque année 5% de chance d'avoir un sinistre et un mauvais conducteur 15%. On considère qu'il y a autant de bons que de mauvais conducteurs.

Étude de la trajectoire d'un bon conducteur au cours du temps

Pour tout $n \in \mathbb{N}$, on note X_n la variable aléatoire qui à un bon conducteur associe son taux d'assurance lors de la n-ème année.

1. Justifier que $(X_n)_{n\in\mathbb{N}}$ est une chaîne de Markov et représenter son graphe.

2. Déterminer sa matrice de transition A.

3. Représenter avec `Scilab` une trajectoire d'un bon conducteur sur 30 années.

4. Pour tout $n \in \mathbb{N}$, on pose $U_n = \begin{pmatrix} P(X_n = A) \\ P(X_n = B) \\ P(X_n = C) \end{pmatrix}$. On va calculer la loi stationnaire U de cette chaîne de Markov en utilisant que la suite $(U_n)_{n\in\mathbb{N}}$ tend vers U.
   (a) Diagonaliser A. En déduire la limite de $(A^n)_{n\in\mathbb{N}}$.
   (b) En déduire la loi stationnaire U de la chaîne de Markov.

Étude de la trajectoire d'un mauvais conducteur au cours du temps

5. (a) Déterminer la matrice de transition B dans ce cas.
   (b) Calculer la loi stationnaire V de cette chaîne de Markov. On utilisera cette fois-ci que V est l'unique vecteur colonne stochastique solution de $BV = V$.

Qualité du système Bonus-Malus mis en place par l'assureur

6. Calculer la probabilité, pour un conducteur au taux C après un nombre d'années assez important, d'être un mauvais conducteur. Le niveau C est-il bien discriminant ?

7. Calculer la probabilité, pour un conducteur au taux A après un nombre d'années assez important, d'être un bon conducteur. Que pensez-vous du système de Bonus-Malus mis en place par l'assureur ?

4 Exemple du Google PageRank

En 1997, deux étudiants de Stanford, Larry Page et Sergueï Brin ont l'idée d'utiliser les chaînes de Markov pour créer un moteur de recherche. Le problème des moteurs de recherche de l'époque est le suivant :

- Une fois qu'on a associé une liste de mots-clés à chaque page du Web, comment les « trier » par ordre d'importance ?
- Comment choisir, pour un mot clé donné les pages les plus susceptibles d'intéresser l'internaute ?

L'idée des fondateurs de Google est d'analyser pour cela les liens existants entre les différentes pages : un lien d'une page i vers une page j est ainsi vu comme un « vote de confiance » pour la page j de la part de la page i.

Pour mesurer le nombre de liens vers une page, et donc sa popularité, ils définissent un algorithme appelé *PageRank* dont voici le fonctionnement : on numérote l'ensemble des pages d'Internet de 1 à N (N est bien évidemment très grand, de l'ordre de 10^{13}). On modélise les déplacements d'un internaute par une chaîne de Markov $(X_n)_{n\in\mathbb{N}}$, où X_n est le numéro de la page où se situe l'internaute après n déplacements. On considère que :

- à l'instant 0, l'internaute choisit une page internet au hasard parmi les N pages ;
- lorsque l'internaute est sur une page j,
  - avec probabilité $c \in]0,1[$, il quitte la page et se dirige vers une page quelconque du web choisie aléatoirement ;
  - avec probabilité $1-c$, il se déplace aléatoirement vers une des ℓ_j pages accessibles depuis la page j (on considère qu'une page sans lien pointe sur elle-même).

1. Justifier que la matrice de transition A associée est donnée par :

$$\forall (i,j) \in [\![1,N]\!]^2, \quad A_{i,j} = \begin{cases} \dfrac{c}{N} + \dfrac{1-c}{\ell_j} & \text{si la page } j \text{ pointe vers la page } i \\ \dfrac{c}{N} & \text{sinon.} \end{cases}$$

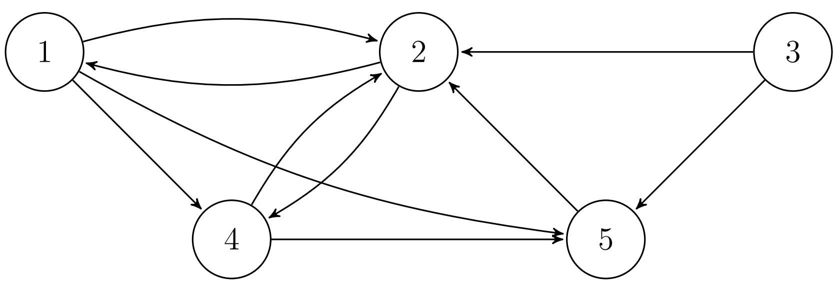
2. Le paramètre c est confidentiel chez Google. Une étude démontre cependant qu'un internaute relance sa navigation au hasard au bout de 6 pages en moyenne. Justifier qu'il est pertinent de choisir $c = 0,15$, ce qu'on fera dans la suite.

L'algorithme PageRank de Google consiste alors à déterminer la loi stationnaire U associée à cette chaîne de Markov, ce qui correspond à chercher un vecteur propre associé à la valeur propre 1 de A. Cette recherche est difficile de part la taille gigantesque de A qui rend impossible un calcul direct. On utilise pour cela des algorithmes d'approximation, exploitant notamment le fait que $A - \frac{c}{N} \times \text{ones}(N,N)$ possède beaucoup de zéros.

On définit alors le *PageRank* de la page i par le nombre $U(i) = \lim\limits_{n\to+\infty} P(X_n = i)$. Il s'agit de la probabilité, en surfant au hasard, d'arriver sur cette page. On remarquera que plus une page possède de liens pointant vers elle, plus son PageRank est élevé, même si ce n'est pas le seul facteur à prendre en compte. Ainsi, une page avec un PageRank élevé est vue par Google comme plus pertinente étant donné qu'elle possède davantage de liens qui pointent vers elle.

3. Afin d'augmenter le PageRank de mon site, je vais avoir besoin de davantage de liens pointant vers lui. Vaut-il mieux pour moi avoir un lien sur chacun des 27 blogs de mes étudiants, ou un seul lien sur la page d'accueil du site du Monde ?

4. On considère un réseau de $N = 5$ pages différentes, les liens entre elles étant représentés dans le graphe ci dessous :



(a) Écrire la matrice de transition associée.

(b) Vérifier que la chaîne de Markov admet une unique loi stationnaire que l'on déterminera.

(c) Y-a-t-il convergence en loi de la chaîne de Markov (X_n) vers la loi stationnaire ?

(d) Déterminer le PageRank de chaque page internet, et les ranger par indice de popularité. Commenter.