

Séance à distance 3

Chaînes de Markov

1 Introduction

Dans cette séance, on étudie des *chaînes de Markov*. Une chaîne de Markov est la donnée d'un ensemble fini d'états (numéroté de 1 à N), et, pour chacun des états $1 \leq i \leq N$, une loi de *probabilités de transition*: une famille de nombres $0 \leq p_{i \rightarrow j} \leq 1$, indexée par $1 \leq j \leq N$, qui décrit une loi de probabilités:

$$\text{pour tout } 1 \leq i \leq n, \quad p_{i \rightarrow 1} + p_{i \rightarrow 2} + \dots + p_{i \rightarrow N} = \sum_{j=1}^N p_{i \rightarrow j} = 1.$$

Un exemple de chaîne de Markov avec deux états est :

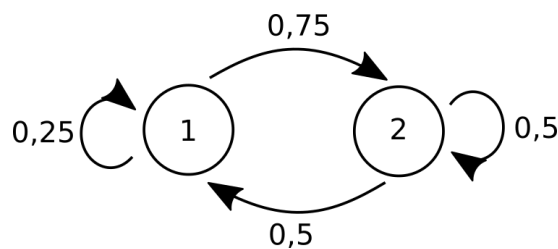
$$\begin{array}{ll} p_{1 \rightarrow 1} = 0,25 & p_{1 \rightarrow 2} = 0,75 \\ p_{2 \rightarrow 1} = 0,5 & p_{2 \rightarrow 2} = 0,5. \end{array}$$

Une chaîne de Markov est donc un tableau de nombres, tel que chaque nombre est positif et la somme sur chaque ligne vaut 1. On écrira donc plus synthétiquement la chaîne de Markov précédente comme

$$\begin{pmatrix} 0,25 & 0,75 \\ 0,5 & 0,5 \end{pmatrix}$$

Notons que la somme des nombres sur les colonnes d'une chaîne de Markov n'a aucune raison de valoir 1, elle peut être plus petite ou plus grande.

On peut aussi représenter cette chaîne de Markov par le dessin suivant:



On associe, à une chaîne de Markov, un *processus* aléatoire, c'est-à-dire, une suite définie par récurrence, mais où chaque étape de la récurrence, au lieu d'être déterministe, comporte un facteur aléatoire. Ce processus $(X_n)_{n \in \mathbb{N}}$ se déroule comme suit:

- On part, au temps $n = 0$, d'un état $1 \leq i_0 \leq n$ de la chaîne de Markov. On pose $X_0 = i_0$.
- Au temps n , on est dans un certain état $1 \leq X_n \leq n$. Alors, au temps $n + 1$, on passe à l'état j (c'est-à-dire $X_{n+1} = j$) avec probabilité $p_{X_n \rightarrow j}$. Dans le langage des probabilités conditionnelles, on a

$$\text{pour tous } 1 \leq i, j \leq N, \quad \mathbb{P}[X_{n+1} = j | X_n = i] = p_{i \rightarrow j}.$$

Pour déterminer la probabilité d'une certaine trajectoire, on raisonne donc selon un arbre de probabilités. Dans l'exemple présenté plus haut, posons $X_0 = 1$. Alors

$$\mathbb{P}[X_1 = 1] = 0.25 \quad \mathbb{P}[X_1 = 2] = 0.75.$$

On peut continuer :

$$\begin{aligned} \mathbb{P}[X_1 = 1, X_2 = 1] &= 0.0625 & \mathbb{P}[X_1 = 1, X_2 = 2] &= 0.1875 \\ \mathbb{P}[X_1 = 2, X_2 = 1] &= 0.375 & \mathbb{P}[X_1 = 2, X_2 = 2] &= 0.375. \end{aligned}$$

Question 1. Considérons une chaîne de Markov générale. Au temps $n \in \mathbb{N}$, on note, pour tout $1 \leq i \leq N$,

$$p_i = \mathbb{P}[X_n = i]$$

la probabilité que, en ce temps, on soit au site i .

Montrer alors que, pour tout $1 \leq j \leq N$,

$$\mathbb{P}[X_{n+1} = j] = \sum_{i=1}^N p_i p_{i \rightarrow j}.$$

Question 2. Dans l'exemple ci-dessus, pour $n \leq 4$, calculer $\mathbb{P}[X_n = 1]$ et $\mathbb{P}[X_n = 2]$.

Question 3. Reprendre le calcul précédent avec $X_0 = 2$. Que constatez-vous ?

Ce TD est assorti d'un notebook Jupyter dans lequel la méthode de calcul précédente est implémentée et généralisée. Étant donnée une chaîne de Markov T (avec N états), un entier n et un état $1 \leq X_0 \leq N$, la fonction `evolution_Markov(T,n,X0)` renvoie la liste des probabilités que $X_n = j$, pour j allant de 1 à N .

T est un tableau de nombres, et on le représente en python de la manière suivante. Rappelons qu'une liste s'écrit entre crochets, avec des virgules: $[1, 2, 4, 8]$. Un tableau de nombres est une liste de listes (la liste des lignes); par exemple, dans cette convention, le tableau

$$\begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$$

s'écrit `[[1,0],[0,2]]`. Attention, comme d'habitude, les points sont les séparateurs décimaux, et les virgules servent à séparer des nombres entre eux.

Question 4. Créer une variable T qui représente la chaîne de Markov donnée en exemple (rappel: on écrit `T=...`) puis appeler la fonction `evolution_Markov` pour retrouver les résultats ci-dessus. Augmentez la valeur de n pour confirmer (ou infirmer) votre impression sur le comportement de cette chaîne de Markov.

2 Convergence des chaînes à deux états

Question 5. Choisir d'autres valeurs pour les quatre coefficients $p_{i \rightarrow j}$ dans le notebook. Observez-vous un comportement différent ? Que se passe-t-il si un coefficient est nul ? Deux coefficients ?

Question 6. Montrer que, si $p_{2 \rightarrow 1}$ et $p_{1 \rightarrow 2}$ ne sont pas tous les deux nuls, et si

$$\mathbb{P}[X_n = 1] = \frac{p_{2 \rightarrow 1}}{p_{1 \rightarrow 2} + p_{2 \rightarrow 1}} \quad \mathbb{P}[X_n = 2] = \frac{p_{1 \rightarrow 2}}{p_{1 \rightarrow 2} + p_{2 \rightarrow 1}},$$

alors il en va de même pour X_{n+1} . Autrement dit, la loi de probabilité ci-dessus est un point d'équilibre de la chaîne de Markov.

Comparer avec le résultat des questions précédentes. Cette loi de probabilités semble-t-elle toujours être le seul point d'équilibre ?

On se propose de montrer que, si les quatre coefficients sont tous non nuls, alors la loi de probabilité ci-dessus est un point d'équilibre stable, c'est-à-dire, quelle que soit la valeur de X_0 , les probabilités $\mathbb{P}[X_n = 1], \mathbb{P}[X_n = 2]$ convergent respectivement vers $\frac{p_{2 \rightarrow 1}}{p_{1 \rightarrow 2} + p_{2 \rightarrow 1}}$ et $\frac{p_{1 \rightarrow 2}}{p_{1 \rightarrow 2} + p_{2 \rightarrow 1}}$.

Étant donné X_0 , on pose $q_n = \mathbb{P}[X_n = 1]$. On a ainsi $1 - q_n = \mathbb{P}[X_n = 2]$.

Question 7. Montrer que

$$q_{n+1} = (1 - p_{1 \rightarrow 2})q_n + p_{2 \rightarrow 1}(1 - q_n),$$

et en déduire que

$$q_{n+1} - \frac{p_{2 \rightarrow 1}}{p_{1 \rightarrow 2} + p_{2 \rightarrow 1}} = (1 - p_{1 \rightarrow 2} - p_{2 \rightarrow 1}) \left(q_n - \frac{p_{2 \rightarrow 1}}{p_{1 \rightarrow 2} + p_{2 \rightarrow 1}} \right).$$

Question 8. Montrer que $-1 < 1 - p_{1 \rightarrow 2} - p_{2 \rightarrow 1} < 1$, et en déduire que, quand $n \rightarrow +\infty$, on a

$$q_n \rightarrow \frac{p_{2 \rightarrow 1}}{p_{1 \rightarrow 2} + p_{2 \rightarrow 1}}.$$

En déduire également que $\mathbb{P}[X_n = 2] \rightarrow \frac{p_{1 \rightarrow 2}}{p_{1 \rightarrow 2} + p_{2 \rightarrow 1}}$.

Question 9. Justifier l'affirmation suivante :

"Si on connaît les k premières décimales de $\mathbb{P}[X_n = 1]$, mais que n est au moins aussi grand qu'une certaine constante fois k , alors on ne peut pas savoir si $X_0 = 1$ ou si $X_0 = 2$."

Si $p_{2 \rightarrow 1}$ et $p_{1 \rightarrow 2}$ ne sont pas tous les deux nuls, alors on a démontré la propriété suivante :

On dit qu'une chaîne de Markov possède la **propriété d'oubli** lorsqu'au bout d'un certain temps n , la valeur de $\mathbb{P}[X_n = i]$ ne dépend presque plus de X_0 .

Plus précisément, une chaîne de Markov à N possède la propriété d'oubli lorsqu'il existe $\mu_1, \dots, \mu_N$ (fixes) tels que, pour tout X_0 , et pour tout $1 \leq i \leq N$, on a $\mathbb{P}[X_n = i] \rightarrow \mu_i$ lorsque $n \rightarrow +\infty$.

Attention à l'ordre des quantificateurs : les valeurs de $\mu_1, \dots, \mu_N$ doit être la même pour toutes les valeurs de X_0 .

Question 10. Montrer que les chaînes

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

ne possèdent pas la propriété d'oubli.

3 Chaînes de Markov irréductibles et apériodiques

Dans la section précédente, on a vu que le comportement d'une chaîne de Markov à deux états pouvait être singulier si trop de probabilités de transition sont nulles. Pour qu'une chaîne de Markov à un nombre quelconque d'état possède la propriété d'oubli, il n'est pas nécessaire que *toutes* les probabilités de transition soient non-nulles, mais il y a un certain nombre de cas singuliers à décrire.

Question 11. Décrire le comportement en temps long des chaînes de Markov suivantes:

$$\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix} \quad \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \quad \begin{pmatrix} 0,25 & 0,75 & 0 \\ 0,5 & 0,5 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \begin{pmatrix} 0 & 0,5 & 0,5 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$\begin{pmatrix} 0 & 0,5 & 0 & 0,5 \\ 0,5 & 0 & 0,5 & 0 \\ 0 & 0,5 & 0 & 0,5 \\ 0,5 & 0 & 0,5 & 0 \end{pmatrix} \quad \begin{pmatrix} 0 & 0,5 & 0 & 0 & 0,5 \\ 0,5 & 0 & 0,5 & 0 & 0 \\ 0 & 0,5 & 0 & 0,5 & 0 \\ 0 & 0 & 0,5 & 0 & 0,5 \\ 0,5 & 0 & 0 & 0,5 & 0 \end{pmatrix}.$$

Il est suggéré de faire des dessins et d'utiliser le notebook!

Parmi ces chaînes, laquelle ou lesquelles possède(nt) la propriété d'oubli ?

Pour éliminer les phénomènes qui apparaissent dans la question précédente, on pose les définitions suivantes.

Probabilité itérée Pour tout $n \in \mathbb{N}$ et tous sites i, j d'une chaîne de Markov, on écrit $p_{i \rightarrow j}^{(n)}$ la probabilité que $X_n = j$ si on choisit $X_0 = i$.

Chemin Un *chemin* de longueur n entre deux états i et j d'une chaîne de Markov est une famille de n états $k_0, k_1, \dots, k_n$, avec $k_0 = i$ et $k_n = j$, telle que pour tout m , la probabilité $p_{k_m \rightarrow k_{m+1}}$ de passer de k_m à k_{m+1} soit non-nulle.

Puits Un *puits* d'une chaîne de Markov est un sous-ensemble $\mathcal{J}$ d'états tels que, pour tous i et j dans $\mathcal{J}$, il existe un chemin allant de i à j et un chemin de j à i .

Irréductibilité Une chaîne de Markov est *irréductible* si elle a un seul puits.

Périodes Une *période* d'un site i d'une chaîne de Markov est la longueur d'un chemin qui va de i à i .

Apériodicité Une chaîne de Markov est dite *apériodique* lorsque pour tout site i , le pgcd de toutes les périodes de i est 1.

Question 12. Montrer que, s'il existe un chemin de longueur n entre i et j , alors $p_{i \rightarrow j}^{(n)} > 0$.

Question 13. Parmi les chaînes précédentes, lesquelles sont irréductibles ? Lesquelles sont apériodiques ?

Le résultat suivant est fondamental dans l'étude des chaînes de Markov, et on se propose d'en démontrer une partie.

Théorème 1. Une chaîne de Markov irréductible possède la propriété d'oubli si et seulement si elle est apériodique.

On va avoir besoin du résultat suivant (admis).

Théorème 2. Soient $a_1, \dots, a_n$ des entiers premiers entre eux (pas forcément deux à deux, mais le plus grand diviseur de tous les $a_1, \dots, a_n$ est 1). Alors il existe $C > 0$ tel que, pour tout $m \geq C$, on peut écrire $m = a_1 j_1 + a_2 j_2 + \dots + a_n j_n$ avec $j_1, j_2, \dots, j_n \in \mathbb{N}$.

Question 14. Montrer que, si une chaîne de Markov est apériodique, alors pour tout site i , l'ensemble $\mathcal{P}_i$ contient tous les entiers à partir d'un certain rang.

Question 15. Montrer que, pour toute chaîne de Markov apériodique et irréductible, il existe n tel que tous les coefficients $p_{i \rightarrow j}^{(n)}$ soient strictement positifs.

Pour simplifier, on suppose à présent, et jusqu'à la fin du devoir, que tous les coefficients $p_{i \rightarrow j}$ dans la matrice sont strictement positifs. Cela revient à "sauter" des étapes de la chaîne de Markov à chaque itération.

Question 16. Le but de cette question est de démontrer qu'il n'existe pas plus d'une famille de coefficients $(p_1, \dots, p_N)$ telle que, si à un instant n on a, pour tout i , $\mathbb{P}[X_n = i] = p_i$, alors on a, pour tout i , $\mathbb{P}[X_{n+1} = i] = p_i$. Autrement dit, il n'existe pas plus d'un point d'équilibre pour l'action de la chaîne de Markov sur les lois de probabilités.

1. Montrer que $(p_1, \dots, p_N)$ est une suite de coefficients de probabilité stable si, et seulement si, toutes les conditions algébriques suivantes sont vérifiées:

a) Pour tout $1 \leq i \leq N$, on a $0 \leq p_i \leq 1$.

b) On a $\sum_{i=1}^N p_i = 1$.

c) Pour tout $1 \leq j \leq N$, on a

$$p_j = \sum_{i=1}^N p_i p_{i \rightarrow j}.$$

2. Montrer qu'au moins l'un des coefficients p_i est strictement positif. En déduire qu'en réalité, tous les coefficients p_i sont strictement positifs.
3. On suppose qu'on a deux suites de coefficients $(p_1, \dots, p_N)$ et $(q_1, \dots, q_N)$ vérifiant les conditions ci-dessus. Montrer que, pour tout $t \in \mathbb{R}$, la suite de coefficients

$$(tp_1 + (1-t)q_1, tp_2 + (1-t)q_2, \dots, tp_N + (1-t)q_N)$$

vérifie les conditions b) et c) ci-dessus.

4. On suppose que les suites $(p_1, \dots, p_N)$ et $(q_1, \dots, q_N)$ ne sont pas identiques. Montrer alors qu'il existe $t \in \mathbb{R}$ tel que, pour tout $1 \leq i \leq N$, on a $0 \leq tp_i + (1-t)q_i \leq 1$, et qu'il existe i tel qu'on ait $tp_i + (1-t)q_i = 0$ ou 1 .
5. Conclure par l'absurde.

On a démontré l'unicité de la limite, pour la propriété d'oubli; pour démontrer son existence, il faut utiliser des outils d'algèbre linéaire.