Exercice 1: ★

Déterminer une probabilité P sur $[\![1,n]\!]$ telle que pour tout k, $P(\{k\})$ soit proportionnelle à k.

Exercice 2: ★★

On prélève au hasard un échantillon de k pièces dans une production totale de N pièces comprenant en tout n pièces défectueuses. Pour tout $j \leq k$, on note A_j l'évènement "il y a exactement j pièces défectueuses dans l'échantillon". Calculer la probabilité de A_j.

Exercice 3: ★

Soit (Ω, P) un espace probabilisé.

(1) Soient $A, B \in \mathscr{P}(\Omega)$. Montrer que $P(A\cap B) \geq P(A) + P(B) - 1$.

(2) Soient $A_1, \dots, A_n \in \mathscr{P}(\Omega)$. Montrer que

$$P(A_1 \cap A_2 \cap \cdots \cap A_n) \geq \sum_{i=1}^{n} P(A_i) - (n-1).$$

Exercice 4: ★

Dans une classe de n élèves, on note A_n l'évènement "Deux élèves au moins ont la même date d'annivesaire" (on supposera que personne n'est né un 29 février). Exprimer la probabilité de A_n en fonction de n. À partir de combien d'élèves cette probabilité devient-elle supérieure à 0,5 ? supérieure à 0,95 ?

Exercice 5: ★

On s'intéresse à la répartition des sexes des enfants d'une famille de n enfants. On prend comme modélisation

$$\Omega_n = \{f, g\}^n$$

muni de l'équiprobabilité. On considère les évènements :

— A : "la famille a des enfants des deux sexes",
— B : "la famille a au plus une fille".

(1) Montrer que pour $n \geq 2$, $P(A) = \dfrac{2^n - 2}{2^n}$ et $P(B) = \dfrac{n+1}{2^n}$.

(2) En déduire que A et B ne sont indépendants que si $n = 3$.

Exercice 6: ★★

Un avion a disparu et la région où il s'est écrasé est divisée pour sa recherche en trois zones de même probabilité. Pour $i = 1, 2, 3$, notons $1 - \alpha_i$ la probabilité que l'avion soit retrouvé par une recherche dans la zone i s'il est effectivement dans cette zone. Les constantes α_i représentent les probabilités de manquer l'avion et sont généralement attribuables à l'environnement de la zone (relief, végétation,...). On notera A_i l'évènement "l'avion est dans la zone i", et R_i l'évènement "l'avion est retrouvé dans la zone i".

(1) Pour $i = 1, 2, 3$, déterminer les probabilités que l'avion soit dans la zone i sachant que la recherche dans la zone i a été infructueuse.

(2) Étudier brièvement les variations de ces probabilités conditionnelles considérées comme fonctions de α_i et commenter les résultats obtenus.

Exercice 7: ★★★

Dans une station de métro est disposée une rangée de 16 sièges. Un sociologue observe la répartition des voyageurs assis sur cette rangée. Il représente son observation par une chaîne de caractères 'O' (pour occupé) et 'L' (pour libre) :

OLOLLOLOLLLOLOLO

Notre sociologue se demande si les voyageurs ont choisi leur place au hasard parmi les places disponibles ou s'ils ont cherché à s'isoler. Si les voyageurs choisissent au hasard parmi les places disponibles, toutes les répartitions des 7 personnes sur les 16 places sont équiprobables.

Pour se faire une idée, on compte les *séries*. Une série est ici un bloc de lettres identiques encadré (sauf s'il est en début ou en fin) par des lettres différentes. Il y a donc deux types de séries : les séries de O et les séries de L. Dans l'observation réalisée, il y a 13 séries, dont 7 séries de O.

(1) Quelle est la probabilité qu'une chaîne comportant m lettres L et n lettres O construite au hasard comporte exactement r séries de O ?

(2) Pouvez-vous aider le sociologue ?

Exercice 8: ★★★

Un train se compose de dix wagons citernes contenant un produit dangereux. Chacun des wagons peut, avec une probabilité de 0,1 et indépendament des autres, avoir un défaut. Avant le départ, les wagons sont examinés par deux contrôleurs. Chacun d'eux vérifie tous les wagons, sans échanger d'information avec son collègue pendant le contrôle. On admet que chaque contrôleur peut déceler le défaut (s'il y en a un) d'un wagon donné avec une probabilité de 0,7. Un seul défaut suffit pour que le train soit retardé. Trouver les probabilités des évènements suivants :

(1) Le train est retardé.

(2) Le train part avec au moins un wagon défectueux.

Exercice 9: ★★★

Combien de fois faut-il lancer un dé équilibré pour avoir au moins une chance sur deux d'obtenir un 6 ?

Exercice 10: ★★★

On considère N coffres. Avec une probabilité p, un trésor a été placé dans l'un de ces coffres, chaque coffre pouvant être choisi de façon équiprobable. On a ouvert $N-1$ coffres sans trouver le trésor. Quelle est la probabilité pour qu'il figure dans le dernier coffre ?

Exercice 11: ★★★

On considère $N+1$ urnes numérotées de 0 à N. L'urne k contient k boules blanches et $N-k$ boules noires.

On choisit une urne au hasard, chaque choix étant équiprobable. Dans l'urne choisie, on tire des boules avec remise.

(1) Quelle est la probabilité que la $(n+1)$-ème boule tirée soit blanche sachant que les n précédentes l'étaient toutes.

(2) Que devient cette probabilité quand $N \longrightarrow +\infty$?

Exercice 12: ★★

Une succession d'individus $A_1, \dots, A_n$ se transmet une information binaire du type “oui” ou “non”.

Chaque individu A_k transmet l'information qu'il a reçu avec la probabilité p à l'individu A_{k+1} ou la transforme en son inverse avec la probabilité $q = 1-p$.

Calculer la probabilité p_n pour que l'information reçue par A_n soit identique à celle émise par A_1.
On suppose $0 < p < 1$. Quelle est la limite de (p_n) ?

Exercice 13: ★★★

Soit n un entier naturel supérieur ou égal à 2. On considère la propbabilité P uniforme sur $\Omega = [\![1, n]\!]$.
Pour un entier p divisant n, on introduit l'événement $A_p = \{k \in \Omega \mid p \text{ divise } k\}$.

(1) Calculer $P(A_p)$.

(2) Soient p et q deux diviseurs de n. On suppose que p et q sont premiers entre eux. Montrer que les événements A_p et A_q sont indépendants.

(3) On note $B = \{k \in [\![1, n]\!] \mid k \wedge n = 1\}$. Montrer que

$$P(B) = \prod_{p \text{ diviseur premier de } n} \left(1 - \frac{1}{p}\right).$$

Exercice 14: ★★★★

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

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

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

— On notera I_N la matrice *identité* de $M_N(\mathbb{R})$.

— Pour $A = (A_{i,j}) \in M_n(\mathbb{R})$, on note ${}^tA = (A_{j,i})$ la matrice transposée de A.

On **admet** le résultat suivant :

$$\forall A, B \in M_N(\mathbb{R}),\ {}^t(AB) =^t B^tA,\ {}^t(A + B) =^t A +^t B.$$

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 strictement positifs.

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

— Une matrice $Q = (Q_{i,j})_{1 \leq i,j \leq n} \in M_N(\mathbb{R})$ 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 $V \in M_{N,1}(\mathbb{R})$ est dit *invariant* par $Q \in M_N(\mathbb{R})$ si $QV = V$.

— Soit $Q = (Q_{i,j}) \in M_N(\mathbb{R})$. La suite de matrices (Q^n) est *convergente vers la matrice* Q_∞ si pour tout i, j, la suite $\left((Q^n)_{i,j}\right)_n$ converge vers $(Q_\infty)_{i,j}$.

Préliminaires

Soit $V = \begin{pmatrix} v_1 \\ \vdots \\ v_N \end{pmatrix} \in M_{N,1}(\mathbb{R})$ tel que $\left|\sum_{j=1}^{N} v_j\right| = \sum_{j=1}^{N} |v_j| \quad (E)$.

(1) Montrer que pour tout réel x, $|x| - x \geq 0$.

(2) Étude du cas $N = 3$. On supposera dans cette question uniquement $N = 3$.

(a) Montrer que $(|v_1|+|v_2|+|v_3|)^2-(v_1+v_2+v_3)^2 = 2(|v_1v_2|-v_1v_2)+2(|v_1v_3|-v_1v_3)+2(|v_2v_3|-v_2v_3)$.

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

(c) Conclure que $V = |V|$ ou $V = -|V|$.

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

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 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-ième page pointe ; lorsque $d_j = 0$, on pose pour tout $i \neq j$, $A_{i,j} = 0$ et $A_{j,j} = 1$; lorsque $d_j > 0$, on pose $A_{i,j} = 1/d_j$ si j pointe vers i, 0 sinon.

Soit $\rho \in [0, 1[$, on définit la matrice de Google $G = (G_{i,j})$ en posant

$$\forall i, j,\ G_{i,j} = \rho A_{i,j} + \frac{1-\rho}{N}.$$

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 $(1 \leq j \leq N)$ étant définis par le système d'équations

$$\sum_{j=1}^{N} p_j = 1 \text{ et } \forall i \in [\![1, N]\!],\ \sum_{j=1}^{N} G_{i,j} p_j = p_i. \qquad (S)$$

À la question “que mesure exactement pour une page j donnée ce fameux PageRank p_j ?”, ses 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 justifier d'une part l'existence et l'unicité de la solution du système (S) et de fournir d'autre part une interprétation probabiliste de ce système permettant de donner un sens mathématique aux affirmations de Brin et Page.

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

(4) Montrer que G est une matrice strictement positive.

(5) Soit $j \in [\![1, N]\!]$ tel que $d_j = 0$, montrer que $\displaystyle\sum_{i=1}^{N} G_{i,j} = 1$.

(6) Soit $j \in [\![1, N]\!]$ tel que $d_j > 0$. En écrivant

$$\sum_{i=1}^{N} G_{i,j} = \sum_{\substack{i \in [\![1, N]\!] \\ j \text{ pointe vers } i}} G_{i,j} + \sum_{\substack{i \in [\![1, N]\!] \\ j \text{ ne pointe pas vers } i}} G_{i,j}$$

montrer que $\displaystyle\sum_{i=1}^{N} G_{i,j} = 1$.

(7) Que peut-on en déduire pour G ?

(8) Montrer que le vecteur $\begin{pmatrix} p_1 \\ \vdots \\ p_N \end{pmatrix}$, en admettant qu'il existe, est invariant par G.

Modèle d'un surfeur sur le Web. Dans toute cette partie, (Ω, P) désigne un espace probabilisé. On considère un internaute surfant sur le Web en utilisant Google. Pour tout $n \in \mathbb{N}$ et tout $j \in [\![1, N]\!]$, on note A_n^j l'événement “la n-ième page visitée par le surfeur est la page j”.

On modélise le comportement du surfeur en supposant que

$$\forall n \in \mathbb{N}^*, \forall i, j \in [\![1, N]\!],\ P(A_n^i | A_{n-1}^j) = G_{i,j}.$$

(9) On note $V_n = \begin{pmatrix} P(A_n^1) \\ P(A_n^2) \\ \vdots \\ P(A_n^N) \end{pmatrix} \in M_{N,1}(\mathbb{R})$. Montrer que V_n est bien un vecteur de probabilités.

(10) Exprimer, pour tout $(i, j) \in [\![1, N]\!]$, $P(A_n^i \cap A_{n-1}^j)$ à l'aide de G et V_{n-1}.

(11) En déduire que $V_n = GV_{n-1}$.

(12) Montrer que pour tout $k \in \mathbb{N}$, $V_k = G^k V_0$.

Matrices stochastiques

Étude d'un exemple. On considère une matrice $Q \in M_2(\mathbb{R})$ stochastique et strictement positive. La matrice Q peut se mettre sous la forme

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

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

(13) Déterminer l'ensemble des vecteurs $V \in M_{2,1}(\mathbb{R})$ vérifiant $QV = V$.

(14) Montrer qu'il existe un unique vecteur de probabilité V_∞ invariant par Q et le calculer.

(15) Prouver que pour tout entier $n \geq 1$,

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

(16) En déduire que la suite (Q^n) converge vers la matrice dont les deux colonnes sont égales à V_∞.

Existence d'un vecteur de probabilité invariant. Dans cette section, $Q \in M_N(\mathbb{R})$ est une matrice stochastique.

(17) On note U le vecteur colonne de $M_{N,1}(\mathbb{R})$ dont toutes les composantes valent 1. Calculer ${}^tQU$.

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

(19) Déduire des deux questions précédentes qu'il existe un vecteur $X \in M_{N,1}(\mathbb{R})$ tel que $QX = X$.

Soit $V \in M_{N,1}(\mathbb{R})$ tel que $QV = V$.

(20) Montrer que $Q|V| - |V|$ est positif.

(21) Montrer que le vecteur $|V|$ est invariant par Q. (On pourra sommer les composantes du vecteur $Q|V| - |V|$.)

(22) Prouver l'existence d'au moins un vecteur de probabilité invariant par Q.

Unicité d'un vecteur de probabilité invariant. Dans cette section, $Q \in M_N(\mathbb{R})$ est une matrice stochastique strictement positive. On sait d'après (22) qu'il existe au moins un vecteur de probabilité invariant par Q noté $V = \begin{pmatrix} v_1 \\ \vdots \\ v_N \end{pmatrix}$.

(23) Justifier que V est strictement positif.

On considère à présent un autre vecteur de probabilité $W = \begin{pmatrix} w_1 \\ \vdots \\ w_N \end{pmatrix}$ invariant par Q. On pose $\alpha = \min\{w_i/v_i \mid i \in [\![1, N]\!]\} = w_k/v_k$ et $U = W - \alpha V$.

(24) Montrer que U est invariant par Q.

(25) Montrer que U est positif mais pas strictement positif.

(26) En déduire que $W = \alpha V$.

(27) Conclure que $W = V$.

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

(29) Démontrer que $p_i \geq (1-\rho)/N$ pour tout $i \in [\![1, N]\!]$.

(30) 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 $N = 3$?

Validation du PageRank

Dans toute cette partie, on admet que la suite de matrices (G^n) (où G est la matrice de Google) converge vers la matrice G_∞ dont toutes les colonnes sont égales à $V = \begin{pmatrix} p_1 \\ \vdots \\ p_N \end{pmatrix}$.

(31) Montrer que pour tout $i \in [\![1, N]\!]$, $\lim\limits_{n \to +\infty} P(A_n^i) = p_i$.

(32) En quoi ce résultat donne-t-il un sens à l'assertion un peu vague "le PageRank d'un epage donnée représente la chance qu'un internaute se retrouve sur la page en question lorsqu'il surfe" ?