Probabilités – 3 IF

TD 8 : Chaînes de Markov

Rappel de cours

Une **chaîne de Markov** à k états est une suite de variables aléatoires $(X_n)_{n\in\mathbb{N}}$ à valeurs dans $E = [1, \ldots k]$ (ensemble des états), telle que :

$$\forall n, \forall i_0, \ldots, i_{n+1}, \mathbb{P}\left[X_{n+1} = i_{n+1} | \bigwedge_{k=0}^{n} X_k = i_k\right] = \mathbb{P}\left[X_{n+1} = i_{n+1} | X_n = i_n\right]$$

On dit que le système est **sans mémoire**.

Si de plus, on a

$$\forall n, \mathbb{P}[X_{n+1} = i_{n+1} | X_n = i_n] = \mathbb{P}[X_1 = i_{n+1} | X_0 = i_n]$$

on dit que la chaîne de Markov est **homogène**.

Dans ce cas, la probabilité de passer de l'état i à l'état j ne dépend pas de n. On la note $p_{i,j}$ et on appelle **matrice de transition** la matrice

$$G = (p_{i,j})_{i,j\in[1,k]^2}$$

$\pi_n \cdot G^k = \pi_{n+k}$

i est récurrent si $\sum_{k>0} p_{i,i}^k = \infty$ sinon, i est transitoire

Le graphe associé à cette matrice est appelé **diagramme sagital** de la chaîne de Markov.

THM

Si G est la matrice de transition d'une chaîne de Markov ergodique, il existe une unique loi de probabilité $\pi^\star$ sur $[1, k]$ telle que

$$\underline{\pi^\star G = \pi^\star}$$

$\underline{\sum_{i=1}^{k} \pi_i^* = 1}$ $(*)$

On l'appelle la **loi invariante** de G et elle vérifie:

$$\forall \pi, \pi \cdot G^n \underset{n\to\infty}{\longrightarrow} \pi^\star$$

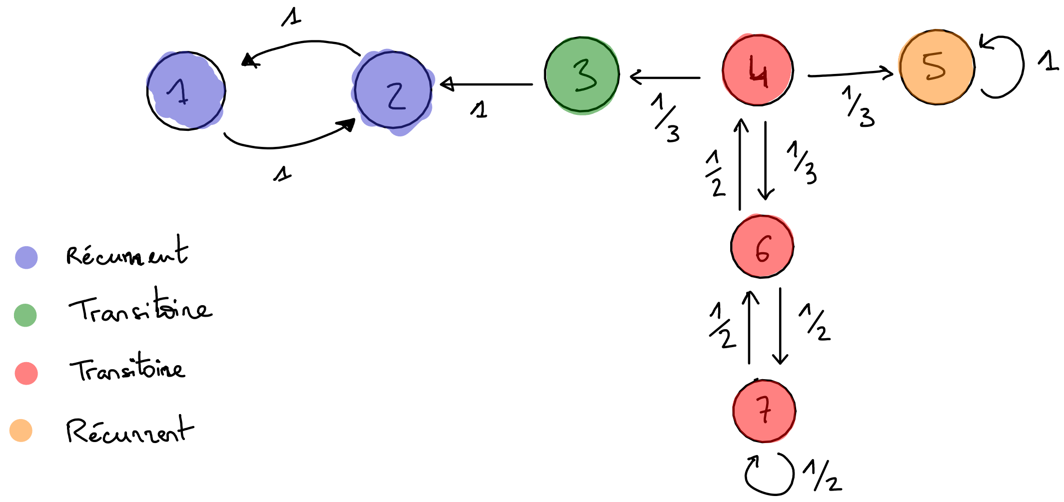
Exercice 1

Soit X_n une chaîne de Markov homogène de matrice de transition

$$G = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{3} & 0 & \frac{1}{3} & \frac{1}{3} & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{2} & 0 & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix}$$

Tracer le diagramme sagittal correspondant, et partitionner l'espace d'états en classes transitoires et récurrentes.

relation d'équivalence sur les états : accessibilité entre états.



Exercice 2. (DS 2015)
Lors d'un appel à un moteur de recherche, celui-ci doit trier les différents sites référencés. Le principe PageRank utilisé par Google utilise la modélisation par les chaînes de Markov. Nous évoquons ici cette modélisation.
Dans un premier temps à l'aide d'une méthode de fouilles de données, le moteur de recherche récupère un nombre fini de sites internet correspondant à la recherche effectuée. La procédure PageRank consiste alors à attribuer à ces sites un rang qui correspond à leur importance de référencement.
Supposons qu'il y ait N sites notés $\{1, \ldots, N\}$ sélectionnés à la première étape. Soit X_k le k-ème site internet visité par un utilisateur fictif. Le site contient des liens vers les autres sites. Soit K_i le nombre total de liens vers d'autres sites contenus sur la page i. Nous noterons

$$p_{i,j} = \begin{cases} \frac{1}{K_i} \text{ si le site } i \text{ contient un lien vers le site } j \\ 0 \ \text{ sinon} \end{cases}$$

On ne prend pas en compte l'auto-référencement ici.
On suppose que l'utilisateur clique de façon aléatoire sur les liens, et qu'ainsi $\mathbb{P}[X_{k+1}] = j | X_k = i] = p_{i,j}$ pour tout $k \in \mathbb{N}$, que le site j ait déjà été visité ou non. Ce problème peut donc être modélisé par une chaîne de Markov homogène. Soit G la matrice de transition.
Dans la suite nous considérons :

$$G = \begin{pmatrix} 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 \\ 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 1 & 0 & 0 & 0 \\ 1/4 & \frac{1}{4} & \frac{1}{4} & 0 & \frac{1}{4} \\ 1/3 & 0 & \frac{1}{3} & \frac{1}{3} & 0 \end{pmatrix}$$

La chaîne associée est ergodique.

1. Compléter le premier vecteur colonne de la matrice de transition.

2. Que vaut N ici ? Que vaut $p_{2,3}$?

$N = 5$ (nbr de site).

$p_{2,3} = \frac{1}{2}$

3. Quel site internet a le plus de liens entrants directs ?

Le site n°3 (4 liens entrants). Le nombre de liens entrants sur le site j est le nombre de coefficients non nuls sur la j^e colonne.

4. Parmi les lois suivantes, laquelle est une loi stable pour cette chaîne de Markov ?

$\pi(1) = (\,0.2 \quad 0.2 \quad 0.2 \quad 0.2 \quad 0.2\,)$ — NON : $\pi(1) \cdot G \neq \pi(1)$

× $\pi(2) = (\,0.09 \quad 0.34 \quad 0.29 \quad 0.09 \quad 0.19\,)$ — $\pi(2) \cdot G = (0.09 \quad 0.34 \quad 0.29 \quad 0.09 \quad 0.19)$ OK

$\pi(3) = (\,0.15 \quad 0.64 \quad 0.27 \quad 0.54 \quad 0.19\,)$ — NON : somme > 1

$\pi(4) = (\,-0.09 \quad 0.48 \quad 0.29 \quad 0.54 \quad -0.14\,)$ — NON : probabilités négatives

× Aucune.

Indication : on pourra procéder par élimination.

5. Quel site internet doit être proposé en premier par le moteur de recherche selon vous ? Commentez.

Le site 2, car à long terme, pour une navigation aléatoire, c'est celui sur lequel on a le plus de chance de rester. À noter, ce n'est pas celui avec le plus grand nombre de liens entrants.

Exercice 3 : Modèle élémentaire de gestion de stock.

La demande de consommation d'un bien entre les instants n et $n+1$ est représentée par une variable aléatoire positive Y_{n+1} . Le niveau du stock du bien est égal à X_n à l'instant n. La gestion du stock se fait de la façon suivante :
— si $X_n - Y_{n+1} \geq m$, alors le stock n'est pas réapprovisionné, et $X_{n+1} = X_n - Y_{n+1}$.
— si $X_n - Y_{n+1} < m$, alors le stock est réapprovisionné et porté au niveau M : $X_{n+1} = M$.
On suppose que X_0 est de loi quelconque sur $\{m, \ldots, M\}$, et que les $(Y_n)_{n \in \mathbb{N}}$ et X_0 sont indépendantes, les $(Y_n)_{n \in \mathbb{N}}$ sont de même loi.
Pour l'application numérique, on pose que les $(Y_n)_{n \in \mathbb{N}}$ suivent une loi uniforme sur $\{0, 1, 2\}$, et que $M = 3$ et $m = 1$.

1. Montrer que X_n est une chaîne de Markov homogène et calculer la matrice de transition G.

→ Ensemble des états : $E = \{m, \dots, M\}$.

$Y_{n+1} > x_n - m$

$x_0 \cdots x_1 \cdots x_n \to x_{n+1}$ M

⊛ Soient $x_0, \dots, x_{n+1} \in E$

$$\underbrace{\mathbb{P}\left[X_{n+1} = x_{n+1} \mid \bigwedge_{i=0}^{n} X_i = x_i\right] = \begin{cases} \mathbb{P}\left[Y_{n+1} = x_n - x_{n+1}\right] & \text{si } m \leq x_{n+1} \leq x_n \\ \mathbb{P}\left[Y_{n+1} > x_n - m\right] & \text{si } x_{n+1} = M \\ 0 & \text{sinon} \end{cases}}_{\mathbb{P}\left[X_{n+1} = x_{n+1} \mid X_n = x_n\right]}$$

$$G = \begin{array}{c|c} & \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \end{array} & \begin{pmatrix} 1/3 & 0 & 2/3 \\ 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \end{pmatrix} \end{array}$$

Donc $(X_n)_n$ est une chaîne de Markov.

HOMOGENÉITÉ

On veut mq $\forall n$, $\underbrace{\mathbb{P}\left[X_{n+1} = i \mid X_n = j\right]}_{\begin{cases} \mathbb{P}[Y_{n+1} = j-i] & \text{si} \dots \\ \mathbb{P}[Y_{n+1} > j-m] & \text{si} \dots \\ 0 & \text{sinon} \end{cases}} = \underbrace{\mathbb{P}\left[X_1 = i \mid X_0 = j\right]}_{\begin{cases} \mathbb{P}[Y_1 = j-i] & \text{si} \dots \\ \mathbb{P}[Y_1 > j-m] & \text{si} \dots \\ 0 & \text{sinon} \end{cases}}$

Y_{n+1} a la même distribution de proba que Y_1

2. Trouver une (la) probabilité invariante ν. Quelle est la probabilité en régime stationnaire d'être en rupture de stock ?

On veut trouver $\lim_{k\to\infty} G^k$ pour pouvoir lire les coefficients de ν sur les colonnes de la matrice limite.

$$G = \frac{1}{3}\begin{pmatrix} 1 & 0 & 2 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$$

$$G^2 = \frac{1}{9}\begin{pmatrix} 3 & 2 & 4 \\ 3 & 2 & 4 \\ 3 & 2 & 4 \end{pmatrix}$$

$$G^3 = \frac{1}{27}\begin{pmatrix} 9 & 6 & 12 \\ 9 & 6 & 12 \\ 9 & 6 & 12 \end{pmatrix} = \frac{1}{9}\begin{pmatrix} 3 & 2 & 4 \\ 3 & 2 & 4 \\ 3 & 2 & 4 \end{pmatrix} = G^2$$

D'où $\forall k \geq 2$, $G^k = G^2$ et $\lim_{k\to\infty} G^k = G^2$

Ainsi, $\forall \pi$, $\lim_{k\to\infty} \pi \cdot G^k = \frac{1}{9}(3 \quad 2 \quad 4) = \nu$

En régime stationnaire, avec les valeurs proposées, on est en rupture de stock si et seulement si le stock est de 1 et la demande de 2. La probabilité de cette situation est donnée par :

$$\mathbb{P}[X = 1 \wedge Y = 2] = \mathbb{P}[X = 1] \times \mathbb{P}[Y = 2] \quad \text{(indépendance)}$$

$$= \nu(1) \times \frac{1}{3}$$

$$= \frac{1}{9}$$