Devoir surveillé 4

le samedi 28 janvier

L'usage d'une calculatrice est interdit pour cette épreuve. Si, au cours de l'épreuve, un candidat repère ce qui lui semble être une erreur d'énoncé, il le signale sur sa copie et poursuit sa composition en expliquant les raisons des initiatives qu'il a été amené à prendre.

Notations :

— Pour $x \in \mathbb{R}$, on note $|x|$ la valeur absolue de x.

— Pour tout vecteur $V = \begin{pmatrix} v_1 \\ \vdots \\ v_N \end{pmatrix} \in \mathcal{M}_{N,1}(\mathbb{R})$, on note $|V| = \begin{pmatrix} |v_1| \\ \vdots \\ |v_N| \end{pmatrix} \in \mathcal{M}_{N,1}(\mathbb{R})$. et $\|V\|_1 = \sum_{j=1}^{N} |v_j|$.

— On notera I_N la matrice identité de $\mathcal{M}_N(\mathbb{R})$.

— Pour $A \in \mathcal{M}_N(\mathbb{R})$, on note ${}^tA$ la matrice transposée de A.

Définitions :

— Une matrice est dite *positive* si tous ses cœfficients sont des nombres réels positifs ou nuls. Elle est dite *strictement positive* si tous ses coefficients sont des nombres réels strictement positifs.

— Un vecteur colonne $V \in \mathcal{M}_{N,1}(\mathbb{R})$ est dit *de probabilité* si V est positif et si $\|V\|_1 = 1$.

— Une matrice $Q = (Q(i,j))_{1\leq i\leq N, 1\leq j\leq N} \in \mathcal{M}_N(\mathbb{R})$ est dite *stochastique* si elle est positive et si, pour tout $1 \leq j \leq N$, $\sum_{i=1}^{N} Q(i,j) = 1$.

— Un vecteur $V \in \mathcal{M}_{N,1}(\mathbb{R})$ est dit *invariant* par une matrice $Q \in \mathcal{M}_N(\mathbb{R})$ si $QV = V$.

PARTIE I : étude d'un premier exemple

Soit $A = \begin{pmatrix} \frac{1}{2} & \frac{1}{4} & 0 \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ 0 & \frac{1}{4} & \frac{1}{2} \end{pmatrix}$.

1. Montrer que A est stochastique. Est-elle strictement positive ?
2. Déterminer $\mathrm{Sp}(A)$, l'ensemble des valeurs propres de A.
3. Déterminer le sous-espace propre de A associé à la valeur propre 1.
4. En déduire qu'il existe un unique vecteur de probabilité invariant par A et le donner.

PARTIE II : étude d'un second exemple

On note E l'ensemble des fonctions polynomiales définies sur $]0, +\infty[$ à valeurs réelles, et de degré inférieur ou égal à 5.

Pour tout $k \in [\![0,5]\!]$, on note f_k la fonction de E définie par $f_k : \begin{array}{ccc}]0,+\infty[& \longrightarrow & \mathbb{R} \\ x & \longmapsto & x^k \end{array}$. On rappelle que la famille $(f_k)_{0\leq k\leq 5}$ est une base de E que l'on appellera $\mathcal{B}$.

Enfin, on définit sur E l'application φ par, pour tout $P \in E$, $\varphi(P) = Q$ où

$$\forall x > 0, \quad Q(x) = P(x) + 2\,x^5\,P\left(\frac{1}{x}\right)$$

1. Montrer que φ est un endomorphisme de E et donner sa matrice M dans la base $\mathcal{B}$. Vérifier que $\frac{1}{3}M$ est stochastique.
2. Montrer que φ est diagonalisable.
3. (a) Vérifier que $M^2 = 2M + 3I_6$. En déduire que $\mathrm{Sp}(\varphi) \subset \{-1, 3\}$.
   (b) Déterminer une base et la dimension de chacun des sous-espaces propres de φ.
   (c) Déterminer une matrice P inversible et une matrice D diagonale telle que $M = PDP^{-1}$.

PARTIE III : existence d'un vecteur de probabilité invariant

Soit Q une matrice stochastique de $\mathcal{M}_N(\mathbb{R})$, et soit $U = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$ la matrice colonne de $\mathcal{M}_{N,1}(\mathbb{R})$ ne comportant que des 1.

1. (a) Montrer que ${}^tQ\,U = U$.

   (b) Montrer que si $Q - I_N$ est inversible, alors ${}^tQ - I_N$ est inversible.

   (c) Déduire des questions précédentes que 1 est valeur propre de Q.

2. Soit λ une valeur propre de Q telle que $|\lambda| = 1$. Soit V un vecteur propre associé à cette valeur propre λ.

   (a) Donner les coefficients du vecteur $Q\,|V|$. En utilisant l'inégalité triangulaire, montrer que le vecteur colonne $Q\,|V| - |V|$ est positif.

   (b) Établir que la somme des composantes du vecteur $Q\,|V| - |V|$ est nulle.

   (c) En déduire que Q possède un vecteur de probabilité invariant.

On admettra pour la suite que si Q est une matrice stochastique strictement positive, alors il existe un unique vecteur de probabilité invariant par Q. On le notera V_∞.

PARTIE IV : un troisième exemple (le PageRank de Google)

En 1998, Sergey Brin et Larry Page, co-fondateurs de Google, ont introduit la notion de PageRank. Le PageRank est un indice mesurant la notoriété de chacune des pages Web référencées dans Google. Bien que les outils de calcul de cet indice soient maintenus secrets, le principe mathématique sur lequel repose ce calcul est public et peut-être résumé comme suit.

On numérote de 1 à N les pages Web référencées dans Google (on pense que $N = 10^9$ est un bon ordre de grandeur). On dira qu'une page $j \in [\![1, N]\!]$ pointe vers une autre page $i \in [\![1, N]\!]$ s'il existe un lien dans la page j permettant de rejoindre la page i en cliquant dessus.

Pour tout $j \in [\![1, N]\!]$, on note d_j le nombre de pages vers lesquelles la j-ème page pointe.

— Lorsque $d_j = 0$, alors, pour tout $i \in [\![1, N]\!]$, on pose $A(i,j) = \begin{cases} 0 & \text{si } i \neq j \\ 1 & \text{si } i = j \end{cases}$.

— Lorsque $d_j > 0$, alors, pour tout $i \in [\![1, N]\!]$, on pose $A(i,j) = \begin{cases} \frac{1}{d_j} & \text{si } j \text{ pointe vers } i \\ 0 & \text{sinon} \end{cases}$.

Si $\rho \in [0, 1[$, on définit la matrice de Google $G = (G(i,j))_{\substack{1 \le i \le N \\ 1 \le j \le N}}$ par

$$\forall (i,j) \in [\![1, N]\!]^2, \quad G(i,j) = \rho A(i,j) + \frac{(1-\rho)}{N}$$

c'est à dire $G = \rho\, A + (1-\rho)\, J$ où $J = \dfrac{1}{N}\begin{pmatrix} 1 & \cdots & 1 \\ \vdots & & \vdots \\ 1 & \cdots & 1 \end{pmatrix} \in \mathcal{M}_N(\mathbb{R})$ (il semblerait que ρ soit de l'ordre de 0.85 en pratique).

Pour tout $j \in [\![1, N]\!]$, le PageRank d'une page j est un nombre réel positif ou nul noté $p(j)$. Les $p(j)$, pour $1 \le j \le N$, sont par ailleurs définis par le système d'équations suivant :

$$\sum_{j=1}^{N} p(j) = 1 \quad \text{et} \quad \forall i \in [\![1, N]\!], \quad \sum_{j=1}^{N} G(i,j)\, p(j) = p(i) \qquad (\mathcal{S})$$

A la question « que mesure exactement pour une page $j \in [\![1, N]\!]$ donnée ce fameux PageRank $p(j)$? », leurs concepteurs assurent qu'il s'agit de la « chance » qu'un surfeur se retrouve sur la page j en question. L'objectif de la suite de ce sujet est de lever un coin du voile entourant le mystère du PageRank en justifiant d'une part de l'existence et de l'unicité de la solution du système $(\mathcal{S})$ et en fournissant d'autre part une interprétation probabiliste de ce système permettant de donner un sens mathématique aux affirmations de Brin et Page.

A : étude de la matrice de Google

1. On suppose uniquement dans cette question que $N = 4$ et on considère le diagramme suivant modélisant les différents liens entre les 4 pages considérées.

   Donner la matrice A correspondante, puis la matrice de Google G en prenant $\rho = \dfrac{2}{3}$.

Dorénavant, N est à nouveau quelconque (de l'ordre de 10^9 si vous préférez).

2. Montrer que la matrice de Google G est une matrice strictement positive.
3. Pour tout $j \in [\![1, N]\!]$, montrer que
$$\sum_{i=1}^{N} G(i,j) = 1$$
   indication : on pourra distinguer les cas $d_j = 0$ et $d_j > 0$.
   Que peut-on en déduire pour G ?
4. Montrer que le système d'équations $(\mathcal{S})$ possède une unique solution.

B : modèle du surfeur sur le Web

Dans toute cette partie $(\Omega, \mathcal{T}, P)$ désigne un espace probabilisé et toutes les variables aléatoires utilisées sont toutes définies sur cet espace. On rappelle que l'on numérote de 1 à N les N pages Web référencées dans Google. On considère un internaute surfant sur le Web en utilisant Google, on note X_0 la première page visitée et X_n la page sur laquelle il se retrouve au bout de n opérations (soit de click sur un lien dans une page soit d'abandon au profit d'une autre adresse). Pour tout $n \in \mathbb{N}$, la variable aléatoire X_n est à valeurs dans $[\![1, N]\!]$ et on admettra que la suite de variables aléatoires $(X_n)_{n\in\mathbb{N}}$ vérifie

$$\forall n \geq 1\,, \quad \forall\,(i,j) \in [\![1,N]\!]^2\,, \quad P_{[X_{n-1}=j]}\,(X_n = i) = P\,(X_n = i | X_{n-1} = j) = G\,(i,j)$$

où G est la matrice de Google.

Pour tout $n \in \mathbb{N}$, on note $V_n = \begin{pmatrix} P(X_n = 1) \\ P(X_n = 2) \\ \vdots \\ P(X_n = N) \end{pmatrix} \in \mathcal{M}_{N,1}(\mathbb{R})$.

1. Vérifier que, pour tout $n \in \mathbb{N}$, V_n est un vecteur de probabilité.
2. Montrer que, pour tout $n \in \mathbb{N}$, $V_{n+1} = GV_n$.
3. En déduire, pour tout $n \in \mathbb{N}$, une expression de V_n à l'aide de G^n et V_0.
4. Utiliser le résultat de la partie V (oui, celle qui suit) pour établir que, pour tout $j \in [\![1, N]\!]$, $\lim\limits_{n\to+\infty} P(X_n = j) = p(j)$.

PARTIE V : méthode de la puissance pour calculer V_∞

Dans cette partie, Q désigne une matrice stochastique strictement positive.

A : valeurs propres de Q

1. Montrer que, pour tout $V \in \mathcal{M}_{N,1}(\mathbb{R})$, $\|QV\|_1 \leq \|V\|_1$, avec de plus $\|QV\|_1 = \|V\|_1$ quand V est positif.
2. En déduire que toutes les valeurs propres λ de Q vérifient $|\lambda| \leq 1$.
3. Soient λ une valeur propre de Q telle que $|\lambda| = 1$ et $V = \begin{pmatrix} v_1 \\ \vdots \\ v_N \end{pmatrix} \in \mathcal{M}_{N,1}(\mathbb{R})$ un vecteur propre associé tel que $\|V\|_1 = 1$. On sait d'après la partie III que $|V| = V_\infty$, où V_∞ est l'unique vecteur de probabilité invariant par Q.
   (a) Établir l'identité
   $$\left|\sum_{j=1}^{N} Q(1,j)v_j\right| = \sum_{j=1}^{N} Q(1,j)|v_j|$$
   (b) En déduire que tous les v_j sont de même signe.
   indication : utiliser sans démonstration l'identité
   $$\left(\sum_{j=1}^{N} Q(1,j)|v_j|\right)^2 - \left(\sum_{j=1}^{N} Q(1,j)v_j\right)^2 = \sum_{\substack{1\leq j\leq N \\ 1\leq k\leq N \\ j\neq k}} Q(1,j)Q(1,k)\big(|v_j||v_k| - v_j v_k\big)$$
   (c) En déduire que V est colinéaire à $|V|$, puis que $\lambda = 1$.

On vient donc d'établir que toutes les valeurs propres de Q sont de module inférieur ou égal à 1 et que 1 est la seule valeur propre de module 1 (et que son sous-espace propre est de dimension 1).

B : convergence vers le vecteur de probabilité invariant

On suppose dans cette question que Q est diagonalisable. On note $\lambda_1, \dots, \lambda_N$ ses valeurs propres non nécessairement distinctes avec, compte-tenu de la partie V.A, $\lambda_1 = 1$ et, pour tout $k \in [\![2, N]\!]$, $|\lambda_k| < 1$. Soit V_0 un vecteur de probabilité positif quelconque.

1. Montrer que
   $$\forall n \in \mathbb{N}, \quad \|Q^n V_0\|_1 = 1$$
2. Justifier l'existence d'une famille $\mathcal{C} = (W_1, \dots, W_N)$ de vecteurs colonnes de $\mathcal{M}_{N,1}(\mathbb{R})$ telle que, pour tout $k \in [\![1, N]\!]$, $QW_k = \lambda_k W_k$, ainsi que l'existence de réels $\alpha_1, \dots, \alpha_N$ tels que
   $$V_0 = \alpha_1 W_1 + \cdots + \alpha_N W_N$$
3. On dira qu'une suite de vecteurs colonnes $C_n = \begin{pmatrix} C_n(1) \\ \vdots \\ C_n(N) \end{pmatrix} \in \mathcal{M}_{N,1}(\mathbb{R})$ converge vers un vecteur colonne $C = \begin{pmatrix} C(1) \\ \vdots \\ C(N) \end{pmatrix} \in \mathcal{M}_{N,1}(\mathbb{R})$ si, pour tout $j \in [\![1, N]\!]$, $\lim\limits_{n\to+\infty} C_n(j) = C(j)$.
   (a) Montrer que la suite $(Q^n V_0)_{n\in\mathbb{N}}$ converge vers $\alpha_1 W_1$.
   (b) En déduire que la suite $(Q^n V_0)_{n\in\mathbb{N}}$ converge vers V_∞.