CAPES externe de mathématiques

Session 2017

Première composition - Problème n°2

Correction proposée par G. Dupont.

Ce corrigé n'a rien d'officiel. Il a été rédigé dans le seul but de proposer une aide aux étudiants travaillant sur les épreuves de CAPES. Il peut contenir des erreurs et est certainement améliorable en plusieurs points. Vous pouvez faire part de vos remarques sur le site `http://maths-concours.fr`.

Introduction

Thème du problème. Le problème n°2 de la deuxième composition du CAPES de mathématiques 2017 (commune aux options mathématiques et informatique) traite de marches aléatoires sur des graphes orientés, thème qui était attendu aux épreuves écrites puisque cette notion est entrée dans les programmes de Terminale en 2012 et n'était toujours pas tombée. Il s'agissait implicitement d'étudier quelques propriétés des chaînes de Markov simples homogènes, mais aucune connaissance préalable n'était requise à ce sujet.

Description de l'épreuve et mise en perspective. Le problème se décompose en quatre parties.

La partie **A** concerne, après quelques résultats préliminaires, l'étude de marches aléatoires isotropes sur un tétraèdre puis sur une pyramide tronquée. Dans le premier cas on observe le classique phénomène de convergence en loi des chaînes de Markov (ergodiques) vers la loi invariante. Dans le second cas, on observe que cette convergence en loi n'a pas lieu (la chaîne est ici périodique).

La partie **B** consiste en une étude très succincte des propriétés des matrices stochastiques.

Les parties **C** et **D** concernent l'algorithme *PageRank* de Brin-Page implémenté dans le moteur de recherche Google afin de déterminer la pertinence d'une page web. Plus spécifiquement, la partie **C** propose l'étude d'un modèle simplifié pour s'approprier la philosophie de l'algorithme. La partie **D** est quant à elle dédiée à l'étude d'un modèle plus élaboré, dans lequel nous observons à nouveau le phénomène de convergence en loi d'une chaîne de Markov vers sa loi invariante, laquelle fournit une solution effective au problème de départ.

Pré-requis. Les objets impliqués dans ce problème sont :

— Calcul matriciel ;

— Probabilités conditionnelles ;

— Suites réelles.

Difficulté. Le problème ne contient aucune difficulté technique, d'autant que le candidat est guidé de très près dans tout le problème - les réponses y sont en général particulièrement courtes et directes. Très peu de prérequis sont nécessaires pour aborder ce problème, à l'exception de la question **A.II.5** qui nécessite de connaître ses théorèmes sur les suites récurrentes linéaires d'ordre 2. Les principales difficultés auxquelles sont confrontés les candidats sont l'organisation et la gestion des nombreuses informations fournies par l'énoncé, et l'originalité de ce sujet relativement à ce qui se pratiquait lors des précédentes années.

Enoncé

Notations

$\mathbb{N}$ désigne l'ensemble des entiers naturels et $\mathbb{R}$ l'ensemble des nombres réels.
Pour m et n deux entiers naturels, $[\![m,n]\!]$ désigne l'ensemble des entiers k tels que $m \leq k \leq n$.
Pour n entier naturel non nul, $\mathcal{M}_n(\mathbb{R})$ désigne l'ensemble des matrices à n lignes et n colonnes, à coefficients réels.
On note I_n la matrice identité d'ordre n.
Le module d'un nombre complexe z est noté $|z|$.

Notations propres à ce corrigé

$(\Omega, \mathcal{A}, \mathbb{P})$ désigne un espace probabilisé sur lequel toutes les variables aléatoires considérées sont définies.
Si $A \in \mathcal{M}_n(\mathbb{R})$ est une matrice et $i, j \in [\![1,n]\!]$, on note $[A]_{i,j}$ le coefficient de A à la ligne i et à la colonne j.
Si X est un vecteur (ligne ou colonne), on note $[X]_i$ sa i-ème composante.

Définitions

Soit $(X^{(k)})_{k\geq 0}$ une suite de vecteurs de $\mathbb{R}^n$ et $X = (x_1, \ldots, x_n)$ un vecteur de $\mathbb{R}^n$. Pour tout entier naturel k, on pose $X^{(k)} = (x_1^{(k)}, \ldots, x_n^{(k)})$. On dit que la suite $(X^{(k)})_{k\geq 0}$ converge vers X si pour tout $i \in [\![1,n]\!]$, la suite $(x_i^{(k)})_{k\geq 0}$ converge vers x_i.
Soit $(A^{(k)})_{k\geq 0}$ un suite de matrices de $\mathcal{M}_n(\mathbb{R})$ et $A = (a_{i,j})_{1\leq i,j\leq n}$ une matrice de $\mathcal{M}_n(\mathbb{R})$. Pour tout entier naturel k, on pose $A^{(k)} = (a_{i,j}^{(k)})_{1\leq i,j\leq n}$. On dit que la suite $(A^{(k)})_{k\geq 0}$ converge vers A si pour tous $i, j \in [\![1,n]\!]$, la suite $(a_{i,j}^{(k)})_{k\geq 0}$ converge vers $a_{i,j}$.
Soit G un graphe orienté fini et soient i et j deux sommets de ce graphe. On dit que j est un sommet voisin de i s'il existe une arête orientée de G reliant i à j.

Partie A : Une marche aléatoire sur un graphe

On considère un graphe orienté fini dont les sommets sont numérotés de 1 à n. Un point se déplace aléatoirement d'un sommet à un autre de ce graphe au cours d'étapes, le nombre d'étapes pouvant tendre vers l'infini. À chaque étape, le point se déplace du sommet où il se trouve vers l'un des sommets voisins de façon équiprobable. Ceci entraîne notamment que la probabilité de passer du sommet i au sommet j ne dépend pas du rang de l'étape.
Pour $1 \leq i, j \leq n$, on note $a_{i,j}$ la probabilité que le point passe du sommet i au sommet j ; en particulier, s'il n'y a pas d'arête reliant i à j, $a_{i,j} = 0$. La matrice dont le coefficient de la ligne i et de la colonne j est égal à $a_{i,j}$ est notée A. Cette matrice s'appelle la matrice de transition du graphe.

Pour $k \in \mathbb{N}$, on note $P^{(k)}$ le vecteur ligne $(p_1^{(k)}, \ldots, p_n^{(k)})$, où $1 \leq i \leq n$, $p_i^{(n)}$ est la probabilité que le point soit sur le sommet i à l'étape de rang k.

A.I. Résultats généraux.

A.I.1. Justifier que, pour tout entier naturel k, $p_1^{(k)} + \cdots + p_n^{(k)} = 1$.
Posons, pour tout $k \in \mathbb{N}$, la variable aléatoire X_k qui indique la position du point à l'étape de rang k. On a donc $X_k(\Omega) = [\![1,n]\!]$ et, pour tout $i \in [\![1,n]\!]$,
$$p_i^{(k)} = \mathbb{P}\left(X_k = i\right).$$

Ainsi,

$$\sum_{i=1}^{n} p_i^{(k)} = \sum_{i=1}^{n} \mathbb{P}\left(X_k = i\right) = \mathbb{P}\left(\bigsqcup_{i=1}^{n} (X_k = i)\right) = \mathbb{P}\left(X_k \in X_k(\Omega)\right) = \mathbb{P}\left(\Omega\right) = 1.$$

□

A.I.2. Montrer que, pour tout entier naturel k, $P^{(k+1)} = P^{(k)}A$.
Soit $i \in [\![1, n]\!]$. On a

$$\begin{aligned} p_i^{(k+1)} &= \mathbb{P}\left(X_{k+1} = i\right) \\ &= \sum_{j=1}^{n} \mathbb{P}\left(X_k = j\right) \mathbb{P}\left(X_{k+1} = i | X_k = j\right) && \text{(probabilités totales)} \\ &= \sum_{j=1}^{n} p_j^{(k)} a_{j,i} \\ &= [P^{(k)}A]_i. \end{aligned}$$

Ainsi, $P^{(k+1)} = P^{(k)}A$.

□

A.I.3. En déduire, pour tout entier naturel k, une expression de $P^{(k)}$ en fonction de A, k et $P^{(0)}$.
On montre, par récurrence sur k (dont la rédaction est laissée au lecteur), que

$$P^{(k)} = P^{(0)}A^k.$$

□

A.I.4. On suppose que la suite de vecteurs $(P^{(k)})_{k\geq 0}$ converge vers un vecteur $P = (p_1, \ldots, p_n)$. Montrer que $PA = P$ et $p_1 + \cdots + p_n = 1$.
On a, pour tout $k \in \mathbb{N}$, la relation

$$P^{(k+1)} = P^{(k)}A.$$

En passant à la limite en k, on obtient

$$P = \lim_{k:+\infty} P^{(k+1)} = \lim_{k:+\infty} P^{(k)}A.$$

Mais le produit matriciel étant continu (car polynomial en les entrées des matrices), on a

$$\lim_{k:+\infty}(P^{(k)}A) = (\lim_{k:+\infty} P^{(k)})A = PA$$

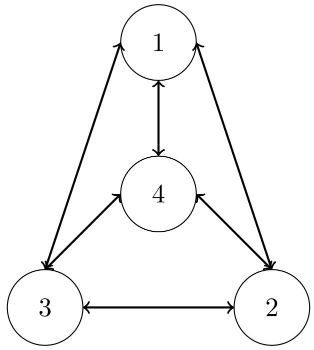
et donc

$$P = PA.$$

□

A.II. Marche aléatoire sur un tétraèdre

Dans cette question, on suppose que G est le graphe ci-dessous :



On remarque que, lorsque le point est sur l'un des sommets du graphe, il a la même probabilité de se rendre sur chacun des trois autres sommets du graphe On suppose qu'au départ, le point est sur le sommet 1, de sorte que :

$$P^{(0)} = (1, 0, 0, 0).$$

On pose

$$U = \begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix}.$$

A.II.1. Exprimer la matrice de transition A en fonction de U.

On a

$$A = \begin{pmatrix} 0 & 1/3 & 1/3 & 1/3 \\ 1/3 & 0 & 1/3 & 1/3 \\ 1/3 & 1/3 & 0 & 1/3 \\ 1/3 & 1/3 & 1/3 & 0 \end{pmatrix} = \frac{1}{3}U.$$

□

A.II.2. Calculer U^2 et U^3.

Un calcul immédiat donne

$$U^2 = 3I_4 + 2U$$

et

$$U^3 = 6I_4 + 7U.$$

□

A.II.3. Montrer qu'il existe deux suites $(\alpha_k)_{k\geq 0}$ et $(\beta_k)_{k\geq 0}$ telles que pour tout entier naturel k :

$$U^k = \begin{pmatrix} \alpha_k & \beta_k & \beta_k & \beta_k \\ \beta_k & \alpha_k & \beta_k & \beta_k \\ \beta_k & \beta_k & \alpha_k & \beta_k \\ \beta_k & \beta_k & \beta_k & \alpha_k \end{pmatrix}.$$

Montrer de plus que, pour tout entier naturel k,

$$\begin{cases} \alpha_{k+1} = 3\beta_k \\ \beta_{k+1} = \alpha_k + 2\beta_k. \end{cases}$$

Montrons qu'il existe deux suites $(\alpha_k)_{k\geq 0}$ et $(\beta_k)_{k\geq 0}$ telles que pour tout entier naturel k, on a

$$U^k = \alpha_k I_4 + \beta_k U.$$

On procède par récurrence sur k. On a

$$U^0 = I$$

de sorte que $\alpha_0 = 1$ et $\beta_0 = 0$ conviennent.

Soit $k \in \mathbb{N}$ tel que de tels α_k et β_k existent. On a alors

$$\begin{aligned} U^{k+1} &= U(U^k) \\ &= U(\alpha_k I_4 + \beta_k U) \\ &= \alpha_k U + \beta_k U^2 \\ &= \alpha_k U + \beta_k (3I_4 + 2U) \\ &= 3\beta_k I_4 + (\alpha_k + 2\beta_k)U, \end{aligned}$$

de sorte que $\alpha_{k+1} = 3\beta_k$ et $\beta_{k+1} = \alpha_k + 2\beta_k$ conviennent.

□

A.II.4. En déduire que, pour tout entier naturel k, $\beta_{k+2} = 2\beta_{k+1} + 3\beta_k$.
On a

$$\beta_{k+2} = \alpha_{k+1} + 2\beta_{k+1} = 3\beta_k + 2\beta_{k+1}.$$

□

A.II.5. En déduire que, pour tout entier naturel k,

$$\beta_k = \frac{3^k - (-1)^k}{4} \qquad \textbf{et} \qquad \alpha_k = \frac{3^k + 3(-1)^k}{4}.$$

L'équation caractéristique associée à la relation de récurrence $\beta_{k+2} = 2\beta_{k+1} + 3\beta_k$ est

$$X^2 - 2X - 3 = 0,$$

les solutions sont donc -1 et 3.

Ainsi, il existe $\lambda, \mu \in \mathbb{R}$ tels que, pour tout $k \in \mathbb{N}$, on ait

$$\beta_k = \lambda 3^n + \mu(-1)^n.$$

Par ailleurs, on a $\beta_0 = 0$ et $\beta_1 = 1$ de sorte que

$$\begin{cases} \lambda + \mu &= 0 \\ 3\lambda - \mu &= 1. \end{cases}$$

Il s'ensuit que $\lambda = \frac{1}{4}$ et $\mu = -\frac{1}{4}$ de sorte que, pour tout $k \in \mathbb{N}$,

$$\beta_k = \frac{3^k - (-1)^k}{4}.$$

Puisque, pour tout $k \in \mathbb{N}$, on a la relation $\alpha_{k+1} = 3\beta_k$, il s'ensuit que,

$$\alpha_{k+1} = \frac{3^{k+1} - 3(-1)^k}{4}$$

et donc, pour tout $k \in \mathbb{N}^*$,

$$\alpha_k = \frac{3^k + 3(-1)^k}{4}.$$

En outre, $\alpha_0 = 1$ donc la relation précédente reste vraie pour $k = 0$.

□

A.II.6. En déduire, pour tout entier naturel k une expression de $P^{(k)}$.

On a

$$\begin{aligned}
P^{(k)} &= P^{(0)} A^k \\
&= P^{(0)} \left(\frac{1}{3}\right)^k U^k \\
&= \frac{1}{3^k} P^{(0)} U^k \\
&= \frac{1}{3^k} \left(\alpha_k, \beta_k, \beta_k, \beta_k\right) \\
&= \frac{1}{3^k} \left(\frac{3^k + 3(-1)^k}{4}, \frac{3^k - (-1)^k}{4}, \frac{3^k - (-1)^k}{4}, \frac{3^k - (-1)^k}{4}\right) \\
&= \frac{1}{4} \left(1 + 3\left(\frac{-1}{3}\right)^k, 1 - \left(\frac{-1}{3}\right)^k, 1 - \left(\frac{-1}{3}\right)^k, 1 - \left(\frac{-1}{3}\right)^k\right)
\end{aligned}$$

□

A.II.7. Montrer que la suite de vecteurs $(P^{(k)})_{k\geq 0}$ converge et déterminer la limite de $P^{(k)}$ lorsque k tend vers $+\infty$.

On a

$$\lim_{k:+\infty} 1 + 3\left(\frac{-1}{3}\right)^k = 1$$

et

$$\lim_{k:+\infty} 1 - \left(\frac{-1}{3}\right)^k = 1$$

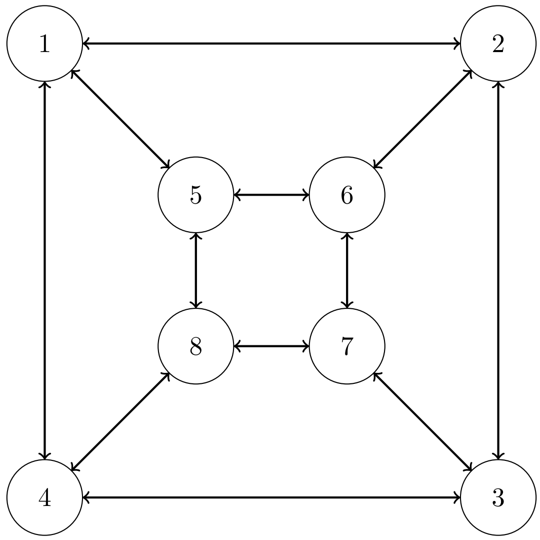
de sorte que

$$\lim_{k:+\infty} P^{(k)} = \left(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right).$$

□

A.III. Marche aléatoire sur une pyramide tronquée à base carrée.

Dans cette question, on suppose que G est le graphe ci-dessous



On rappelle que, lorsque le point est sur l'un des sommets du graphe, il a la même probabilité de se rendre sur chacun des sommets auquel il est relié. On suppose qu'au départ, le point est sur le sommet 1, de sorte que

$$P^{(0)} = (1, 0, 0, 0, 0, 0, 0, 0).$$

On note $X = \{1, 3, 6, 8\}$ et $Y = \{2, 4, 5, 7\}$.

A.III.1. Donner la matrice de transition de ce graphe et calculer

$$(1, 1, 1, 1, 1, 1, 1, 1)A.$$

Chaque sommet a trois voisins. Ainsi,

$$A = \begin{pmatrix} 0 & 1/3 & 0 & 1/3 & 1/3 & 0 & 0 & 0 \\ 1/3 & 0 & 1/3 & 0 & 0 & 1/3 & 0 & 0 \\ 0 & 1/3 & 0 & 1/3 & 0 & 0 & 1/3 & 0 \\ 1/3 & 0 & 1/3 & 0 & 0 & 0 & 0 & 1/3 \\ 1/3 & 0 & 0 & 0 & 0 & 1/3 & 0 & 1/3 \\ 0 & 1/3 & 0 & 0 & 1/3 & 0 & 1/3 & 0 \\ 0 & 0 & 1/3 & 0 & 0 & 1/3 & 0 & 1/3 \\ 0 & 0 & 0 & 1/3 & 1/3 & 0 & 1/3 & 0 \end{pmatrix}$$

et

$$(1, 1, 1, 1, 1, 1, 1, 1)A = (1, 1, 1, 1, 1, 1, 1, 1).$$

$\square$

A.III.2. Montrer que si le point se trouve sur un sommet de la partie X à une étape donnée, il se trouvera sur un sommet de Y à l'étape suivante, et que s'il se trouve sur un sommet de Y à une étape donnée, il se trouvera sur un sommet de X à l'étape suivante.

Si E, F sont des sous-ensembles de l'ensemble des sommets, on note $p_{E,F}$ la probabilité qu'un point se situant sur un sommet de E à une étape donnée se retrouve sur un sommet de F à l'étape suivante. En reprenant les notations utilisées à la partie **A.I**, on a, pour tout $k \in \mathbb{N}$,

$$\begin{aligned}
p_{X,Y} &= 1 - p_{X,X} \\
&= 1 - \mathbb{P}\left(X_{k+1} \in X \mid X_k \in X\right) \\
&= 1 - \sum_{j \in X} \mathbb{P}\left(X_{k+1} = j | X_k \in X\right) \\
&= 1 - \frac{1}{\mathbb{P}\left(X_k \in X\right)} \sum_{j \in X} \mathbb{P}\left(\left(X_{k+1} = j\right) \cap \left(X_k \in X\right)\right) \\
&= 1 - \frac{1}{\mathbb{P}\left(X_k \in X\right)} \sum_{i,j \in X} \mathbb{P}\left(\left(X_{k+1} = j\right) \cap \left(X_k = i\right)\right) \\
&= 1 - \frac{1}{\mathbb{P}\left(X_k \in X\right)} \sum_{i,j \in X} a_{i,j}.
\end{aligned}$$

Or, pour tout $(i, j) \in X \times X$, on a $a_{i,j} = 0$ (aucun sommet de X n'est relié à un autre sommet de X), donc $p_{X,X} = 0$ et $p_{X,Y} = 1$. De la même manière, $p_{Y,X} = 1 - p_{Y,Y} = 1$ car aucun sommet de Y n'est relié à un autre sommet de Y. $\square$

A.III.3.a. Démontrer que les coefficients de $P^{(k)}$ dont les indices sont des éléments de X sont nuls si k est impair, et que les coefficients de $P^{(k)}$ dont les indices sont des éléments de Y sont nuls si k est pair.

On le montre par récurrence sur k. Avec les notations de la question précédente, il s'agit de montrer que $\mathbb{P}\left(X_k \in X\right) = 0$ si k est impair et $\mathbb{P}\left(X_k \in Y\right) = 0$ si k est pair.

On a

$$\mathbb{P}\left(X_0 \in Y\right) = 1 - \mathbb{P}\left(X_0 \in X\right) \leq 1 - \mathbb{P}\left(X_0 = 1\right) = 0$$

et

$$\mathbb{P}\left(X_1 \in X\right) = \mathbb{P}\left(X_1 \in X | X_0 \in X\right) = p_{X,X} = 0.$$

Supposons la propriété vraie jusqu'à un rang k, montrons qu'elle est vraie au rang suivant. Si $k+1$ est impair, alors k est pair de sorte que X_k prend ses valeurs dans X et

$$\mathbb{P}\left(X_{k+1} \in X\right) = \mathbb{P}\left(X_{k+1} \in X | X_k \in X\right) = p_{X,X} = 0.$$

Si $k+1$ est pair, alors k est impair de sorte que X_k prend ses valeurs dans Y et

$$\mathbb{P}\left(X_{k+1} \in Y\right) = \mathbb{P}\left(X_{k+1} \in Y | X_k \in Y\right) = p_{Y,Y} = 0.$$

$\square$

A.III.3.b. La suite de vecteurs $(P^{(k)})_{k \geq 0}$ converge-t-elle ?
Raisonnons par l'absurde. Si la suite $(P^{(k)})_{k \geq 0}$ converge vers un vecteur P, alors les sous-suites $(P^{(2k)})_{k \geq 0}$ et $(P^{(2k+1)})_{k \geq 0}$ convergent aussi vers P. Mais, pour $i \in X$ et $k \in \mathbb{N}$, on a $p_i^{(2k+1)} = 0$ donc $p_i = 0$. De même, pour

$i \in Y$ et $k \in \mathbb{N}$, on a $p_i^{(2k)} = 0$ donc $p_i = 0$. Ainsi, $P = 0$. Or, pour tout $k \in \mathbb{N}$, on a

$$\sum_{k=1}^{n} p_i^{(k)} = 1$$

donc, en passant à la limite en k, on a

$$\sum_{k=1}^{n} p_i = 1,$$

une contradiction.

□

A.IV. On revient au cas général d'un graphe G à n sommets. Parmi les trois liens logiques condition nécessaire, condition suffisante et condition nécessaire et suffisante, quel est celui qui relie les deux propositions suivantes ?

(i) **la suite de vecteurs $(P^{(k)})_{k\geq 0}$ converge ;**

(ii) **il existe un vecteur $P = (p_1, \ldots, p_n)$ de $\mathbb{R}^n$, avec $p_1, \ldots, p_n$ positifs ou nuls et $p_1 + \cdots + p_n = 1$, tel que $PA = P$.**

La réponse, qui devra être soigneusement justifiée, sera présentée sous deux formes : une phrase rédigée en français et une proposition mathématiques comportant une implication ou une équivalence.

On a (i) $\Rightarrow$ (ii) mais la réciproque est fausse. Autrement dit, (ii) est une condition nécessaire à la convergence de la suite $(P^{(k)})_{k\geq 0}$ mais elle n'est pas suffisante.

En effet, si $(P^{(k)})_{k\geq 0}$ converge vers un vecteur P, on a pour tout $i \in [\![1, n]\!]$, $p_i^{(k)} \geq 0$ donc $p_i \geq 0$ et les deux autres propriétés suivent de la question A.I.4.

La réciproque est fausse. En effet, il suffit de considérer le graphe G de **A.III** et le vecteur

$$P = \frac{1}{8}(1, 1, 1, 1, 1, 1, 1).$$

Il satisfait bien aux propriétés demandées en vertu de la question **A.III.1** mais nous avons montré à la question **A.III.3.b** que la suite $(P^{(k)})_{k\geq 0}$ diverge.

□

Partie B : Matrices stochastiques et densités de probabilité

Soit $X = (x_1, \ldots, x_n) \in \mathbb{R}^n$. On dit que X est une densité de probabilité si pour tout $i \in [\![1, n]\!]$, $x_i \geq 0$ et $x_1 + \cdots + x_n = 1$.

Soit A une matrice de $\mathcal{M}_n(\mathbb{R})$. On dit que A est une matrice stochastique si chaque ligne de A est une densité de probabilité.

B.I. Soit $A \in \mathcal{M}_n(\mathbb{R})$, dont tous les coefficients sont positifs ou nuls. Montrer que A est une matrice stochastique si et seulement si

$$A \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}.$$

Soit

$$U = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}.$$

On a, pour tout $i \in [\![1, n]\!]$,

$$[AU]_i = \sum_{j=1}^{n} a_{i,j} u_j = \sum_{j=1}^{n} a_{i,j}.$$

Ainsi

$$AU = U \Leftrightarrow (\forall i \in [\![1, n]\!], \quad \sum_{j=1}^{n} a_{i,j} = 1)$$

et donc

$$AU = U \Leftrightarrow A \text{ est stochastique.}$$

$\square$

B.II. Montrer que la matrice de transition d'un graphe (définie dans la partie A) est une matrice stochastique et que, pour tout entier naturel k, le vecteur $P^{(k)}$, lui aussi défini dans la partie A est une densité de probabilité.

Avec les notations de la partie **A**, pour tout $i \in [\![1, n]\!]$, on a

$$\sum_{j=1}^{n} a_{i,j} = \sum_{j=1}^{n} \mathbb{P}\left(X_{k+1} = j | X_k = i\right) = \mathbb{P}\left(\Omega | X_k = i\right) = 1.$$

Ainsi, A est une matrice stochastique.

De la même manière, pour tout $k \in \mathbb{N}$, on a

$$p_i^{(k)} = \mathbb{P}\left(X_k = i\right) \geq 0$$

et

$$\sum_{i=1}^{n} p_i^{(k)} = \sum_{i=1}^{n} \mathbb{P}\left(X_k = i\right) = 1$$

donc $P^{(k)}$ est une densité de probabilité.

$\square$

B.III. Soit $A \in \mathcal{M}_n(\mathbb{R})$ une matrice stochastique et $X \in \mathbb{R}^n$ une densité de probabilité. Montrer que XA est une densité de probabilité.

Si $X = (x_1, \dots, x_n)$ est une densité de probabilité, on a pour tout $i \in [\![1, n]\!]$,

$$[XA]_i = \sum_{j=1}^{n} x_j a_{j,i}$$

donc

$$\begin{aligned}\sum_{i=1}^{n} [XA]_i &= \sum_{i=1}^{n} \sum_{j=1}^{n} x_j a_{j,i} \\ &= \sum_{j=1}^{n} x_j \sum_{i=1}^{n} a_{j,i} \\ &= \sum_{j=1}^{n} x_j = 1\end{aligned}$$

de sorte que XA est encore une densité de probabilité.

$\square$

B.IV. Soient A et B deux matrices stochastiques de $\mathcal{M}_n(\mathbb{R})$.

B.IV.1. Montrer que AB est une matrice stochastique.
On a, pour tout $i \in [\![1, n]\!]$,

$$\begin{aligned}\sum_{j=1}^{n} [AB]_{i,j} &= \sum_{j=1}^{n} \sum_{k=1}^{n} a_{i,k} b_{k,j} \\ &= \sum_{k=1}^{n} a_{i,k} \sum_{j=1}^{n} b_{k,j} \\ &= \sum_{k=1}^{n} a_{i,k} = 1.\end{aligned}$$

$\square$

B.IV.2. Soit $\alpha \in [0, 1]$. Montrer que $\alpha A + (1 - \alpha)B$ est une matrice stochastique.
On a, pour tout $i \in [\![1, n]\!]$,

$$\begin{aligned}\sum_{j=1}^{n} [\alpha A + (1 - \beta)B]_{i,j} &= \sum_{j=1}^{n} \alpha a_{i,j} + (1 - \alpha) b_{i,j} \\ &= \alpha \sum_{j=1}^{n} a_{i,j} + (1 - \alpha) \sum_{i=1}^{n} b_{i,j} \\ &= \alpha + (1 - \alpha) = 1.\end{aligned}$$

$\square$

B.V. Soit $(X^{(k)})_{k\geq 0}$ une suite de vecteurs de $\mathbb{R}^n$, convergeant vers un vecteur X. Montrer que, si pour tout entier naturel k, $X^{(k)}$ est une densité de probabilité, alors X est une densité de probabilité.
On a, pour tout $k \in \mathbb{N}$,

$$\forall i \in [\![1, n]\!], \quad x_i^{(k)} \geq 0.$$

En passant à la limite en k, on a donc

$$\forall i \in [\![1, n]\!], \quad x \geq 0.$$

En outre, pour tout $k \in \mathbb{N}$,

$$\sum_{i=1}^{n} x_i^{(k)} = 1$$

et donc, en passant à la limite en k, on obtient

$$\sum_{i=1}^{n} x_i = 1.$$

□

B.VI. Soit $(A^{(k)})_{k\geq 0}$ une suite de matrices de $\mathcal{M}_n(\mathbb{R})$ convergeant vers une matrice A. Montrer que, si pour tout entier naturel k, $A^{(k)}$ est une matrice stochastique, alors A est une matrice stochastique.
On a, pour tout $i \in [\![1, n]\!]$ et $k \in \mathbb{N}$,

$$\sum_{j=1}^{n} a_{i,j}^{(k)} = 1$$

et, en passant à la limite en k, on obtient

$$\sum_{j=1}^{n} a_{i,j} = 1$$

de sorte que A est stochastique.

□

C. PageRank, un premier modèle

Le moteur de recherche Google annonce environ 339 000 résultats à la requête ≪ capes mathématiques ≫. Sur la première page de résultats, il figure en premier le site `capes-math.org`, puis le site `www.devenirenseignant.gouv.fr`. Le succès de ce moteur de recherche tient à sa rapidité, son exhaustivité et la qualité du classement qu'il effectue entre les différentes pages, à partir d'une ≪ mesure d'importance ≫attribuée à chacune. Cette mesure d'importance d'une page s'appelle sa pertinence. On se propose d'étudier deux algorithmes permettant de déterminer la pertinence de chaque page web. Le second algorithme, connu sous le nom de *PageRank*, a été inventé par Sergey Brin et Larry Page.

Le mot ≪ PageRank ≫lui-même est une marque déposée par Google. Le procédé a été breveté par l'université de Stanford qui en a accordé une licence exclusive à Google.
On modélise le web par un graphe orienté à n sommets représentant chacun une page et dont les arêtes représentent les liens entre celles-ci (liens hypertextes, citations d'articles, etc.) ; la notation $i \longrightarrow j$ traduit le fait que la page i pointe vers la page j.
Pour tout entier $i \in [\![1, n]\!]$, λ_i désigne le nombre de liens issus de la page i et μ_i désigne la pertinence

de cette page.
On impose à la suite $(\mu_j)_{1\leq j\leq n}$ des pertinences des pages du web de vérifier les deux conditions suivantes :

(i) **la pertinence d'une page dépend des pertinences des pages qui pointent vers elle de manière affine ;**

(ii) **plus il y a de liens issus d'une page, plus la contribution de cette page dans la pertinence des pages vers lesquelles elle pointe est faible.**

C.I. Montrer que toute suite $(\mu_j)_{1\leq j\leq n}$ de nombres réels positifs ou nuls satisfaisant aux n conditions

$$(*) \qquad \forall j \in [\![1, n]\!], \quad \mu_j = \sum_{i\longrightarrow j} \frac{1}{\lambda_i}\mu_i$$

vérifie les conditions (i) et (ii).
La pertinence μ_j d'une page j est donnée par

$$\mu_j = \sum_{i\longrightarrow j} \frac{1}{\lambda_i}\mu_i.$$

Elle dépend donc affinement de chaque μ_i où $i \longrightarrow j$, c'est-à-dire de la pertinence des pages qui pointent vers elle. En outre, plus une page i possède de liens sortants, plus λ_i augmente et donc $\frac{1}{\lambda_i}$ diminue. Ainsi, la contribution $\frac{1}{\lambda_i}\mu_i$ de cette page à la pertinence μ_j est plus faible.

Les solutions de (*) vérifient donc (i) et (ii).

□

Toute solution $(\mu_j)_{1\leq j\leq n}$ du système linéaire (*) formée de réels positifs ou nuls fournit les pertinences des n pages du web. Dans la pratique, la résolution de ce système est impossible, le nombre n de pages du web étant beaucoup trop élevé. Pour cette raison, on cherche une autre méthode pour obtenir les pertinences.

C.II. Soient $i, j \in [\![1, n]\!]$.

Si $i \neq j$, on pose

$$a_{i,j} = \begin{cases} 0 & \textbf{s'il n'y a pas de lien de la page } i \textbf{ vers la page } j \text{ ;} \\ \dfrac{1}{\lambda_i} & \textbf{si } i \longrightarrow j. \end{cases}$$

Si $i = j$, dans le cas où il n'y a aucun lien qui permette de sortir de la page i, on convient d'écrire $a_{i,i} = 1$, pour indiquer que si on arrive sur cette page, on y reste ; sinon, on note $a_{i,i} = 0$.

Pour $1 \leq i, j \leq n$, le coefficient $a_{i,j}$ peut s'interpréter comme la probabilité qu'une personne navigant sur le web et se trouvant sur la page i suive le lien vers la page j en choisissant au hasard entre les λ_i liens disponibles à partir de la page i s'il y en a et reste sur la page i sinon. Soit A la matrice de $\mathcal{M}_n(\mathbb{R})$ dont le coefficient de la ligne i et de la colonne j est $a_{i,j}$.

Montrer que A est une matrice stochastique.
Soit $i \in [\![1,n]\!]$. Si i ne pointe vers aucune page, alors

$$\sum_{j=1}^{n} a_{i,j} = a_{i,i} = 1.$$

Sinon,

$$\begin{aligned}\sum_{j=1}^{n} a_{i,j} &= \sum_{i\longrightarrow j} \frac{1}{\lambda_i} \\ &= \frac{1}{\lambda_i}\sum_{i\longrightarrow j} 1 \\ &= \frac{1}{\lambda_i}\cdot\lambda_i = 1\end{aligned}$$

□

C.III. Montrer que la suite finie $(\mu_j)_{1\leq j\leq n}$ vérifie les conditions (*) introduites à la question C.I si et seulement si la matrice ligne $M = (\mu_1, \mu_2, \dots, \mu_n)$ vérifie $M = MA$.
On a, pour tout $j \in [\![1,n]\!]$,

$$[MA]_j = \sum_{i=1}^{n} \mu_i a_{i,j} = \sum_{i\longrightarrow j} \mu_i \cdot \frac{1}{\lambda_i}$$

Ainsi, $(\mu_j)_{1\leq j\leq n}$ vérifie (*) si et seulement si $MA = M$.

□

C.IV. On note dans la suite $P^{(k)}$ la densité de probabilité après k clics : le coefficient d'indice i de la matrice ligne $P^{(k)}$ est la probabilité que le « surfeur »se trouve sur la page i après k clics.

C.IV.1. Vérifier que pour tout entier naturel k, $P^{(k+1)} = P^{(k)}A$.
On identifie le surfeur à un point se déplaçant aléatoirement sur le graphe orienté modélisant le réseau internet. La matrice A construite précédemment n'est autre que la matrice de transition correspondant à ce graphe et la réponse à la question suit alors de **A.I.2**.

□

C.IV.2. Montrer que si la suite $(P^{(k)})_{k\geq 0}$ converge vers une densité de probabilité limite P, alors cette densité de probabilité vérifie les conditions (*) et fournit donc une mesure de la pertinence des pages du web.
C'est une conséquence directe de **C.IV.1**, **A.I.4**, **C.III** et **C.I**.

□

Partie D : PageRank, un second modèle

On a vu dans la partie A que l'existence de cette densité de probabilité limite n'est pas toujours assurée. De plus, il peut arriver que certaines pages ne comportent aucun lien vers d'autres pages.

Selon le premier modèle, lorsque le surfeur arrive sur l'une d'entre elles, il lui est impossible de la quitter. Ce modèle n'étant pas conforme à la réalité, on étudie un second modèle qui introduit la possibilité de quitter à chaque clic une page quelconque pour se diriger vers une autre choisie au hasard, et ce avec une probabilité fixe $\alpha \in]0,1[$, appelée *facteur d'amortissement*.

Non seulement ce modèle permet de mieux rendre compte de la réalité, mais en plus on va montrer qu'il garantit l'existence d'une densité de probabilité limite fournissant une mesure de la pertinence des n pages.

On considère les matrices

$$B = (1-\alpha)A + \frac{\alpha}{n}J$$

où J est la matrice de $\mathcal{M}_n(\mathbb{R})$ dont tous les coefficients sont égaux à 1, A la matrice stochastique définie dans la partie C et L la matrice ligne possédant n coefficients tous égaux à 1 :

$$J = \begin{pmatrix} 1 & \cdots & 1 \\ \vdots & & \vdots \\ 1 & \cdots & 1 \end{pmatrix}, \qquad L = (1, \dots, 1).$$

D.I. Montrer que B est une matrice stochastique.

Pour tout $i \in [\![1,n]\!]$, on a

$$\begin{aligned} \sum_{j=1}^{n} [B]_{i,j} &= (1-\alpha)\sum_{j=1}^{n} a_{i,j} + \frac{\alpha}{n}\sum_{j=1}^{n}[J]_{i,j} \\ &= (1-\alpha) + \frac{\alpha}{n}\sum_{j=1}^{n} 1 \\ &= (1-\alpha) + \alpha = 1. \end{aligned}$$

□

D.II. On suppose qu'il existe un vecteur ligne $N = (\nu_1, \dots, \nu_n)$ de réels positifs ou nuls vérifiant $N = NB$, tel que $\nu_1 + \cdots + \nu_n = 1$. Montrer que la suite $(\nu_1, \dots, \nu_n)$ vérifie les conditions (i) et (ii) énoncées dans la partie C.

On a

$$\begin{aligned} N = NB &\Leftrightarrow \forall j \in [\![1,n]\!], \nu_j = \sum_{i=1}^{n} \nu_i b_{i,j} \\ &\Leftrightarrow \forall j \in [\![1,n]\!], \nu_j = \sum_{i=1}^{n} \nu_i \left((1-\alpha)a_{i,j} + \frac{\alpha}{n}\right) \\ &\Leftrightarrow \forall j \in [\![1,n]\!], \nu_j = (1-\alpha)\sum_{i=1}^{n} \nu_i a_{i,j} + \frac{\alpha}{n}\sum_{i=1}^{n}\nu_i \\ &\Leftrightarrow \forall j \in [\![1,n]\!], \nu_j = (1-\alpha)\sum_{i \longrightarrow j} \nu_i \cdot \frac{1}{\lambda_i} + \frac{\alpha}{n}. \end{aligned}$$

Ainsi, pour tout $j \in [\![1, n]\!]$, ν_j dépend affinement des autres ν_i et la contribution de chaque i à ν_j est décroissante en λ_i. Ainsi, N satisfait bien les conditions (i) et (ii).

□

D.III. Montrer que si Q est une densité, alors $QJ = L$.
On a, pour tout $i \in [\![1, n]\!]$,

$$[QJ]_i = \sum_{j=1}^{n} [Q]_j [J]_{j,i} = \sum_{j=1}^{n} [Q]_j = 1.$$

□

D.IV. Justifier que la matrice colonne U de taille n dont tous les coefficients sont égaux à 1 vérifie $AU = U$. En déduire que 1 est valeur propre de la matrice A.
Pour tout $i \in [\![1, n]\!]$, on a

$$\begin{aligned}[AU]_i &= \sum_{j=1}^{n} a_{i,j} [U]_j \\ &= \sum_{i \longrightarrow j} \frac{1}{\lambda_i} = 1 = [U]_i.\end{aligned}$$

Ainsi $AU = U$, ce qui signifie que U est un vecteur propre de A associé à la valeur propre 1.

□

D.V. Soit $\lambda \in \mathbb{C}$ une valeur propre de A et $Z = \begin{pmatrix} z_1 \\ \vdots \\ z_n \end{pmatrix}$, où $z_i \in \mathbb{C}$ pour tout $i \in [\![1, n]\!]$, un vecteur propre associé. On note

$$|Z| = \max_{1 \leq j \leq n} |z_j|.$$

D.V.1. À l'aide de l'inégalité triangulaire appliquée à une coordonnée bien choisie du vecteur AZ, montrer que $|\lambda| \leq 1$.
On a $AZ = \lambda Z$ donc, pour tout $i \in [\![1, n]\!]$, on a

$$[AZ]_i = \lambda z_i.$$

Soit i_0 tel que

$$|Z| = |z_{i_0}|,$$

l'indice d'une composante de module maximal. On a donc

$$|[AZ]_{i_0}| = |\lambda z_{i_0}|$$

mais

$$|[AZ]_{i_0}| = \left| \sum_{j=1}^{n} a_{i_0,j} z_j \right| \leq \sum_{j=1}^{n} |a_{i_0,j}| \, |z_j| \leq \sum_{j=1}^{n} |a_{i_0,j}| \, |z_{i_0}| = |Z|$$

et $|\lambda z_{i_0}| = |\lambda| \cdot |Z|$. Ainsi

$$|\lambda| \cdot |Z| \leq |Z|.$$

Puisque Z est non-nul, on a $|Z| \neq 0$ et donc $|\lambda| \leq 1$.

□

D.V.2. En déduire que, si β est un nombre réel strictement supérieur à 1, la matrice $A - \beta I_n$ est inversible.

On vient de montrer à la question **D.V.1** que les valeurs propres de A sont toutes de module inférieur ou égal à 1. Mais, pour tout nombre complexe β, on a

$$\begin{aligned} \beta \text{ valeur propre de } A &\Leftrightarrow \mathrm{Ker}(A - \beta I_n) \neq \{0\} \\ &\Leftrightarrow A - \beta I_n \text{ non inversible.} \end{aligned}$$

Ainsi, si $\beta > 1$, alors β n'est pas valeur propre et donc $A - \beta I_n$ est inversible.

□

D.V.3. En déduire que si γ est un nombre réel strictement compris entre 0 et 1, la matrice $I_n - \gamma A$ est inversible.

Puisque $\gamma \neq 0$, on a l'équivalence (obtenue en multipliant par $-\frac{1}{\gamma}$) :

$$I_n - \gamma A \text{ inversible} \Leftrightarrow -\frac{1}{\gamma} I_n + A \text{ inversible}$$

mais $\gamma \in]0, 1[$ donc $\frac{1}{\gamma} > 1$. Le résultat suit donc de la question **D.V.2**.

□

D.VI.1. Justifier que la matrice $I_n - (1 - \alpha)A$ est inversible. Cela permet de poser

$$H = \frac{\alpha}{n} L(I_n - (1 - \alpha)A)^{-1}.$$

On a $\alpha \in]0, 1[$ donc $(1 - \alpha) \in]0, 1[$ et $I_n - (1 - \alpha)A$ est inversible en vertu de **D.V.3**.

□

D.VI.2. Vérifier que $H = (1 - \alpha)HA + \frac{\alpha}{n} L$.

On a

$$H = \frac{\alpha}{n} L(I_n - (1 - \alpha)A)^{-1}$$

donc en multipliant à droite par $(I_n - (1 - \alpha)A)$, on obtient

$$H(I_n - (1 - \alpha)A) = \frac{\alpha}{n} L$$

et donc

$$H - (1 - \alpha)A = \frac{\alpha}{n} L$$

ou encore

$$H = (1 - \alpha)A + \frac{\alpha}{n} L.$$

□

D.VII. Soit Q une densité de probabilité. On définit la suite $(Q^{(k)})_{k\geq 0}$ par $Q^{(k)} = QB^k$ pour tout entier naturel k.

D.VII.1. Démontrer, pour tout entier naturel k, l'égalité $(Q^{(k+1)} - H) = (1-\alpha)(Q^{(k)} - H)A$.
Pour tout $k \in \mathbb{N}$, on a

$$\begin{aligned}
Q^{(k+1)} - H &= Q^{(k)}B - H \\
&= Q^{(k)}\left((1-\alpha)A + \frac{\alpha}{n}J\right) - H \\
&= (1-\alpha)Q^{(k)}A + \frac{\alpha}{n}Q^{(k)}J - (1-\alpha)HA - \frac{\alpha}{n}L && \text{(d'après } \textbf{D.VI.2.}\text{)} \\
&= (1-\alpha)Q^{(k)}A + \frac{\alpha}{n}L - (1-\alpha)HA - \frac{\alpha}{n}L && \text{(d'après } \textbf{D.III.}\text{)} \\
&= (1-\alpha)Q^{(k)}A - (1-\alpha)HA \\
&= (1-\alpha)(Q^{(k)} - H)A.
\end{aligned}$$

□

D.VII.2. En déduire que, pour tout entier naturel k,

$$Q^{(k)} - H = (1-\alpha)^k(Q-H)A^k.$$

C'est une récurrence immédiate sur k à partir de la relation établie à la question **D.VII.1**.

□

D.VIII. En utilisant le fait que A^k est une matrice stochastique, montrer que les suites formées par les coefficients des matrices lignes $(Q-H)A^k$ sont bornées. Que peut-on en déduire pour la suite des densités de probabilité $Q^{(k)}$?
Commençons par remarquer que A^k est effectivement stochastique en vertu de la question **B.IV.1** et d'une récurrence sur k. Ses coefficients sont en outre positifs de sorte qu'ils sont tous majorés par 1. Ainsi, la suite $(A^k)_{k\in\mathbb{N}}$ est bornée et il en est de même de la suite $((Q-H)A^k)_{k\in\mathbb{N}}$.

Par ailleurs $(1-\alpha) \in]0,1[$ donc $(1-\alpha)^k \xrightarrow[k\longrightarrow+\infty]{} 0$. Il s'ensuit que la suite de terme général $(1-\alpha)^k(Q-H)A^k$ tend vers zéro.

Il suit alors de la question D.VII.2 que

$$Q^{(k)} \xrightarrow[k\longrightarrow+\infty]{} H.$$

□

D.IX. Que peut-on en conclure pour le vecteur H ?
Pour tout $k \in \mathbb{N}$, on a $Q^{(k)} = QB^k$ donc $Q^{(k+1)} = Q^{(k)}B$. En passant à la limite en k, on obtient

$$H = HB$$

et H est une densité de probabilité comme limite de suite de densités de probabilité (voir **B.V**). Il suit alors de la question **D.II** que H vérifie les conditions (i) et (ii).

□

D.X. Quel est l'avantage de cette méthode par rapport à celle exposée dans la partie C ?
Outre les avantages de modélisation déjà explicités dans l'introduction de la partie **D**, cette méthode a l'avantage de ne pas nécessiter la résolution du système (*) pour déterminer la pertinence des pages sous la contrainte des conditions (i) et (ii). Le calcul théorique de H nécessite néanmoins théoriquement l'inversion de la matrice $I_n - (1-\alpha)A$, mais comme nous venons de l'établir à la question **D.IX**, on peut obtenir une approximation de H en calculant les $Q^{(k)}$, ce qui s'obtient avec de simples exponentiations de la matrice B.

□

D.XI. Le calcul de l'inverse de la matrice $I_n - (1-\alpha)A$ permettant de calculer H étant très coûteux, on se contente d'approcher H par $Q^{(k)} = QB^k$ pour des valeurs de k suffisamment grandes. Quelle justification mathématique peut-on apporter à cette pratique ?
On a établi à la question **D.IX** que la suite $Q^{(k)}$ tend vers H de sorte qu'il est possible d'approcher H par $Q^{(k)}$ de façon aussi précise qu'on le souhaite. Néanmoins, puisque ce que l'on cherche est en fait une densité de probabilité N telle que $N = NB$ (voir **D.II**), il reste à vérifier que $Q^{(k)}$ n'est pas trop éloigné de $NQ^{(k)}$, ce qui est assuré par la continuité du produit matriciel, en effet :

$$Q^{(k)} - Q^{(k)}B \xrightarrow[k\longrightarrow+\infty]{} H - HB = 0.$$

Ainsi, pour k suffisamment grand, $Q^{(k)}$ donnera bien une solution approchée du problème.

□

Fin du problème