Exercices 2 : Chaînes de Markov

Une chaîne de Markov (à temps discret) est un processus stochastique $(X_i)_{i\in\mathbb{N}}$ sur un alphabet $\mathcal{A}$ régi par la propriété de Markov. C'est à dire que la loi de probabilité en $n+1$ dépend uniquement de l'état en n (et pas de n). Formellement, cette chaîne de Markov sera décrite par un vecteur (horizontal) de distribution initiale $p \in [0,1]^{\mathcal{A}}$ pour $P(X_0 = i) = p_i$ et des probabilités de transition $A_{i,j} := P(X_{n+1} = j \mid X_n = i)$ pour tout $n \in \mathbb{N}$ et $i, j \in \mathcal{A}$.

1. Montrer que pour tout $n \in \mathbb{N}$,
$$P(X_n = i) = (pA^n)_i \,.$$
2. Observer que A est une matrice stochastique, c'est-à-dire que toutes ses entrées sont positives et la somme de chaque ligne est égale à 1.

Le théorème suivant est un résultat fondamental dans la théorie des chaînes de Markov que l'on va admettre pour la suite.

Théorème (Perron-Frobenius)

Soit une matrice stochastique A telle que toutes les coordonnées de A^n sont strictement positives pour un $n \in \mathbb{N}$, alors

(i) Il existe un unique vecteur propre à gauche π pour la valeur propre 1 ($\pi A = \pi$) et celui-ci a toutes ses composantes strictement positives.

(ii) De plus, 1 est la valeur propre dominante : toutes les autres valeurs propres λ de A satisfont $|\lambda| < 1$.

3. Si A est la matrice d'adjacence d'un graphe, à quoi correspond les coordonnées non nulles de A^n ?
4. Comment interpréter la condition sur les coordonnées strictement positives du théorème ?

Le vecteur $\pi = (\pi_1, \pi_2, \ldots, \pi_d)$ défini donc une loi de probabilité que l'on appelle distribution stationnaire. En effet, d'après la question précédente, si l'on prend π comme distribution initiale, on a pour tout n, $P(X_n = i) = \pi_i$.

5. On appelle rayon spectral d'une matrice la limite $\rho(A) = \lim_{n\to\infty} \|A^n\|^{\frac{1}{n}}$ avec la norme sur les matrices induite par une norme donnée (par exemple L^2) sur les vecteurs par
$$\|A\| = \max_{v\in\mathbb{R}^{\mathcal{A}}} \frac{\|Av\|}{\|v\|}.$$
   (i) Justifier que pour le produit de deux matrices $\|AB\| \leq \|A\|\|B\|$.

   (ii) Montrer que pour une matrice A conjuguées à une matrice diagonale à coefficients $(\lambda_i)_{i\in\mathcal{A}}$,
$$\|A^n\|^{\frac{1}{n}} \to \max_{i\in\mathcal{A}} |\lambda_i|.$$

(iii) Rappeler le résultat de décomposition de Dunford (diagonale–nilpotente) des matrices et en déduire que $\rho(A)$ est égal à la valeur absolue maximale des valeurs propres de A.

6. Rappeler comment on déduit du théorème de Bezout le fait suivant :
   Si P_1 et P_2 sont deux polynômes premiers entre eux dans $\mathbb{R}[X]$, et $P_1 \cdot P_2$ annule A, alors $\ker P_1(A) \oplus \ker P_2(A)$ de plus les projections sur ces espaces sont des polynômes de la matrice A que l'on note $Q_1(A), Q_2(A)$.

   En déduire qu'il existe H un hyperplan de $\mathbb{R}^{\mathcal{A}}$ stable par A et supplémentaire de π. C'est à dire que $H \cdot A \subset H$ et $\mathbb{R}^{\mathcal{A}} = \mathbb{R}\pi \oplus H$.
   Que peut-on dire du rayon spectral de la matrice A restreinte à H ?
7. Montrer $Q_1(1) = 1$ et $Q_2(1) = 0$. En déduire que pour tout vecteur de probabilité $p \in \mathbb{R}^d$, on a une unique décomposition $p = \pi + v$ avec $v \in H$.
8. Montrer que $\|pA^n - \pi\| = o(r^n)$ pour un $r \in]0, 1[$.

Donc pour un n assez grand, la distribution de probabilité est exponentiellement proche de π. Dans la suite on suppose donc que $p = \pi$ et on cherche à calculer l'entropie conjointe du processus $X_0, X_1, \ldots, X_n$.

9. Calculer $H(X_1|X_0)$.
10. Montrer que $H(X_{n+1}|X_n, \ldots, X_0) = H(X_1|X_0)$ (on commencera avantageusement par le cas n=1).
11. En déduire une formule pour $H(X_0, \ldots, X_n)$.

Exercice (Quelques caculs explicites)

X \ Y	0	1
0	1/3	1/3
1	0	1/3

Calculer :

1. $H(X), H(Y)$.
2. $H(X|Y), H(Y|X)$.
3. $H(X, Y)$.
4. $H(Y) - H(Y|X)$.
5. $I(X; Y)$.
6. Vérifier que $H(X|Y) \leq H(X)$.