BANQUE COMMUNE D'EPREUVES

Concepteur : ESSEC

OPTION LETTRES ET SCIENCES HUMAINES

Filière B/L

MATHEMATIQUES

Mercredi 7 mai 2008, de 14h à 18h

La présentation, la lisibilité, l'orthographe, la qualité de la rédaction, la clarté et la précision des raisonnements entreront pour une part importante dans l'appréciation des copies.
Les candidats sont invités à encadrer dans la mesure du possible les résultats de leurs calculs.
Ils ne doivent faire usage d'aucun document ; l'utilisation de toute calculatrice et de tout matériel électronique est interdite. Seule l'utilisation d'une règle graduée est autorisée.
Si au cours de l'épreuve un candidat repère ce qui lui semble être une erreur d'énoncé, il le signalera sur sa copie et poursuivra sa composition en expliquant les raisons des initiatives qu'il sera amené à prendre.

L'exercice et le problème sont indépendants.

Exercice

Soit n un entier strictement positif et f_n l'application définie sur $\mathbb{R}$ par

$$f_n(x) = x\exp(x) - n$$

pour tout x réel.

1. Montrer que l'équation $f_n(x) = 0$ admet une unique solution réelle notée u_n et que cette solution est strictement positive.
2. Montrer que pour $n \geq 3$
$$1 \leq u_n \leq \ln(n).$$
3. Montrer que pour tout n strictement positif, $\ln(u_n) + u_n = \ln(n)$.
4. Montrer que
$$u_n \underset{n\to+\infty}{\sim} \ln(n).$$
5. Donner un équivalent simple de $u_n - \ln(n)$ lorsque n tend vers $+\infty$.

Problème

Tout au long du problème N désigne un entier naturel supérieur ou égal à 2.

Notations :

- Pour $x \in \mathbb{R}$, on note $|x|$ la *valeur absolue* de x.
- Pour tout vecteur $V = \begin{pmatrix} v_1 \\ \cdot \\ \cdot \\ v_N \end{pmatrix} \in \mathcal{M}_{N,1}(\mathbb{R})$, on note $|V| = \begin{pmatrix} |v_1| \\ \cdot \\ \cdot \\ |v_N| \end{pmatrix} \in \mathcal{M}_{N,1}(\mathbb{R})$

  et $\|V\|_1 = \sum_{j=1}^{N} |v_j|$.
- On notera I_N la matrice *identitée* de $\mathcal{M}_N(\mathbb{R})$.
- Pour $A \in \mathcal{M}_N(\mathbb{R})$, on note ${}^tA$ la matrice *transposée* de A.

Résultat admis :

- Pour $(A, B) \in \mathcal{M}_N(\mathbb{R})^2$, on a ${}^t(AB) = {}^tB\,{}^tA$.

Définitions :

- Une matrice est dite *positive* si tous ses coefficients 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 $\sum_{i=1}^{N} Q(i,j) = 1$ pour tout $1 \leq j \leq N$.
- Un vecteur $V \in \mathcal{M}_{N,1}(\mathbb{R})$ est dit *invariant* par $Q \in \mathcal{M}_N(\mathbb{R})$ si $QV = V$.
- Soit $Q \in \mathcal{M}_N(\mathbb{R})$. La suite de matrice $(Q^n)_{n \geq 0}$ est convergente vers la matrice Q_∞ si pour tout $(i,j) \in \{1, ..., N\}^2$, $(Q^n(i,j))_{n \geq 0}$ converge vers $Q_\infty(i,j)$.

Préliminaires

Soit $V = \begin{pmatrix} v_1 \\ \cdot \\ \cdot \\ v_N \end{pmatrix} \in \mathcal{M}_{N,1}(\mathbb{R})$ tel que

$$\left| \sum_{j=1}^{N} v_j \right| = \sum_{j=1}^{N} |v_j| . \tag{E}$$

P1. Montrer que, pour tout x réel, $|x| - x \geq 0$.

P2. Etude du cas $N = 3$. On supposera donc, dans cette question P2 uniquement, $N = 3$.

P2.a Montrer que

$$(|v_1| + |v_2| + |v_3|)^2 - (v_1 + v_2 + v_3)^2 = 2(|v_1 v_2| - v_1 v_2) + 2(|v_1 v_3| - v_1 v_3) + 2(|v_2 v_3| - v_2 v_3).$$

P2.b Montrer à l'aide de (E), que si $|v_j v_{j'}| > 0$ pour $(j, j') \in \{1, 2, 3\}^2$ tel que $j \neq j'$, alors v_j et $v_{j'}$ ont même signe.

P2.c Conclure que $V = |V|$ ou $V = -|V|$.

P3. Montrer que $V = |V|$ ou $V = -|V|$ dans le cas général où N est un entier quelconque vérifiant $N \geq 2$.

Partie I Google et PageRank

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 $d_j = 0$ pour chaque couple de pages (i, j) posons $A(i, j) = 0$ si $i \neq j$ et $A(j, j) = 1$. Lorsque $d_j > 0$, posons soit $A(i, j) = 1/d_j$ si j pointe vers i soit $A(i, j) = 0$ sinon. Si $\rho \in [0, 1[$, on définit la matrice de Google $G = (G(i, j))_{1 \leq i \leq N, 1 \leq j \leq N}$ par

$$G(i, j) = \rho A(i, j) + \frac{(1-\rho)}{N}, \text{ pour tout } 1 \leq i \leq N \text{ et } 1 \leq j \leq N. \tag{D}$$

Pour tout $j \in \{1, ..., N\}$, le PageRank d'une page j est un nombre réel positif ou nul noté $p(j)$. Les $p(j)$, $1 \leq j \leq N$ sont par ailleurs définis par le système d'équations

$$\sum_{j=1}^{N} p(j) = 1 \text{ et } \sum_{j=1}^{N} G(i, j) p(j) = p(i) \text{ pour tout } 1 \leq i \leq N. \tag{S}$$

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. Le but 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 (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. Etude de la matrice G de Google

On démontre dans cette section quelques propriétés simples de la matrice G de Google.

I.A.1 Montrer que $G \in \mathcal{M}_N(\mathbb{R})$ est une matrice strictement positive.

I.A.2 Soit $j \in \{1, ..., N\}$ tel que $d_j = 0$, montrer que $\sum_{i=1}^{N} G(i, j) = 1$.

I.A.3 Soit $j \in \{1, ..., N\}$ tel que $d_j > 0$. En écrivant que

$$\sum_{i=1}^{N} G(i,j) = \sum_{\substack{i \in \{1,...,N\} \\ j \text{ pointe vers } i}} G(i,j) + \sum_{\substack{i \in \{1,...,N\} \\ j \text{ ne pointe pas vers } i}} G(i,j)$$

montrer que $\sum_{i=1}^{N} G(i,j) = 1$.

I.A.4 Que peut-on en déduire pour G ?

I.A.5 Montrer que le vecteur $\begin{pmatrix} p(1) \\ \cdot \\ \cdot \\ p(N) \end{pmatrix} \in \mathcal{M}_{N,1}(\mathbb{R})$ défini en (S), en admettant qu'il existe, est invariant par G.

B. Modèle du surfeur sur le Web

Dans toute cette partie $(\Omega, \mathcal{F}, P)$ 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 clic sur un lien dans une page soit d'abandon au profit d'une autre adresse). Pour tout $n \in \mathbb{N}$, X_n est à valeurs dans $\{1, 2, ..., N\}$ et on admettra que la suite de variables aléatoires $(X_n)_{n \geq 0}$vérifie pour tout entier $n \geq 1$ et tout $(x_{n-1}, x_n) \in \{1, 2, ..., N\}^2$

$$P_{\{X_{n-1}=x_{n-1}\}}(X_n = x_n) = P(X_n = x_n \mid X_{n-1} = x_{n-1}) = G(x_n, x_{n-1})$$

où G est la matrice de Google. On considère n un entier naturel strictement positif.

I.B.1 On note V_n le vecteur de $\mathcal{M}_{N,1}(\mathbb{R})$ dont pour tout $i \in \{1, 2, ..., N\}$, la $i^{\text{ème}}$ composante est définie par $(v_n)_i = P(X_n = i)$. Vérifier que V_n est bien un vecteur de probabilité.

I.B.2 Exprimer pour tout $(i,j) \in \{1, 2, ..., N\}^2$ $P(\{X_n = i\} \cap \{X_{n-1} = j\})$ à l'aide de G et V_{n-1}.

I.B.3 En déduire que $V_n = GV_{n-1}$.

I.B.4 Montrer que pour tout k entier naturel $V_k = G^k V_0$.

Partie II Matrices stochastiques

Le but de cette deuxième partie est de prouver l'existence et l'unicité d'un vecteur de probabilité invariant pour une matrice stochastique strictement positive.

A. Etude d'un exemple

On considère une matrice $Q \in \mathcal{M}_2(\mathbb{R})$ stochastique et strictement positive. Q peut se mettre sous la forme

$$Q = \begin{pmatrix} 1-q & q' \\ q & 1-q' \end{pmatrix}$$

avec $q \in]0, 1[$ et $q' \in]0, 1[$.

II.A.1 Déterminer l'ensemble des vecteurs $V \in \mathcal{M}_{2,1}(\mathbb{R})$ vérifiant $QV = V$.

II.A.2 Montrer qu'il existe un unique vecteur de probabilité $V_\infty \in \mathcal{M}_{2,1}(\mathbb{R})$ invariant par Q et le calculer.

II.A.3 Prouver que pour tout entier $n \geq 1$

$$Q^n = \frac{1}{q+q'}\begin{pmatrix} q' & q' \\ q & q \end{pmatrix} + \frac{(1-q-q')^n}{q+q'}\begin{pmatrix} q & -q' \\ -q & q' \end{pmatrix}.$$

II.A.4 En déduire que $(Q^n)_{n\geq 1}$ converge lorsque n tend vers l'infini vers une matrice dont les deux vecteurs colonnes sont égaux à V_∞.

On cherchera à généraliser ce résultat dans la partie III.

B. Existence d'un vecteur de probabilité invariant

Dans cette section II.B, $Q \in \mathcal{M}_N(\mathbb{R})$ est une matrice stochastique.

II.B.1 On note U le vecteur élément de $\mathcal{M}_{N,1}(\mathbb{R})$ dont toutes les composantes valent 1. Calculer ${}^tQU$.

II.B.2 Montrer que si $Q - I_N$ est inversible, alors ${}^tQ - I_N$ est aussi inversible.

II.B.3 Déduire des deux questions précédentes que 1 est valeur propre de Q.

Soit λ une valeur propre réelle de Q telle que $|\lambda| = 1$ et $V \in \mathcal{M}_{N,1}(\mathbb{R})$ un vecteur propre de Q associé à la valeur propre λ.

II.B.4 Prouver que le vecteur $Q|V| - |V|$ est positif.

II.B.5 Montrer que le vecteur $|V|$ est invariant par Q.

Indication : on pourra sommer les composantes du vecteur $Q|V| - |V|$.

II.B.6 Prouver l'existence d'au moins un vecteur de probabilité invariant par Q.

C. Unicité d'un vecteur de probabilité invariant

Dans cette section II.C, $Q \in \mathcal{M}_N(\mathbb{R})$ est une matrice stochastique strictement positive. On sait d'après II.B.6 qu'il existe au moins un vecteur de probabilité invariant par Q noté

$$V_\infty = \begin{pmatrix} (v_\infty)_1 \\ \cdot \\ \cdot \\ (v_\infty)_N \end{pmatrix}.$$

Cette section II.C permettra de démontrer l'unicité d'un tel vecteur.

II.C.1 Prouver que si V est un vecteur positif invariant par Q, alors soit $V = 0_{\mathcal{M}_{N,1}(\mathbb{R})}$ soit V est strictement positif.

II.C.2 Justifier que V_∞ est strictement positif.

On considère à présent un autre vecteur de probabilité noté

$$W_\infty = \begin{pmatrix} (w_\infty)_1 \\ \cdot \\ \cdot \\ (w_\infty)_N \end{pmatrix}$$

invariant par Q. Puis, on définit

$$\begin{aligned} \alpha &= \min\left\{ \frac{(w_\infty)_i}{(v_\infty)_i} \mid 1 \leq i \leq N \right\} = \frac{(w_\infty)_{i_0}}{(v_\infty)_{i_0}} \text{ et} \\ V &= W_\infty - \alpha V_\infty. \end{aligned}$$

II.C.3 Montrer que V est invariant par Q.

II.C.4 Montrer que V est positif mais pas strictement positif.

II.C.5 En déduire que $W_\infty = \alpha V_\infty$.

II.C.6 En conclure que $W_\infty = V_\infty$.

On reprend jusqu'à la fin de cette partie les notations de la partie I sur Google et la notion de PageRank.

II.C.7 Montrer que le système (S) définissant le PageRank admet bien une et une seule solution.

II.C.8 Démontrer que $p(i) \geq (1-\rho)/N$ pour tout $i \in \{1, ..., N\}$.

II.C.9 Le rôle du paramètre ρ est essentiel pour assurer l'unicité de la solution du système (S). Que se passerait-il pour $\rho = 1$ et disons $N = 3$ pour simplifier à l'extrême? (Songer qu'il peut exister des pages Web qui ne pointent vers aucune autre!).

Partie III Validation du PageRank

Soit $Q \in \mathcal{M}_N(\mathbb{R})$ une matrice stochastique strictement positive. On admettra dans cette partie III que $(Q^n)_{n\geq 1}$ converge vers une matrice Q_∞.

A. Etude de Q_∞

Il s'agit dans cette partie d'obtenir une généralisation du résultat de la partie II.A.

III.A.1 Prouver que pour tout $n \geq 1$, Q^n est stochastique puis que Q_∞ est stochastique.

III.A.2 Démontrer que $QQ_\infty = Q_\infty$.

III.A.3 En déduire que chacun des vecteurs colonnes de Q_∞ est invariant par Q. En déduire que Q_∞ est une matrice dont les vecteurs colonnes sont tous égaux à V_∞ où V_∞ est l'unique vecteur de probabilité invariant par Q.

B. Application au modèle du surfeur

On reprend, pour la fin du sujet, les notations du PageRank de Google décrit dans l'introduction de la partie I et du modèle du surfeur décrit dans la partie I.B. On considère X_∞ la variable aléatoire à valeurs dans $\{1, ..., N\}$ définie par

$$P(X_\infty = i) = p(i) \text{ pour tout } i \in \{1, 2, ..., N\}$$

III.B.1 Montrer que pour tout $i \in \{1, ..., N\}$

$$\lim_{n\to+\infty} P(X_n = i) = P(X_\infty = i).$$

III.B.2 En quoi ce résultat donne-t-il du sens à l'assertion un peu vague: "le PageRank d'une page donnée représente la chance qu'un internaute se retrouve sur la page en question lorsqu'il surfe"?