Probabilités
Cours 4 - Chaînes de Markov sur un espace fini d'états

A. Ridard

A propos de ce document

- Pour naviguer dans le document, vous pouvez utiliser :
  - le menu (en haut à gauche)
  - les différents liens
- Pour signaler une erreur, vous pouvez envoyer un message à l'adresse suivante : anthony.ridard@univ-ubs.fr

Plan du cours

Dans cette section, on considère l'ensemble (des états) $E = \{1, \ldots, n\}$**.**

Définition (Matrice stochastique)

On dit que la matrice $A = \left(a_{ij}\right)_{i,j \in E}$ est stochastique si :

- elle est positive : $\forall i, j \in E,\ a_{ij} \geq 0$
- la somme sur chaque colonne vaut 1 : $\forall j \in E,\ \sum\limits_{i \in E} a_{ij} = 1$

- Le réel a_{ij} peut s'interpréter comme la probabilité de passer de l'état j à l'état i
- La matrice A peut se représenter à l'aide d'un graphe probabiliste (orienté et pondéré)

Dans la littérature, c'est plutôt la somme de chaque ligne qui vaut 1.
Mon choix [a] est moins cohérent avec la matrice d'un graphe, mais plus avec la matrice d'une application linéaire !

a. Pour suivre l'article « Comment fonctionne Google ? »

Représenter les graphes probabilistes pour les matrices stochastiques suivantes :

$$A_1 = \begin{pmatrix} 0.65 & 0.45 \\ 0.35 & 0.55 \end{pmatrix} \quad A_2 = \begin{pmatrix} 0.9 & 0.7 & 0.8 \\ 0.05 & 0 & 0 \\ 0.05 & 0.3 & 0.2 \end{pmatrix}$$

$$A_3 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \qquad A_4 = \begin{pmatrix} 0.9 & 0.7 & 0.6 \\ 0.1 & 0.3 & 0.2 \\ 0 & 0 & 0.2 \end{pmatrix}.$$

Définition (Vecteur stochastique ou loi de probabilité)

On dit qu'un vecteur $\begin{pmatrix} \mu_1 \\ \vdots \\ \mu_n \end{pmatrix}$ est stochastique si :

- il est positif : $\forall i \in E,\ \mu_i \geq 0$
- la somme de ses coordonnées vaut 1 : $\sum_{i \in E} \mu_i = 1$

Par abus, on confond le vecteur avec sa matrice colonne formée de ses coordonnées (dans la base canonique)

Dorénavant, sauf mention contraire, A désignera une matrice stochastique et μ un vecteur stochastique.

Définition (Loi de probabilité invariante)

On dit que la loi de probabilité μ est invariante pour A si $A\mu = \mu$.

Autrement dit, si μ est un vecteur propre de A associé à la valeur propre 1

Définition (Matrice ergodique)

On dit que la matrice A est ergodique s'il existe une puissance de A strictement positive [a] :

$$\exists k \geq 1,\ A^k > 0$$

a. Une matrice est dite strictement positive lorsque tous ses coefficients sont strictement positifs

- En notant $a_{ij}^{(k)}$ les coefficients de A^k, on a :

$$A^k > 0 \Leftrightarrow \forall i,j \in E,\ a_{ij}^{(k)} > 0$$

- Si A est strictement positive, alors elle est ergodique
- Si A est ergodique, alors elle est irréductible [a]

a. Matrice (positive) avec une seule classe d'équivalence pour la relation de communication :

$$i \leftrightarrow j \Leftrightarrow \exists k,k' \geq 1,\ a_{ij}^{(k)} > 0 \quad \text{et} \quad a_{ji}^{(k')} > 0$$

Les matrices A_1, A_2, A_3, A_4 sont-elles ergodiques ?

Existence et unicité de la loi de probabilité invariante

Si A est ergodique, alors il existe une unique loi de probabilité μ invariante pour A.

- De plus, $\mu > 0$
- Cette propriété reste vraie si A est seulement irréductible
- En fait, c'est une conséquence du théorème de Frobenius [a]

a. Le rayon spectral d'une matrice positive irréductible est une valeur propre simple associée à un vecteur propre strictement positif

Vérifier le deuxième point de la remarque pour la matrice A_3.

Vérifier la propriété pour les matrices A_1, A_2 (TP5).

Théorème d'ergodicité

Si A est ergodique, alors pour tout $i, j \in E$:

$$a_{ij}^{(k)} \xrightarrow[k \to +\infty]{} \mu_i$$

avec μ l'unique loi de probabilité invariante pour A.

Si A est seulement irréductible, alors on a seulement la convergence en moyenne de Césaro :

$$\frac{a_{ij}^{(1)} + a_{ij}^{(2)} + \cdots + a_{ij}^{(k)}}{k} \xrightarrow[k \to +\infty]{} \mu_i$$

Vérifier la propriété pour les matrices A_1, A_2 et la remarque pour la matrice A_3 (TP4).

Définition (Chaîne de Markov)

Une suite de v.a. $(X_k)_{k\in\mathbb{N}}$ est une chaîne de Markov si pour tout $k\in\mathbb{N}$ et tout $i,j,j_0,\dots,j_{k-1}\in E$, on a :

$$\mathrm{P}\left(X_{k+1}=i \mid X_0=j_0,\dots,X_{k-1}=j_{k-1},X_k=j\right)=\mathrm{P}\left(X_{k+1}=i \mid X_k=j\right)$$

« La prédiction du futur connaissant le présent et le passé n'est pas meilleure que celle connaissant uniquement le présent »

Définition (Chaîne de Markov homogène)

Une chaîne de Markov $(X_k)_{k\in\mathbb{N}}$ est dite homogène si pour tout $k\in\mathbb{N}$, on a :

$$\mathrm{P}(X_{k+1}=i \mid X_k=j)=\mathrm{P}(X_1=i \mid X_0=j)$$

La matrice $A=\left(\mathrm{P}(X_1=i \mid X_0=j)\right)_{i,j\in E}$ est sa matrice de transition.

La loi de X_0 est sa loi initiale.

- On se limitera aux chaînes de Markov homogènes
- A est une matrice stochastique
- Une telle chaîne représente l'évolution d'un système dynamique à n états possibles, au cours du temps (discret), étant donnée une loi initiale et en supposant qu'à chaque instant le changement est régi par la matrice de transition A

Loi conjointe

La loi conjointe de $(X_0, X_1, \dots, X_k)$ est déterminée par la matrice de transition $A = \left(a_{ij}\right)_{i,j \in E}$ et la loi initiale :

$$\forall i_0, i_1, \dots, i_k \in E,\ \mathrm{P}(X_0 = i_0, X_1 = i_1, \dots, X_k = i_k) = \mathrm{P}(X_0 = i_0)\, a_{i_1 i_0}\, a_{i_2 i_1} \dots a_{i_k i_{k-1}}$$

Évidemment, on a : $\mathrm{P}(X_0 = i_0, X_1 = i_1, \dots, X_k = i_k) = \mathrm{P}\Big((X_0 = i_0) \cap (X_1 = i_1) \cap \dots \cap (X_k = i_k)\Big)$

Démontrer la propriété.

Lois conditionnelles

La matrice de transition en k pas $\left(\mathrm{P}\left(X_{m+k} = i \mid X_m = j\right)\right)_{i,j \in E}$ est égale à A^k :

$$\forall i,j \in E,\ \mathrm{P}\left(X_{m+k} = i \mid X_m = j\right) = a_{ij}^{(k)}$$

Démontrer la propriété.

Lois marginales

La loi de X_{m+k} est donnée par :

$$\forall i \in E,\ \mathrm{P}(X_{m+k} = i) = \sum_{j \in E} a_{ij}^{(k)} \mathrm{P}(X_m = j)$$

En particulier, la loi de X_k est donnée par :

$$\forall i \in E,\ \mathrm{P}(X_k = i) = \sum_{j \in E} a_{ij}^{(k)} \mathrm{P}(X_0 = j)$$

En notant $\mu^{(l)}$ la loi de X_l, on peut reformuler de la manière suivante :

$$\mu^{(m+k)} = A^k \mu^{(m)} \quad \text{et} \quad \mu^{(k)} = A^k \mu^{(0)}$$

ce qui n'est pas sans rappeler les formules sur les suites géométriques

Démontrer la propriété.

Théorème d'ergodicité

Si A est ergodique, alors la loi de X_k tend vers l'unique loi de probabilité μ invariante pour A :

$$\mu^{(k)} \xrightarrow[k \to +\infty]{\mathscr{L}} \mu$$

Le comportement asymptotique ne dépend pas de la loi initiale !

Doudou le hamster 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 pour Doudou ; 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.

En supposant que Doudou soit en train de dormir (est-ce important ?), prévoir son comportement dans deux heures (TP5).