Terminale Maths expertes

Graphes et chaînes de Markov

Cours

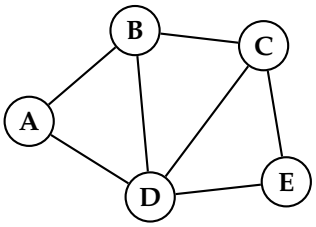
Un graphe représente des objets (les sommets) et les liens entre eux (les arêtes) : réseau routier, réseau social, pages web, états d'un système… Les matrices du chapitre 5 permettent de compter les trajets dans un graphe et d'étudier les **chaînes de Markov**, qui décrivent des évolutions aléatoires.

I. Graphes

1. Vocabulaire

Définition 1 : graphe

Un **graphe** (non orienté) est formé d'un ensemble fini de points, les **sommets**, et de liens entre certains d'entre eux, les **arêtes**.
• L'**ordre** du graphe est son nombre de sommets.
• Deux sommets reliés par une arête sont **adjacents**. Une arête qui relie un sommet à lui-même est une **boucle**.
• Le **degré** d'un sommet est le nombre d'arêtes qui en partent (une boucle compte deux fois).
• Un graphe est **complet** si deux sommets distincts quelconques sont toujours adjacents.



graphe G

Exemple

Le graphe G est d'ordre 5 ; il a 7 arêtes.
A et B sont adjacents ; A et C ne le sont pas.
Degrés : $\deg(A) = 2$, $\deg(B) = 3$, $\deg(C) = 3$, $\deg(D) = 4$, $\deg(E) = 2$.
G n'est pas complet (A et C ne sont pas reliés).

Propriété 1 : lemme des poignées de main

Dans un graphe, la somme des degrés des sommets est égale au double du nombre d'arêtes.
En particulier, le nombre de sommets de degré impair est pair.

Démonstration

Somme des degrés. Lorsqu'on additionne les degrés, chaque arête est comptée deux fois : une fois pour chacune de ses extrémités (une boucle est comptée deux fois pour son unique sommet). La somme vaut donc $2 \times$ (nombre d'arêtes).

Sommets de degré impair. La somme des degrés est paire. La somme des degrés pairs est paire, donc la somme des degrés impairs est paire aussi. Une somme de k nombres impairs a la parité de k : donc k est pair.

Méthode 1 : utiliser la somme des degrés

Énoncé. a. Combien d'arêtes a le graphe complet à n sommets ? **b.** Sept joueurs peuvent-ils organiser un tournoi où chacun joue exactement trois matchs ?

Solution. a. Chaque sommet est relié aux $n-1$ autres : la somme des degrés vaut $n(n-1)$, donc il y a $\dfrac{n(n-1)}{2}$ arêtes. Par exemple, 10 personnes qui se serrent toutes la main échangent 45 poignées de main.

b. On représente les joueurs par des sommets et les matchs par des arêtes. Il y aurait 7 sommets de degré 3 : un nombre impair de sommets de degré impair. C'est impossible d'après la propriété 1.

2. Chaînes et connexité

Définition 2 : chaîne, cycle, connexité

• Une **chaîne** est une suite de sommets dans laquelle deux sommets consécutifs sont adjacents. Sa **longueur** est le nombre d'arêtes parcourues (une même arête peut être parcourue plusieurs fois).
• Une chaîne est **fermée** si son premier et son dernier sommet sont égaux. Un **cycle** est une chaîne fermée dont les arêtes sont toutes distinctes.
• Un graphe est **connexe** si deux sommets quelconques peuvent toujours être reliés par une chaîne.

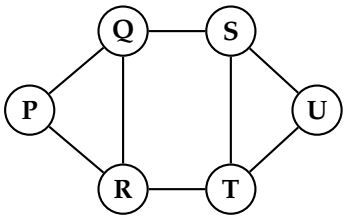
Exemple : dans le graphe G

A – B – C – E est une chaîne de longueur 3 ; A – D – B – D est une chaîne de longueur 3 (l'arête D–B est parcourue deux fois) ; B – C – E – D – B est un cycle de longueur 4. Le graphe G est connexe.

Définition 3 : graphe orienté, graphe pondéré

• Dans un graphe **orienté**, les liens sont des **arcs**, parcourus dans un seul sens et représentés par des flèches. On parle alors de **chemin** (au lieu de chaîne) et de **circuit** (au lieu de cycle).
• Un graphe est **pondéré** si chaque arête (ou chaque arc) porte un nombre, appelé **poids** : une distance, une durée, un coût, une probabilité…

Exercice d'application 1 : vocabulaire des graphes



On considère le graphe ci-contre.

a. Donner son ordre, le degré de chaque sommet et son nombre d'arêtes. Vérifier la propriété 1.

b. Ce graphe est-il complet ? connexe ?

……

……

……

c. Donner une chaîne de longueur 3 reliant P à U, puis un cycle de longueur 4.

……

……

d. Existe-t-il un graphe à 5 sommets dont les degrés sont 1, 2, 2, 3 et 3 ? Justifier.

……

……

II. Matrice d'adjacence

Définition 4 : matrice d'adjacence

Soit un graphe d'ordre n dont les sommets sont numérotés de 1 à n. Sa **matrice d'adjacence** est la matrice carrée $A = (a_{ij})$ d'ordre n où a_{ij} est :
• pour un graphe non orienté, le nombre d'arêtes reliant les sommets i et j ;
• pour un graphe orienté, le nombre d'arcs allant du sommet i vers le sommet j.

Exemple : le graphe G, sommets rangés dans l'ordre A, B, C, D, E

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

La matrice d'un graphe non orienté est **symétrique** ($a_{ij} = a_{ji}$) ; sans boucle, sa diagonale est nulle. La somme des coefficients de la ligne i est le degré du sommet i.

Théorème 1 : nombre de chaînes de longueur n

Soit A la matrice d'adjacence d'un graphe et $n \geqslant 1$ un entier. Le coefficient d'indice (i, j) de A^n est égal au nombre de chaînes (de chemins, si le graphe est orienté) de longueur n allant du sommet i au sommet j.

Démonstration

On note $a_{ij}^{(n)}$ le coefficient d'indice (i, j) de A^n et on raisonne par récurrence sur n.

Initialisation. Une chaîne de longueur 1 de i à j est une arête de i à j : il y en a $a_{ij} = a_{ij}^{(1)}$.

Hérédité. Supposons la propriété vraie pour un entier $n \geqslant 1$. Une chaîne de longueur $n+1$ de i à j se décompose de façon unique en une chaîne de longueur n de i à un sommet k (l'avant-dernier sommet), suivie d'une arête de k à j.

Pour un sommet k fixé, il y a $a_{ik}^{(n)}$ chaînes de longueur n de i à k (hypothèse de récurrence) et a_{kj} arêtes de k à j, donc $a_{ik}^{(n)} \times a_{kj}$ chaînes de ce type. En additionnant sur tous les sommets k :

$$\text{nombre de chaînes de longueur } n+1 \text{ de } i \text{ à } j = \sum_k a_{ik}^{(n)}\, a_{kj} = a_{ij}^{(n+1)},$$

car $A^{n+1} = A^n \times A$ (définition du produit). La propriété est vraie au rang $n+1$.

Conclusion. Elle est vraie pour tout entier $n \geqslant 1$. (La démonstration est la même pour un graphe orienté.)

Méthode 2 : compter des chaînes avec une puissance de la matrice d'adjacence

Énoncé. Dans le graphe G, combien y a-t-il de chaînes de longueur 2 de A à C ? de longueur 3 de A à D ?

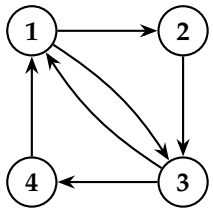
Solution. On calcule (à la main ou à la calculatrice) :

$$A^2 = \begin{pmatrix} 2 & 1 & 2 & 1 & 1 \\ 1 & 3 & 1 & 2 & 2 \\ 2 & 1 & 3 & 2 & 1 \\ 1 & 2 & 2 & 4 & 1 \\ 1 & 2 & 1 & 1 & 2 \end{pmatrix} \quad \text{et} \quad A^3 = \begin{pmatrix} 2 & 5 & 3 & 6 & 3 \\ 5 & 4 & 7 & 7 & 3 \\ 3 & 7 & 4 & 7 & 5 \\ 6 & 7 & 7 & 6 & 6 \\ 3 & 3 & 5 & 6 & 2 \end{pmatrix}.$$

Le coefficient $(1, 3)$ de A^2 vaut 2 : deux chaînes de longueur 2 de A à C, A – B – C et A – D – C.

Le coefficient $(1,4)$ de A^3 vaut 6 : six chaînes de longueur 3 de A à D, à savoir A – B – A – D, A – B – C – D, A – D – A – D, A – D – B – D, A – D – C – D et A – D – E – D.
Les coefficients diagonaux de A^2 redonnent les degrés : une chaîne de longueur 2 de i à i est un aller-retour sur une arête.

Exercice d'application 2 : chemins dans un graphe orienté



Quatre pages web sont reliées par des liens (graphe orienté ci-contre).

On donne $M^3 = \begin{pmatrix} 2 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 2 & 1 \\ 1 & 0 & 1 & 1 \end{pmatrix}$.

a. Écrire la matrice d'adjacence M (pages dans l'ordre 1, 2, 3, 4).

.....

.....

b. Calculer M^2. Combien y a-t-il de chemins de longueur 2 de la page 3 vers elle-même ? Lesquels ?

.....

.....

.....

c. Combien y a-t-il de chemins de longueur 3 de la page 1 vers elle-même ? Les écrire.

.....

.....

d. Peut-on aller de la page 2 à la page 4 en suivant au plus trois liens ? Justifier à l'aide des matrices.

.....

.....

Pour aller plus loin : chaînes eulériennes

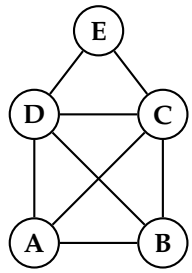
Théorème 2 : théorème d'Euler (admis)

Une chaîne est **eulérienne** si elle passe une fois et une seule par chaque arête du graphe. Soit G un graphe connexe.
• G possède un cycle eulérien si, et seulement si, tous ses sommets sont de degré pair.
• G possède une chaîne eulérienne non fermée si, et seulement si, il a exactement deux sommets de degré impair ; la chaîne commence alors en l'un d'eux et se termine en l'autre.

Remarque : pourquoi la condition est nécessaire

Chaque fois qu'une chaîne eulérienne traverse un sommet, elle utilise deux arêtes : l'une pour arriver, l'autre pour repartir. Un sommet qui n'est ni le départ ni l'arrivée a donc un degré pair.

Exercice d'application 3 : dessiner sans lever le crayon



a. Donner le degré de chaque sommet de la « maison » ci-contre.

b. Peut-on la dessiner d'un seul trait, sans lever le crayon et sans repasser sur un trait ? Si oui, de quel(s) sommet(s) faut-il partir ? Donner un tel tracé.

c. Peut-on le faire en revenant au point de départ ?

.....

.....

.....

.....

III. Chaînes de Markov

1. Définition et matrice de transition

Définition 5 : chaîne de Markov

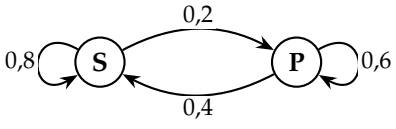
Soit $E = \{1\,;2\}$ ou $E = \{1\,;2\,;3\}$ un ensemble d'**états**. Une suite de variables aléatoires (X_n) à valeurs dans E est une **chaîne de Markov** si, pour tous états i et j, la probabilité de passer de l'état i à l'étape n à l'état j à l'étape $n+1$:
• ne dépend pas des états occupés avant l'étape n (le futur ne dépend du passé qu'à travers le présent) ;
• ne dépend pas de n : on la note $p_{ij} = P_{(X_n=i)}(X_{n+1} = j)$.
La **matrice de transition** de la chaîne est la matrice $P = (p_{ij})$: le coefficient de la ligne i et de la colonne j est la probabilité de passer de i à j. On lui associe un **graphe orienté pondéré** : les sommets sont les états, et un arc de i vers j porte le poids p_{ij} (lorsque $p_{ij} \neq 0$).

Propriété 2 : matrice stochastique

Les coefficients d'une matrice de transition sont positifs et la somme des coefficients de chaque ligne vaut 1.

Démonstration

Ce sont des probabilités, donc ils sont positifs. Sachant que $X_n = i$, la variable X_{n+1} prend exactement une valeur de E : les événements $(X_{n+1} = j)$ forment une partition, donc $\sum_j P_{(X_n=i)}(X_{n+1} = j) = 1$.



Exemple : la météo

Un jour est soit sec (état S), soit pluvieux (état P). Après un jour sec, le lendemain est sec avec la probabilité 0,8 ; après un jour de pluie, le lendemain est sec avec la probabilité 0,4. On note X_n le temps du jour n. La matrice de transition (ordre S, P) est

$$P = \begin{pmatrix} 0{,}8 & 0{,}2 \\ 0{,}4 & 0{,}6 \end{pmatrix}.$$

2. Évolution de la distribution

Définition 6 : distribution

La **distribution** à l'étape n est la matrice ligne π_n qui donne la loi de X_n :
$\pi_n = \big(P(X_n = 1) \;\; P(X_n = 2)\big)$ (ou avec trois coefficients pour trois états). π_0 est la distribution initiale.

Propriété 3 : distribution à l'étape n

Pour tout entier naturel n : $\quad \pi_{n+1} = \pi_n\, P \quad$ et $\quad \pi_n = \pi_0\, P^n$.

Démonstration

Première relation. Les événements $(X_n = 1)$, $(X_n = 2)$ (et $(X_n = 3)$) forment une partition de l'univers. D'après la formule des probabilités totales, pour tout état j :

$$P(X_{n+1} = j) = \sum_i P(X_n = i) \times P_{(X_n=i)}(X_{n+1} = j) = \sum_i P(X_n = i)\, p_{ij}$$

(un terme pour lequel $P(X_n = i) = 0$ est nul). C'est exactement le coefficient j du produit de la ligne π_n par la matrice P : donc $\pi_{n+1} = \pi_n P$.
Seconde relation. Par récurrence : $\pi_0 = \pi_0 P^0$, et si $\pi_k = \pi_0 P^k$, alors $\pi_{k+1} = \pi_k P = \pi_0 P^k P = \pi_0 P^{k+1}$.

Théorème 3 : probabilité de passer de i à j en n étapes

Le coefficient d'indice (i, j) de P^n est la probabilité de passer de l'état i à l'état j en n étapes : $P_{(X_0=i)}(X_n = j)$.

Démonstration

Supposons que la chaîne part de l'état i : π_0 est la matrice ligne dont le coefficient i vaut 1 et les autres 0. D'après la propriété 3, $\pi_n = \pi_0 P^n$; or le produit de cette ligne par P^n est la ligne i de P^n. Son coefficient j est donc $P(X_n = j)$ lorsque $X_0 = i$, c'est-à-dire $P_{(X_0=i)}(X_n = j)$.

Méthode 3 : calculer des distributions successives

Énoncé. Dans l'exemple de la météo, le lundi (jour 0) est sec. Quelle est la probabilité que le mercredi soit sec ?

Solution. $\pi_0 = (1 \;\; 0)$, $\pi_1 = \pi_0 P = (0{,}8 \;\; 0{,}2)$,
$\pi_2 = \pi_1 P = (0{,}8 \times 0{,}8 + 0{,}2 \times 0{,}4 \quad 0{,}8 \times 0{,}2 + 0{,}2 \times 0{,}6) = (0{,}72 \;\; 0{,}28)$.
Le mercredi est sec avec la probabilité 0,72. On peut aussi lire ce nombre dans $P^2 = \begin{pmatrix} 0{,}72 & 0{,}28 \\ 0{,}56 & 0{,}44 \end{pmatrix}$:
c'est le coefficient $(1, 1)$, d'après le théorème 3.

Exercice d'application 4 : abonnés d'une plateforme

Chaque mois, un abonné à une plateforme de vidéos est actif (état A) ou inactif (état I). Un abonné actif le reste le mois suivant avec la probabilité 0,9 ; un abonné inactif redevient actif avec la probabilité 0,3. Au mois 0, la moitié des abonnés sont actifs.

a. Représenter la situation par un graphe orienté pondéré et donner la matrice de transition P.

.....

b. Calculer π_1 et π_2.

.....

.....

c. Déterminer la distribution invariante (voir paragraphe 3) et interpréter.

.....

.....

.....

3. Distribution invariante

Définition 7 : distribution invariante

Une distribution π (matrice ligne à coefficients positifs de somme 1) est **invariante** si $\pi P = \pi$: si la chaîne suit la loi π à une étape, elle la suit encore à l'étape suivante.

Propriété 4 : limite et distribution invariante

Si la suite (π_n) converge vers une matrice ligne π, alors π est une distribution invariante.

Démonstration

Les coefficients de π_n sont positifs et de somme 1 : par passage à la limite, ceux de π aussi. Chaque coefficient de $\pi_n P$ est une somme de termes $P(X_n = i) \times p_{ij}$: il tend vers le coefficient correspondant de πP. En passant à la limite dans $\pi_{n+1} = \pi_n P$, on obtient $\pi = \pi P$.

Théorème 4 : chaîne à deux états

Soit une chaîne de Markov à deux états de matrice $P = \begin{pmatrix} 1-a & a \\ b & 1-b \end{pmatrix}$, avec $0 < a + b < 2$. Alors, quelle que soit la distribution initiale, (π_n) converge vers l'unique distribution invariante

$$\pi = \begin{pmatrix} \dfrac{b}{a+b} & \dfrac{a}{a+b} \end{pmatrix}.$$

Démonstration

On note $\pi_n = (p_n \;\; 1-p_n)$. D'après la propriété 3 :
$p_{n+1} = p_n(1-a) + (1-p_n)\,b = (1-a-b)\,p_n + b$.

Suite arithmético-géométrique. Soit $\ell = \dfrac{b}{a+b}$; on vérifie que $\ell = (1-a-b)\ell + b$. En soustrayant : $p_{n+1} - \ell = (1-a-b)(p_n - \ell)$. La suite $(p_n - \ell)$ est géométrique de raison $q = 1-a-b$, donc $p_n = \ell + q^n(p_0 - \ell)$.

Convergence. Comme $0 < a+b < 2$, on a $-1 < q < 1$, donc $q^n \to 0$: $p_n \to \ell = \dfrac{b}{a+b}$ et $1-p_n \to 1-\ell = \dfrac{a}{a+b}$, quelle que soit la valeur de p_0.

Unicité. Si $(x \;\; 1-x)$ est invariante, la chaîne qui part de cette distribution la garde à chaque étape : la suite (p_n) est constante égale à x ; or elle converge vers ℓ, donc $x = \ell$.

Remarques :

• Dans l'exemple de la météo, $a = 0{,}2$ et $b = 0{,}4$: à long terme, un jour est sec avec la probabilité $\frac{0{,}4}{0{,}6} = \frac{2}{3}$.

• Si $a = b = 1$, $P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$: la chaîne alterne entre les deux états et (π_n) ne converge pas si $\pi_0 = (1 \;\; 0)$.

• **Pour trois états** (admis) : s'il existe une puissance de P dont tous les coefficients sont strictement positifs, la distribution invariante est unique et (π_n) converge vers elle, quelle que soit π_0.

Méthode 4 : déterminer une distribution invariante à trois états

Énoncé. Un robot aspirateur circule entre trois pièces en enfilade A, B, C. Chaque minute, il reste dans sa pièce avec la probabilité $\frac{1}{2}$; sinon il passe dans une pièce voisine, choisie au hasard. Déterminer la distribution invariante.

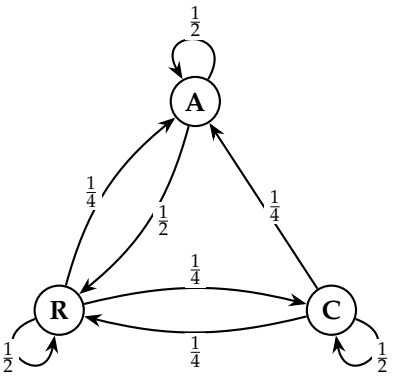
Solution. Depuis A ou C, il n'y a qu'une pièce voisine (B) ; depuis B, deux. $P = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix}$.

$\pi = (x \;\; y \;\; z)$ est invariante si $\pi P = \pi$ et $x+y+z = 1$:

$$\begin{cases} \frac{1}{2}x + \frac{1}{4}y = x \\ \frac{1}{2}x + \frac{1}{2}y + \frac{1}{2}z = y \\ \frac{1}{4}y + \frac{1}{2}z = z \\ x+y+z = 1 \end{cases} \iff \begin{cases} y = 2x \\ y = 2z \\ x+y+z = 1 \end{cases} \iff x = z = \frac{1}{4},\; y = \frac{1}{2}.$$

$\pi = \left(\frac{1}{4} \;\; \frac{1}{2} \;\; \frac{1}{4}\right)$. Comme P^2 a tous ses coefficients strictement positifs, (π_n) converge vers π : à long terme, le robot passe la moitié du temps dans la pièce centrale.

Exercice d'application 5 : une chaîne à trois états



Chaque midi, un étudiant déjeune au restaurant universitaire (R), à la cafétéria (C) ou apporte son repas (A), selon le graphe ci-contre. Le lundi (jour 0), il déjeune au restaurant.

a. Écrire la matrice de transition P (ordre R, C, A).

.....

.....

b. Calculer π_1 et π_2. Quelle est la probabilité qu'il apporte son repas le mercredi ?

.....

.....

c. Déterminer la distribution invariante. En admettant que (π_n) converge, interpréter.

.....

.....

.....

IV. Algorithmique

La fonction `etape` simule une transition : on tire un nombre u au hasard dans $[0\,;1[$ et on parcourt la ligne i de P en cumulant les probabilités. La fonction `frequences` répète la simulation de n étapes pour estimer la loi de X_n ; on peut comparer avec la ligne i de P^n (théorème 3).

Python

```
from random import random
def etape(P, i):
    u = random()
    cumul = 0
    for j in range(len(P)):
        cumul = cumul + P[i][j]
        if u < cumul:
            return j
    return len(P) - 1
```

Python

```
def frequences(P, i, n, essais):
    compte = [0] * len(P)
    for k in range(essais):
        etat = i
        for t in range(n):
            etat = etape(P, etat)
        compte[etat] = compte[etat] + 1
    return [c / essais for c in compte]

P = [[0.8, 0.2], [0.4, 0.6]]
print(frequences(P, 0, 2, 10000))
# environ [0.72, 0.28]
```

L'essentiel du chapitre

Graphes	Ordre, degré, adjacents, complet, chaîne, cycle, connexe ; somme des degrés $= 2\times$ nombre d'arêtes.
Matrice d'adjacence	a_{ij} = nombre d'arêtes (d'arcs) de i à j ; coefficient (i,j) de A^n = nombre de chaînes de longueur n de i à j.
Euler	Graphe connexe : cycle eulérien $\iff$ tous les degrés pairs ; chaîne eulérienne $\iff$ 0 ou 2 sommets de degré impair.
Chaîne de Markov	$p_{ij} = P_{(X_n=i)}(X_{n+1}=j)$; lignes de somme 1 ; $\pi_{n+1} = \pi_n P$, $\pi_n = \pi_0 P^n$.
P^n	Coefficient (i,j) = probabilité de passer de i à j en n étapes.
Invariante	$\pi P = \pi$, coefficients de somme 1 ; deux états : $\pi = \left(\frac{b}{a+b} \quad \frac{a}{a+b}\right)$, limite de (π_n) si $0 < a+b < 2$.