

Ce problème étudie le comportement probabiliste des graphes aléatoires. Dans un premier temps, on introduit des outils généraux permettant de comparer des lois de probabilité, à l'aide de la distance en variation totale et du principe de couplage. Ces notions fournissent un cadre pour mesurer la proximité entre deux lois.

On étudie ensuite un graphe aléatoire, en analysant la proportion de sommets de degré donné. Dans un régime asymptotique approprié, on met en évidence l'émergence d'une loi de Poisson.

La dernière partie propose une illustration numérique de ces résultats à l'aide de courts programmes Python.

Partie I Distance en variation totale

Dans cette partie on considère une partie $\mathcal{K}$ de $\mathbb{N}$, finie ou égale à $\mathbb{N}$. Pour toute partie A de $\mathcal{K}$, on note $\bar{A}$ le complémentaire de A dans $\mathcal{K}$.

On considère également $\mathcal{L} = (p_k)_{k \in \mathcal{K}}$ et $\mathcal{M} = (q_k)_{k \in \mathcal{K}}$ deux lois de probabilités sur $\mathcal{K}$.

Par définition, on a donc :

$$\left\{ \begin{array}{l} (i) \quad \forall k \in \mathcal{K}, p_k \geq 0 \text{ et } \forall k \in \mathcal{K}, q_k \geq 0 \\ (ii) \quad \sum_{k \in \mathcal{K}} p_k = 1 \text{ et } \sum_{k \in \mathcal{K}} q_k = 1 \end{array} \right.$$

A la loi $\mathcal{L}$, on peut associer une unique probabilité P sur $\mathcal{K}$ par :

$$\forall A \in \mathcal{P}(\mathcal{K}), P(A) = \sum_{k \in A} p_k$$

A la loi $\mathcal{M}$, on peut associer une unique probabilité Q sur $\mathcal{K}$ par :

$$\forall A \in \mathcal{P}(\mathcal{K}), Q(A) = \sum_{k \in A} q_k$$

avec en particulier, $\forall k \in \mathcal{K}, p_k = P(\{k\})$ et $q_k = Q(\{k\})$.

Lorsque $\mathcal{K}$ est fini, on définit la *distance en variation totale* entre les lois $\mathcal{L}$ et $\mathcal{M}$ par :

$$D(\mathcal{L}, \mathcal{M}) = \frac{1}{2} \sum_{k \in \mathcal{K}} |p_k - q_k| \quad (*)$$

1) On suppose $\mathcal{K} = \{0, 1\}$. Exprimer $D(\mathcal{L}, \mathcal{M})$ en fonction de p_1 et q_1 .

2) On suppose $\mathcal{K} = \mathbb{N}$. Vérifier que la série $\sum_{k \geq 0} |p_k - q_k|$ est convergente.

On étend donc la définition donnée en $(*)$ au cas où $\mathcal{K} = \mathbb{N}$.

3)a) Montrer : $0 \leq D(\mathcal{L}, \mathcal{M}) \leq 1$.

b) Montrer : $D(\mathcal{L}, \mathcal{M}) = 0 \iff \mathcal{L} = \mathcal{M}$.

4) Vérifier pour toute partie A de $\mathcal{K}$, on a :

$$0 \leq |P(A) - Q(A)| \leq 1.$$

5) Montrer que pour toute partie A de $\mathcal{K}$, on a :

$$2|P(A) - Q(A)| = \left| \sum_{k \in A} (p_k - q_k) \right| + \left| \sum_{k \in \overline{A}} (p_k - q_k) \right|$$

6) En déduire que pour toute partie A de $\mathcal{K}$, on a :

$$|P(A) - Q(A)| \leq D(\mathcal{L}, \mathcal{M}) \quad (**)$$

7) Montrer que la partie $A_0 = \{k \in \mathcal{K} \mid q_k \geq p_k\}$ réalise l'égalité dans (**), c'est à dire que $|P(A_0) - Q(A_0)| = D(\mathcal{L}, \mathcal{M})$

Les questions 6) et 7) montrent donc que

$$D(\mathcal{L}, \mathcal{M}) = \max_{A \subset \mathcal{K}} |P(A) - Q(A)|$$

8) Pour tous réels a et b , on note $\min(a, b)$ le plus petit de ces deux nombres.

a) Etablir : $\min(a, b) = \frac{1}{2}(a + b - |a - b|)$.

b) En déduire la formule : $D(\mathcal{L}, \mathcal{M}) = 1 - \sum_{k \in \mathcal{K}} \min(p_k, q_k)$

9) Soient $S = \{k \in \mathcal{K} \mid p_k > 0\}$ et $T = \{k \in \mathcal{K} \mid q_k > 0\}$.

Montrer : $D(\mathcal{L}, \mathcal{M}) = 1 \iff S \cap T = \emptyset$.

Partie II Couplages

On garde les notations de la partie I.

Soient X et Y deux variables aléatoires discrètes définies sur un espace probabilisé $(\Omega, \mathcal{A}, \mathbf{P})$.

On dit que le couple (X, Y) est un *couplage* de $\mathcal{L}$ et $\mathcal{M}$ si X a pour loi $\mathcal{L}$ et si Y a pour loi $\mathcal{M}$.

Autrement dit, si pour tout $k \in \mathcal{K}$, on a : $\mathbf{P}(X = k) = p_k$ et $\mathbf{P}(Y = k) = q_k$.

On dit que ce couplage est *indépendant* si X et Y sont indépendantes.

On dit que ce couplage est *monotone* si $\mathbf{P}(X \leq Y) = 1$.

1) Quelques exemples de couplage :

a) On considère les lois de probabilité $\mathcal{L} = \mathcal{B}(p)$ et $\mathcal{M} = \mathcal{B}(q)$, où p et q sont dans $[0, 1]$.

Soit (X, Y) un couple de variables aléatoires de loi conjointe donnée par :

Y	0	1
X		
0	$(1-p)(1-q)$	$(1-p)q$
1	$p(1-q)$	pq

Montrer que (X, Y) est un couplage indépendant de $\mathcal{L}$ et $\mathcal{M}$.

b) On considère les lois de probabilité $\mathcal{L} = \mathcal{U}(\{-1, 0, 1\})$ et $\mathcal{M} = \mathcal{B}(p)$ avec $\frac{1}{3} < p < \frac{2}{3}$.

Soit (X, Y) un couple de variables aléatoires de loi conjointe donnée par :

Y	0	1
X		
-1	$\frac{1}{3}$	0
0	$\frac{2}{3} - p$	$p - \frac{1}{3}$
1	0	$\frac{1}{3}$

Montrer que (X, Y) est un couplage monotone de $\mathcal{L}$ et $\mathcal{M}$.

2) Soit (X, Y) un couplage de $\mathcal{L}$ et $\mathcal{M}$. A l'aide de la question I.8)b), montrer :

$$D(\mathcal{L}, \mathcal{M}) \leq \mathbf{P}(X \neq Y).$$

3) Un exemple simple de majoration de la distance en variation totale :

On s'intéresse ici aux lois de probabilité $\mathcal{L} = \mathcal{P}(\alpha)$ et $\mathcal{M} = \mathcal{P}(\beta)$ avec $\alpha > \beta > 0$.

Soient Y et Z deux variables aléatoires indépendantes telles que $Y \hookrightarrow \mathcal{P}(\beta)$ et $Z \hookrightarrow \mathcal{P}(\alpha - \beta)$.

On pose $X = Y + Z$.

a) Justifier que (X, Y) est un couplage de $\mathcal{L}$ et $\mathcal{M}$.

b) Conclure que $D(\mathcal{L}, \mathcal{M}) \leq 1 - e^{-(\alpha - \beta)}$.

Partie III Couplage binomiale - Poisson

Dans la partie II, nous avons vu que la distance en variation totale entre deux lois de probabilité peut être majorée en construisant un couplage adapté de ces deux lois.

Dans cette partie, nous allons appliquer cette méthode à un exemple fondamental : l'approximation de la loi binomiale par la loi de Poisson.

L'objectif sera de construire un couplage de ces deux lois afin d'obtenir une majoration de leur distance en variation totale.

Soit n un entier strictement positif et soit λ un réel strictement positif, strictement plus petit que n .

Notons $\mathcal{L}_n = \mathcal{B}\left(n, \frac{\lambda}{n}\right)$ et $\mathcal{M} = \mathcal{P}(\lambda)$.

Le but est de prouver la majoration suivante :

$$D(\mathcal{L}_n, \mathcal{M}) \leq \frac{\lambda^2}{n} \quad (***)$$

1) Soient $Y_1, \dots, Y_n$ des variables aléatoires mutuellement indépendantes et de même loi de Poisson de paramètre $\frac{\lambda}{n}$.

Rappeler sans démonstration quelle est la loi de $\sum_{i=1}^n Y_i$.

2) Soit f la fonction définie sur $[0, 1]$ par : $f(x) = 1 - (1 - x)e^x$.

Montrer que $\forall x \in [0, 1], f(x) \in [0, 1]$.

Par la suite, on considère $U_1, \dots, U_n$ des variables aléatoires mutuellement indépendantes de même loi de Bernoulli de paramètre $f\left(\frac{\lambda}{n}\right)$.

On suppose que les variables $U_1, \dots, U_n$ sont indépendantes des variables $Y_1, \dots, Y_n$ de la question III.1)

Pour tout $i \in \{1, \dots, n\}$ on pose :

$$X_i = \begin{cases} 0 & \text{si } U_i = 0 \text{ et } Y_i = 0 \\ 1 & \text{sinon} \end{cases}$$

3a) Vérifier que pour tout $i \in \{1, \dots, n\}$, X_i suit la loi de Bernoulli de paramètre $\frac{\lambda}{n}$

b) Montrer que $\mathbf{P}((X_1 = 0) \cap \dots \cap (X_n = 0)) = \mathbf{P}(X_1 = 0) \times \dots \times \mathbf{P}(X_n = 0)$.

c) En déduire que $X_1, \dots, X_n$ sont mutuellement indépendantes, puis

reconnaître la loi de $\sum_{i=1}^n X_i$.

4) Montrer que pour tout $i \in \{1, \dots, n\}$, $\mathbf{P}(X_i \neq Y_i) \leq \frac{\lambda^2}{n^2}$

Indication : on pourra établir que $\forall x \in \mathbf{R}, 1 + x \leq e^x$.

5) Montrer que :

$$\mathbf{P}\left(\sum_{i=1}^n X_i \neq \sum_{i=1}^n Y_i\right) \leq \mathbf{P}\left(\bigcup_{i=1}^n (X_i \neq Y_i)\right)$$

6)a) Montrer pour tout entier $n \geq 1$ et pour tous événements $A_1, \dots, A_n$ l'inégalité de Boole :

$$\mathbf{P}(A_1 \cup \dots \cup A_n) \leq \mathbf{P}(A_1) + \dots + \mathbf{P}(A_n).$$

b) En déduire que :

$$\mathbf{P}\left(\sum_{i=1}^n X_i \neq \sum_{i=1}^n Y_i\right) \leq \frac{\lambda^2}{n}$$

Conclure quant à $(***)$.

7) Quel résultat connu peut-on déduire de $(***)$ lorsque $n \rightarrow +\infty$?

Partie IV Distribution asymptotique des degrés dans le graphe aléatoire d'Erdos-Rényi

Dans cette partie, n désigne un entier supérieur ou égal à 2 et p un réel appartenant à $]0,1[$. On pose $q = 1 - p$.

On considère un graphe aléatoire non orienté $G(n, p)$ construit de la manière suivante : ses n sommets sont numérotés de 1 à n , et pour chaque paire $\{i, j\}$ de sommets distincts, l'arête $\{i, j\}$ est présente avec probabilité p , indépendamment des autres arêtes.

Pour toute couple (i, j) avec $1 \leq i \neq j \leq n$, on considère la variable aléatoire $T_{i,j}$ définie par :

$$T_{i,j} = \begin{cases} 1 & \text{si } \{i, j\} \text{ est une arête} \\ 0 & \text{sinon} \end{cases}$$

Les variables $T_{i,j}$ sont donc indépendantes et suivent toutes une loi de Bernoulli de paramètre p .

Le graphe n'étant pas orienté, on remarquera que $T_{i,j} = T_{j,i}$.

Pour tout $i \in \llbracket 1, n \rrbracket$, on définit la variable aléatoire D_i égale au degré du sommet i .

Pour tout entier d , on définit la variable aléatoire N_d égale au nombre de sommets de degré d .

1) Préliminaire :

Soit $(\Omega, \mathcal{A}, \mathbf{P})$ un espace probabilisé.

Pour tout événement A , on note $\mathbb{1}_A$, la variable aléatoire définie pour tout $\omega \in \Omega$ par :

$$\mathbb{1}_A(\omega) = \begin{cases} 1 & \text{si } \omega \in A \\ 0 & \text{sinon} \end{cases}$$

a) Justifier que $\mathbb{1}_A$ suit la loi de Bernoulli de paramètre $\mathbf{P}(A)$.

b) Montrer que pour tous événements A et B , on a : $\mathbb{1}_A \times \mathbb{1}_B = \mathbb{1}_{A \cap B}$

c) Montrer que pour tout événement A , on a : $(\mathbb{1}_A)^2 = \mathbb{1}_A$

2)a) Montrer que pour tout entier $n \geq 2$ et pour toutes variables aléatoires $Z_1, \dots, Z_n$, on a :

$$(Z_1 + \dots + Z_n)^2 = Z_1^2 + \dots + Z_n^2 + 2 \sum_{1 \leq i < j \leq n} Z_i Z_j.$$

b) Justifier que $N_d = \sum_{i=1}^n \mathbb{1}_{(D_i=d)}$

c) En déduire que $(N_d)^2 = N_d + 2 \sum_{1 \leq i < j \leq n} \mathbb{1}_{(D_i=d) \cap (D_j=d)}$

3)a) Quelle est la loi de N_d lorsque $d > n - 1$?

b) Soit $d \in \llbracket 0, n - 1 \rrbracket$.

Justifier que $N_d(\Omega) \subset \llbracket 0, n \rrbracket$. Y a-t-il égalité de ces ensembles ?

4) Dans cette question, on étudie des propriétés du couple (D_1, D_2) .

On prend $d \in \llbracket 0, n - 1 \rrbracket$ et on pose :

$$S_1 = \sum_{k=3}^n T_{1,k} \quad \text{et} \quad S_2 = \sum_{k=3}^n T_{2,k}$$

a) Justifier que $D_1 = \sum_{k=2}^n T_{1,k} = T_{1,2} + S_1$ et que $D_2 = \sum_{k=2}^n T_{2,k} = T_{1,2} + S_2$.

b) En utilisant les propriétés de bilinéarité de la covariance, montrer que

$$\text{cov}(D_1, D_2) = p(1 - p).$$

c) D_1 et D_2 sont-elles indépendantes ?

d) S_1 et S_2 suivent la même loi. Laquelle ?

e) A l'aide de la formule des probabilités totales pour le système complet $((T_{1,2} = 0), (T_{1,2} = 1))$, établir :

$$\mathbf{P}(D_1 = d) = q\mathbf{P}(S_1 = d) + p\mathbf{P}(S_1 = d - 1).$$

f) Démontrer de même :

$$\mathbf{P}((D_1 = d) \cap (D_2 = d)) = q\mathbf{P}(S_1 = d)^2 + p\mathbf{P}(S_1 = d - 1)^2.$$

g) En déduire l'égalité :

$$\mathbf{P}((D_1 = d) \cap (D_2 = d)) - \mathbf{P}(D_1 = d)^2 = pq[\mathbf{P}(S_1 = d) - \mathbf{P}(S_1 = d - 1)]^2.$$

On admettra par la suite, que pour des raisons de symétrie, toutes les variables D_i ont la même loi, et que tous les couples (D_i, D_j) pour $i \neq j$ ont la même loi.

5) Soit $d \in \llbracket 0, n-1 \rrbracket$.

a) A l'aide des questions IV.1) et IV.2), montrer :

$$E(N_d) = n\mathbf{P}(D_1 = d).$$

b) En appliquant la formule de Koëinig, montrer :

$$V(N_d) = n\mathbf{P}(D_1 = d) + n(n-1)\mathbf{P}((D_1 = d) \cap (D_2 = d)) - n^2\mathbf{P}(D_1 = d)^2.$$

c) En déduire l'inégalité :

$$V\left(\frac{N_d}{n}\right) \leq \frac{1}{n} + \mathbf{P}((D_1 = d) \cap (D_2 = d)) - \mathbf{P}(D_1 = d)^2.$$

d) Conclure des questions précédentes : $V\left(\frac{N_d}{n}\right) \leq \frac{1}{n} + p$.

On suppose dans toute la suite que $\mathbf{p} = \frac{c}{n}$ où c est une constante réelle strictement positive.

6) Soit $d \in \llbracket 0, n-1 \rrbracket$.

a) Vérifier l'égalité : $E\left(\frac{N_d}{n}\right) = \frac{c^d}{d!} \times \frac{(n-1)!}{(n-1-d)!n^d} \times \left(1 - \frac{c}{n}\right)^{n-1-d}$.

b) Montrer que $\lim_{n \rightarrow +\infty} \frac{(n-1)!}{(n-1-d)!n^d} = 1$ et que $\lim_{n \rightarrow +\infty} \left(1 - \frac{c}{n}\right)^{n-1-d} = e^{-c}$.

On vient donc de montrer que $\lim_{n \rightarrow +\infty} E\left(\frac{N_d}{n}\right) = e^{-c} \times \frac{c^d}{d!}$.

7) Dans cette question, on considère $l > 0$ et $\epsilon > 0$ des réels.

$(X_n)_{n \geq 1}$ est une suite de variables aléatoires discrètes.

$(u_n)_{n \geq 1}$ est une suite pour laquelle il existe un entier n_ϵ tel que

$$\forall n \geq n_\epsilon, |u_n - l| \leq \frac{\epsilon}{2}.$$

(i) Etablir l'inclusion : $(|X_n - l| \geq \epsilon) \subset (|X_n - u_n| + |u_n - l| \geq \epsilon)$.

(ii) Pour $n \geq n_\epsilon$, établir l'inclusion : $(|X_n - u_n| + |u_n - l| \geq \epsilon) \subset \left(|X_n - u_n| \geq \frac{\epsilon}{2}\right)$.

8) A l'aide de l'inégalité de Bienaymé-Tchébychev, établir :

$$\forall \epsilon > 0, \exists n_\epsilon \in \mathbf{N}, \forall n \geq n_\epsilon, \mathbf{P}\left(\left|\frac{N_d}{n} - e^{-c} \times \frac{c^d}{d!}\right| \geq \epsilon\right) \leq \frac{4(1+c)}{n\epsilon^2}.$$

Interpréter ce résultat.

9)Python.

On suppose importés le module `numpy` d'alias `np` et le module `numpy.random` d'alias `rd`.

a) Compléter la fonction suivante pour qu'elle retourne la matrice d'adjacence du graphe $G(n, p)$.

```
def graphe_erdos_renyi(n, p):
    M=np.zeros(shape=(n,n))
    for i in range(n):
        for j in range(i+1,n):
            if rd.random()<p:
                M[i,j]=....
                M[j,i]=....
    return M
```

b) Compléter la fonction suivante prenant en entrée la matrice d'adjacence d'un graphe et retournant la liste des degrés des sommets.

```
def degres(M):
    n= .....
    liste=[]
    for i in range(n):
        deg=sum(.....)
        liste.append(deg)
    return liste
```

c) Compléter la fonction suivante prenant en entrée deux entiers n et d , un réel c , et retournant la proportion de sommets de degré d .

```
def proportion(n, c, d):
    p=c/n
    M=graphe_erdos_renyi(n, p)
    liste_degres=degres(M)
    nb=0
    for deg in liste_degres:
        if deg == d :
            nb= .....
    return .....
```

d) Dans la console Python, on tape la commande `proportion(1000,1,2)`. Il en sort la valeur 0,193. Est-ce cohérent avec la question IV.8) ?

On donne $e^{-1} \approx 0,37$.