AP 9

Graphes aléatoires d'Erdös-Renyi
Distance en variation et couplage

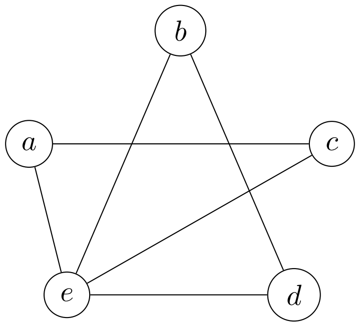
Exercice 1 (Graphes aléatoires d'Erdös-Renyi)

On considère un espace probabilisé $(\Omega, \mathscr{A}, P)$.
Pour définir un graphe aléatoire non orienté G, on se donne :

- $S = \{s_1, \ldots, s_n\}$ un ensemble fini de $n \geq 2$ sommets ;
- pour toute paire de sommets $\{s_i, s_j\}$ avec $i < j$, $T_{i,j}$ une variable de Bernoulli de paramètre p sur $(\Omega, \mathscr{A}, P)$.

Les arêtes du graphe sont les paires $\{s_i, s_j\}$ telles que $T_{i,j} = 1$ avec $i < j$. Les variables $T_{i,j}$ sont supposées indépendantes.

Voici un exemple de graphe aléatoire avec $S = \{a, b, c, d, e\}$ et $p = 0.4$:



1. Quel est le nombre maximal d'arêtes de G ?

2. Écrire une fonction `Python` d'entête `listAdj(S,p)` qui génère la liste des listes d'adjacence d'un tel graphe aléatoire ayant `S` pour liste de sommets.

   Le graphe dessiné ci-dessus correspond par exemple à la liste des listes d'adjacence suivante :

   ```
   [['c', 'e'], ['d', 'e'], ['a', 'e'], ['b', 'e'], ['a', 'b', 'c', 'd']]
   ```

   avec la liste des sommets qui est `S = 'abcde'`.

3. Pour tout $k \in [\![1, n]\!]$, on considère la variable aléatoire D_k égale au degré du sommet s_k. Déterminer la loi de D_k.

4. On dit que s_k est isolé si $D_k = 0$ est réalisé. On note Z_n la variable aléatoire comptant les sommets isolés de G et X_k la variable aléatoire de Bernoulli qui vaut 1 si s_k est isolé et 0 sinon.

   (a) Montrer que : $Z_n = \displaystyle\sum_{k=1}^{n} X_k$. En déduire que : $E(Z_n) = n(1-p)^{n-1}$.

   (b) Montrer que : $Z_n^2 = \displaystyle\sum_{k=1}^{n} X_k + 2 \sum_{1 \leq i < j \leq n} X_i X_j$.

   (c) Justifier que : $P([X_i = 1] \cap [X_j = 1]) = (1-p)^{2n-3}$.
   En déduire que : $E(Z_n^2) = n(1-p)^{n-1} + n(n-1)(1-p)^{2n-3}$.

On suppose désormais que $p = p_n = c\dfrac{\ln(n)}{n}$, avec $c > 0$, $c \neq 1$.

5. (a) Écrire une fonction Python `Z` qui renvoie le nombre de sommets isolés d'un graphe donné par sa liste de listes d'adjacence `lst`. On importera les bibliothèques **`numpy as np`** et **`numpy.random as rd`**.

   (b) On souhaite estimer l'influence de la valeur de c sur le nombre de sommets isolés. En exécutant le script suivant :

```python
list2c = [0.3, 0.5, 0.7, 1.3, 1.5, 1.7]
n = 1000
res = []

for c in list2c:
    s = 0
    for k in range(200):
        if Z(listAdj(range(1,n+1), c*np.log(n)/n))==0:
            s += 1
    res.append(s/200)
print(res)
```

   on obtient `[0.0, 0.0, 0.0, 0.91, 0.975, 0.99]` après de longues minutes.

   Quelle conjecture sur la valeur d'une probabilité pouvez-vous faire lorsque $c < 1$ et $c > 1$? Justifier.

6. (a) Montrer que $(1-p_n)^{n-1} \underset{n\to+\infty}{\sim} (1-p_n)^n$ puis que $(1-p_n)^{n-1} \underset{n\to+\infty}{\sim} n^{-c}$.

   (b) On rappelle que l'inégalité de Markov affirme que si X est une variable aléatoire positive admettant une espérance et $a > 0$,

$$P(X \geq a) \leq \frac{E(X)}{a}.$$

   Si $c > 1$, en déduire la limite de $P(Z_n \geq 1)$ puis de $P(Z_n = 0)$, lorsque $n \to +\infty$.

   (c) On rappelle que l'inégalité de Bienaymé-Tchébichev affirme que si X est une variable aléatoire admettant une variance et $\varepsilon > 0$,

$$P(|X - E(X)| \geq \varepsilon) \leq \frac{V(X)}{\varepsilon^2}.$$

   Montrer que $P(Z_n = 0) \leq \dfrac{V(Z_n)}{E(Z_n)^2}$. En déduire que si $c < 1$, $P(Z_n = 0) \underset{n\to+\infty}{\longrightarrow} 0$?

   (d) Votre conjecture est-elle correcte ?

Exercice 2 (Distance en variation et couplage)

Tout au long du sujet $(\Omega, \mathscr{F}, \mathbf{P})$ désignera un espace probabilisé et les variables aléatoires utilisées plus bas seront toutes définies sur cet espace probabilisé. Sous réserve de son existence, l'espérance mathématique d'une variable aléatoire réelle X sera notée $E(X)$.

Partie 1 : Distance en variation

Dans cette première partie, on considère un ensemble discret $\mathcal{K}$ dont on suppose qu'il est soit fini soit égal à l'ensemble des entiers naturels $\mathbb{N}$. $\mathcal{A}$ désigne l'ensemble de toutes les parties de $\mathcal{K}$ et pour tout $A \in \mathcal{A}$, on note $\bar{A}$ le complémentaire de A dans $\mathcal{K}$.

Soient P et Q deux probabilités sur $\mathcal{K}$. Pour tout $k \in \mathcal{K}$, on pose $p_k = P(\{k\})$ et $q_k = Q(\{k\})$. On rappelle que pour tout $k \in \mathcal{K}$, $p_k \geq 0$, avec $\sum_{k\in\mathcal{K}} p_k = 1$. De plus, toute probabilité P est entièrement déterminée par la donnée de $(p_k)_{k\in\mathcal{K}}$ puisque pour tout $A \in \mathcal{A}$, $P(A) = \sum_{k\in A} p_k$.

Lorsque $\mathcal{K}$ est fini, on définit la distance en variation entre les probabilités P et Q par

$$D(P,Q) = \frac{1}{2}\sum_{k\in\mathcal{K}} |p_k - q_k| \tag{i}$$

1. Lorsque $\mathcal{K} = \{0;1\}$ exprimer $D(P,Q)$ en fonction de p_1 et q_1.
2. Lorsque $\mathcal{K} = \mathbb{N}$, vérifier que la série de terme général $(|p_k - q_k|)_{k\in\mathbb{N}}$ est convergente.

On étend donc la définition de la distance en variation donnée par (i) au cas où $\mathcal{K} = \mathbb{N}$.

3. Vérifier que pour tout $A \in \mathcal{A}$, $|P(A) - Q(A)| \in [0,1]$.
4. Montrer que pour tout $A \in \mathcal{A}$, $2|P(A) - Q(A)| = \left|\sum_{k\in A}(p_k - q_k)\right| + \left|\sum_{k\in\bar{A}}(p_k - q_k)\right|$.
5. En déduire que pour tout $A \in \mathcal{A}$,
$$|P(A) - Q(A)| \leq D(P,Q) \tag{ii}$$
6. Montrer que la partie $A = \{k \in \mathcal{K} \mid q_k \geq p_k\}$ réalise l'égalité dans (ii), c'est-à-dire que
$$|P(A) - Q(A)| = D(P,Q)$$
7. Démontrer la formule $D(P,Q) = 1 - \sum_{k\in\mathcal{K}} \min(p_k, q_k)$.
8. On considère un couple de variables aléatoires (X,Y) tel que X soit de loi P et Y soit de loi Q. Autrement dit, pour tout $k \in \mathcal{K}$,
$$\mathbf{P}(X = k) = p_k \quad \text{et} \quad \mathbf{P}(Y = k) = q_k$$
Montrer que $D(P,Q) \leq \mathbf{P}(X \neq Y)$.

Partie 2 : Couplage binomiale - Poisson

Soit $n \in \mathbb{N}^*$ et $\lambda \in \mathbb{R}_+^*$ tels que $\lambda < n$.

L'objet de cette deuxième partie est d'étudier un exemple, l'approximation de la loi binomiale par la loi de Poisson en terme de distance en variation.

Plus précisément, si d'une part $\mathcal{B}(n, \lambda/n)$ désigne la loi binomiale de paramètres n et λ/n et si d'autre part on note $\mathcal{P}(\lambda)$ la loi de Poisson de paramètre λ, le but est de prouver la majoration suivante :

$$D(\mathcal{B}(n,\lambda/n), \mathcal{P}(\lambda)) \leq \frac{\lambda^2}{n} \tag{iii}$$

où D est définie au (i).

9. Soit $Y_1, \ldots, Y_n$ des variables aléatoires indépendantes et de même loi de Poisson $\mathcal{P}(\lambda/n)$.

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

10. Vérifier que pour tout $x \in [0,1]$,

$$f(x) = 1 - (1-x)\exp(x).$$

appartient à $[0,1]$.

Soit $U_1, \ldots, U_n$ des variables aléatoires indépendantes de même loi de Bernoulli de paramètre $f(\lambda/n)$. On suppose que les variables $U_1, \ldots, U_n$ sont indépendantes des variables $Y_1, \ldots, Y_n$ de la question 9. Pour $i \in [\![1,n]\!]$, on pose $X_i = 0$ si $U_i = Y_i = 0$ et $X_i = 1$ sinon.

11. Vérifier que, pour tout $i \in [\![1,n]\!]$, X_i suit une loi de Bernoulli de paramètre λ/n.

    Donner la loi de $\sum\limits_{i=1}^{n} X_i$.

12. Montrer que pour tout $i \in [\![1,n]\!]$,

$$\mathbf{P}(X_i \neq Y_i) \leq \frac{\lambda^2}{n^2}$$

    On pourra établir que pour tout réel x, $1 + x \leq \exp(x)$.

13. Montrer que :

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

14. 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 (iii).

15. Quel résultat connu peut-on déduire de (iii) lorsque n tend vers l'infini ?
