Informatique

TP 4

Chaînes de Markov

Partie 1 : exemple introductif

Exercice 1 : épidémiologie

On modélise l'évolution d'une maladie en classant les individus en trois groupes : S (individus sains et non immunisés), I (individus immunisés) et M (individus malades).
On note s_n, i_n et m_n les proportions d'individus dans chacun des trois groupes à l'instant n.
On considère qu'à l'instant 0 (avant le début de l'épidémie), tous les individus sont dans l'état S.
On constate que la moitié des individus appartenant au groupe S à l'instant n restent dans ce groupe à l'instant $n+1$, alors que que l'autre moitié tombe malade. Parmi les individus immunisés à l'instant n, 10% passent dans le groupe S à l'instant $n+1$ et les 90% autres restent immunisés. Enfin, 20% des malades à l'instant n le restent à l'instant $n+1$ et 80% guérissent et deviennent immunisés.

1. Représenter le modèle épidémiologique par un graphe orienté valué, c'est-à-dire un ensemble de sommets reliés par des arêtes orientées munies de coefficients.
2. Pour $n \in \mathbb{N}$, exprimer s_{n+1} (resp. i_{n+1} et m_{n+1}) en fonction de s_n, i_n et m_n.
3. On pose :

$$\forall n \in \mathbb{N}, \quad \mathrm{U}_n = \begin{pmatrix} s_n \\ i_n \\ m_n \end{pmatrix}.$$

   a. Préciser le vecteur U_0, qui donne la *loi initiale*.
   b. Déterminer la matrice Q, appelée *matrice de transition*, telle que $\mathrm{U}_{n+1} = \mathrm{QU}_n$ pour tout $n \in \mathbb{N}$.

La commande Scilab `grand(n,'markov',Q',x0)` simule les n premiers états (numérotés 1 pour S, 2 pour I et 3 pour M) d'un individu initialement dans l'état x_0 ; on parle de trajectoire de l'individu.
La commande `grand(n,'markov',Q',X0)` renvoie une matrice dont les lignes sont des trajectoires indépendantes, les états initiaux étant donnés dans le vecteur X_0.
Remarque. On notera que le troisième argument de la commande `grand` est la *transposée* de la matrice de transition.

4. Utiliser la commande `grand` pour simuler les trajectoires de 10 individus sur 15 unités de temps.
5. Pour $n \in \mathbb{N}$, exprimer U_n en fonction de Q et U_0.
6. **a.** Sous Scilab, créer une matrice E à 3 lignes et 15 colonnes, dont la j-ième colonne est donnée par U_j pour tout $j \in [\![1, 15]\!]$. Que voit-on apparaître ?
   b. Sur une même figure, superposer les représentations graphiques des suites (s_n), (i_n) et (m_n). Est-ce cohérent ?

Étant donnée une matrice A diagonalisable, la commande Scilab `[P,D]=spec(A)` renvoie un couple de matrices (P, D) où P est inversible et D diagonale telles que $\mathrm{P}^{-1}\mathrm{AP} = \mathrm{D}$.

7. Déterminer expérimentalement puis mathématiquement la limite de la suite matricielle (Q^n). Que peut-on en déduire pour U_n ?

Partie 2 : chaînes de Markov

On considère une suite $(\mathrm{X}_n)_{n\in\mathbb{N}}$ de variables aléatoires à valeurs dans un ensemble E fini (ou dénombrable), dont les éléments sont appelés *états*. On dit que $(\mathrm{X}_n)_{n\in\mathbb{N}}$ est une *chaîne de Markov* si, pour tout $n \in \mathbb{N}$,

$$\forall (x_0, \dots, x_{n+1}) \in \mathrm{E}^{n+2}, \quad \mathbb{P}(\mathrm{X}_{n+1} = x_{n+1} | \mathrm{X}_0 = x_0, \mathrm{X}_1 = x_1, \dots, \mathrm{X}_n = x_n) = \mathbb{P}(\mathrm{X}_{n+1} = x_{n+1} | \mathrm{X}_n = x_n).$$

On suppose que la chaîne est *homogène*, c'est-à-dire que pour tous $n \in \mathbb{N}$ et $x, y \in \mathrm{E}$, la probabilité $\mathbb{P}(\mathrm{X}_{n+1} = y | \mathrm{X}_n = x)$ ne dépend pas de n. On appelle alors (lorsque E est fini) *matrice de transition* de la chaîne la matrice Q ayant autant de colonnes et de lignes qu'il y a d'états possibles, et dont le coefficient en ligne $y \in \mathrm{E}$ et en colonne $x \in \mathrm{E}$ est $\mathrm{Q}_{y,x} = \mathbb{P}(\mathrm{X}_{n+1} = y | \mathrm{X}_n = x)$.
Remarque. On pourra si on le souhaite numéroter les états de 1 à N pour travailler sur des indices de lignes et de colonnes entiers. C'est ainsi que procède Scilab.
Attention. Dans la littérature, la matrice de transition désigne souvent la transposée de la matrice précédente.

Exercice 2 : matrices stochastiques

1. Justifier que la matrice de transition Q est à coefficients réels positifs et vérifie :

$$\forall x \in \mathrm{E}, \quad \sum_{y \in \mathrm{E}} \mathrm{Q}_{y,x} = 1.$$

On dit que Q est *stochastique par colonnes* (la somme des coefficients sur chaque colonne est égale à 1).

2. a. Justifier que 1 est valeur propre de Q.
b. Montrer que toute valeur propre de Q est de module inférieur ou égal à 1.

Pour tout $n \in \mathbb{N}$, on note U_n le vecteur colonne donnant la loi de X_n : ses coefficients sont les probabilités $\mathbb{P}(\mathrm{X}_n = x)$, $x \in \mathrm{E}$.

Exercice 3 : matrice de transition et loi initiale déterminent la loi de la chaîne

1. Montrer que :

$$\forall n \in \mathbb{N}, \quad \mathrm{U}_{n+1} = \mathrm{Q}\mathrm{U}_n.$$

2. En déduire que :

$$\forall n \in \mathbb{N}, \quad \mathrm{U}_n = \mathrm{Q}^n \mathrm{U}_0.$$

Exercice 4 : lois stationnaires et états stables

On appelle *loi stationnaire* de la chaîne tout matrice colonne Π stochastique (i.e. à coefficients positifs dont la somme est égale à 1) telle que $\mathrm{Q}\Pi = \Pi$.

1. Justifier que si, à un instant n_0, $\mathrm{U}_{n_0} = \Pi$, alors $\mathrm{U}_n = \Pi$ pour tout $n \geqslant n_0$: la distribution de probabilité de X_n est constante à partir du rang n_0 et l'on parle loi stationnaire (ou d'état stable si l'on considère qu'il s'agit du vecteur des proportions d'individus dans chacun des états).

2. Justifier que si la suite (U_n) converge, alors sa limite est une loi stationnaire.

L'étude de la convergence des chaînes de Markov est un sujet difficile. On admet les résultats suivants :

➢ Sous de bonnes hypothèses, on a existence et unicité de la loi stationnaire Π et la fréquence avec laquelle la chaîne passe par un état x donné entre les instants 1 et n converge presque sûrement vers $\Pi(x)$ lorsque $n \to \infty$ (théorème ergodique) :

$$\forall x \in \mathrm{E}, \quad \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \mathbb{1}_{[\mathrm{X}_k = x]} = \Pi(x) \text{ p.s.}$$

➢ Sous des hypothèses encore plus fortes, la suite (X_n) converge en loi vers une variable aléatoire de loi Π :

$$\forall x \in \mathrm{E}, \quad \lim_{n\to\infty} \mathbb{P}(\mathrm{X}_n = x) = \Pi(x).$$

Partie 3 : Deuxième exemple

Exercice 5 : sociologie

On considère une société (fictive, obtenue par simplification du modèle romain) comportant 4 classes sociales : les esclaves, les hommes libres, les citoyens et les hauts fonctionnaires.
On étudie l'évolution au fil des ans d'un individu (ou d'une lignée si l'on préfère, avec les hypothèses simplificatrices que cela implique) : chaque année,

- ➢ sur 98 esclaves, 1 devient libre (achète sa liberté ou est affranchi), les autres restent esclaves ;
- ➢ sur 73 hommes libres, 2 vendent leur liberté (pour racheter leurs dettes), 6 deviennent citoyens (par mérite ou achat de titre), les autres restent libres ;
- ➢ les citoyens sont citoyens à vie ;
- ➢ sur 13 citoyens, 1 est élu (aux élections annuelles) pour devenir haut fonctionnaire pour un mandat de deux ans ;
- ➢ au terme de son mandat de 2 ans, 1 haut fonctionnaire sur 8 est réélu, les autres redeviennent de simples citoyens.

1. Justifier que le modèle décrit ci-dessus est markovien et le représenter par un graphe valué.

2. Écrire la matrice de transition Q associée. Simuler les trajectoires sur 10 ans de 3 esclaves, 3 hommes libres et 3 citoyens.

3. Représenter graphiquement les suites donnant les probabilités d'appartenir à chacun des états pour un esclave sur les 50 premières années. Faire de même pour un homme libre puis un citoyen.

4. Calculer les puissances de Q et voir s'il apparaît des états stables.

Partie 4 : l'algorithme PageRank de Google

On s'intéresse dans cette partie à l'algorithme *PageRank* utilisé par Google pour classer les pages web par indice de popularité.
On numérote les pages du web de 1 à N (l'entier N est de l'ordre de 10^{13}) et l'on modélise les déplacements d'un internaute fictif 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).

Le paramètre c est confidentiel chez Google. La valeur $c = 0{,}15$ est communément admise comme un bon choix : on peut en effet montrer que la valeur $\frac{1}{c} \simeq 6$ correspond au nombre moyen de liens suivis par le surfeur avant de relancer sa navigation à partir de n'importe quelle page au hasard.

Exercice 6 : matrice de transition

Justifier que la matrice de transition Q associée est donnée par :

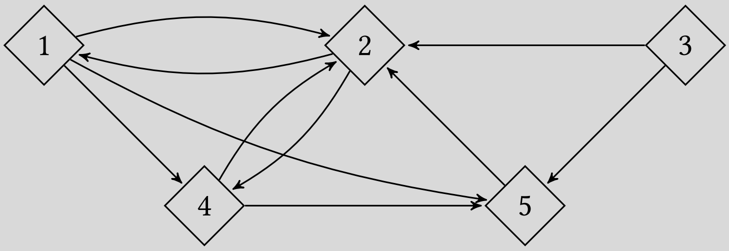
$$\forall (i,j) \in [\![1,\mathrm{N}]\!]^2, \quad Q_{i,j} = \begin{cases} \frac{c}{\mathrm{N}} + \frac{1-c}{\ell_j} & \text{si la page } j \text{ pointe vers la page } i \\ \frac{c}{\mathrm{N}} & \text{sinon} \end{cases}.$$

L'algorithme PageRank de Google consiste alors à déterminer la loi stationnaire associée à cette chaîne, ce qui revient à trouver le vecteur stochastique Π tel que $Q\Pi = \Pi$. Il suffit donc de résoudre un banal système linéaire, la difficulté venant du fait qu'il a plusieurs millions d'inconnues !

Le théorème ergodique permet alors d'interpréter $\Pi(i)$ comme étant la fréquence moyenne de passages de l'internaute par la page i. La valeur sera donc d'autant plus grande que la page est importante (du point de vue de l'architecture du réseau), et réciproquement.

Exercice 7 : un exemple

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



1. Écrire la matrice de transition associée.

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

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

4. Ranger les pages internet par indice de popularité. Commenter.

Partie 5 : système des bonus-malus en assurance

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.

Exercice 8 : système bonus-malus à trois niveaux

On étudie ici un modèle où l'assureur propose à ses assurés une échelle bonus-malus à trois niveaux : taux A (bonus 50%, en bas de l'échelle), taux B (tarif de base) et taux C (malus 100%, en haut de l'échelle). On considère que :

- Les nouveaux assurés font leur entrée dans le système au niveau B puis le niveau est révisé chaque année. Une année sans sinistre est récompensée par une descente d'un niveau dans l'échelle, tandis que chaque sinistre est pénalisé par une remontée d'un niveau.
- Les conducteurs se répartissent équitablement en deux catégories :
  - ➢ les bons, pour lesquels le nombre de sinistre par an obéit à la loi $\mathcal{P}(\lambda_1)$,
  - ➢ et les mauvais, pour lesquels le nombre de sinistre par an obéit à la loi $\mathcal{P}(\lambda_2)$

  où $\lambda_1 < \lambda_2$ sont deux réels. Pour les simulations, on prendra $\lambda_1 = 0{,}05$ et $\lambda_2 = 0{,}15$. Les nombres annuels de sinistres sont supposés indépendants.

Pour tout $n \in \mathbb{N}$, on note X_n le niveau occupé par l'assuré au cours de la période $(n, n+1)$.

1. On suppose que l'assuré est un bon conducteur.

a. Justifier que $(X_n)_{n\in\mathbb{N}}$ est une chaîne de Markov et en déterminer la matrice de transition Q.

b. Déterminer la loi stationnaire Π de (X_n). Y a-t-il convergence de (X_n) vers Π ?

2. Faire le même travail pour les mauvais conducteurs.

3. **a.** Calculer la probabilité, pour un conducteur au tarif C au bout d'un grand nombre d'années, d'être mauvais conducteur. Le niveau C est-il discriminant ?

b. Même question pour le niveau A.

Partie 6 : une chaîne de Markov périodique

Exercice 9 : urne d'Ehrenfest

On considère deux urnes A et B ainsi que N boules numérotées de 1 à N. Initialement, toutes les boules se trouvent dans l'urne A. À chaque instant $n \in \mathbb{N}^*$, on tire au hasard un numéro $i \in [\![1, N]\!]$ et on change d'urne la boule numérotée i. On s'intéresse à l'évolution au cours du temps du nombre de boules dans l'urne A.
On note, pour tout $n \in \mathbb{N}^*$, X_n le nombre de boules dans l'urne A à l'instant n (après déplacement d'une boule). On prendra $N = 5$ pour les simulations.

1. Expliquer pourquoi $(X_n)_{n\in\mathbb{N}}$ est une chaîne de Markov et donner sa matrice de transition Q.
2. Écrire une fonction qui trace une trajectoire de cette chaîne. On pourra utiliser la commande `grand(1,1,'uin',a,b)` qui simule la loi uniforme sur l'intervalle d'entiers $[\![a, b]\!]$ mais on s'interdira l'emploi de la commande `grand(n,'markov',Q',x0)`.
3. Vérifier que la loi invariante Π de cette chaîne est la loi binomiale $\mathcal{B}\left(N, \frac{1}{2}\right)$.
4. Illustrer par simulation le théorème ergodique, i.e. le fait que :

$$\forall k \in [\![0, N]\!], \quad \frac{1}{n} \operatorname{Card}\{j \in [\![1, n]\!] : X_j = k\} \xrightarrow[n\to\infty]{} \Pi(k) \text{ p.s..}$$

5. Y a-t-il convergence en loi de (X_n) vers Π, i.e. a-t-on :

$$\forall k \in [\![0, N]\!], \quad \mathbb{P}(X_n = k) \xrightarrow[n\to\infty]{} \Pi(k)\,?$$

   On pourra émettre une conjecture grâce à Scilab puis la démontrer rigoureusement.