Partie A : Partie aléatoire d'un ensemble fini et indépendance asymptotique

Soit $n \geqslant 2$ et soit $E = \{e_1, \dots, e_n\}$ un ensemble de cardinal n. Soit $N \in [\![0, n]\!]$.

1. Rappeler quel est le nombre de parties de E, puis quel est le nombre de parties de E à N éléments.

On note Ω l'ensemble des parties de E à N éléments. On munit $(\Omega, \mathcal{P}(\Omega))$ (qui est fini d'après la question précédente) de la probabilité uniforme $\mathbb{P}$.

(Exemple : on pourra penser à Ω comme l'ensemble des mains de N cartes possibles lorsque l'on tire au hasard et simultanément N cartes dans un jeu de n cartes)

On note, pour tout $i \in [\![1, n]\!]$,

$$T_i : \left| \begin{array}{ccl} \Omega & \to & \mathbb{R} \\ \omega & \mapsto & \begin{cases} 1 & \text{si } e_i \in \omega \\ 0 & \text{sinon} \end{cases} \end{array} \right.$$

Autrement dit, si l'on note $A_i = \{\omega \in \Omega \mid e_i \in \omega\}$, alors $T_i = \mathbb{1}_{A_i}$ (la variable aléatoire indicatrice de A_i).

(En reprenant l'exemple du jeu de cartes, A_i est l'événement : « la main tirée contient la carte numéro i »)

2. (a) Montrer que les variables aléatoires T_i ($i \in [\![1, n]\!]$) suivent toutes la même loi de Bernoulli. On note p le paramètre de cette loi (on ne demande pas d'expliciter p dans cette question).
   (b) Que vaut $\displaystyle\sum_{i=1}^{n} T_i$?
   (c) En déduire la valeur de p.
3. (a) Cas $N = 0$. Que vaut T_i pour $i \in [\![1, n]\!]$? Les v.a.r. T_i sont-elles indépendantes ?
   (b) Cas $N = n$. Que vaut T_i pour $i \in [\![1, n]\!]$? Les v.a.r. T_i sont-elles indépendantes ?
   (c) Cas $1 \leqslant N \leqslant n - 1$. Les v.a.r. T_i sont-elles indépendantes ?
4. On considère dans cette question que N dépend de n (on utilisera la notation N_n) et que $\displaystyle\lim_{n \to +\infty} N_n = +\infty$.
   (a) Justifier que, pour tout entier $k \geqslant 1$, il existe un entier $n_k \in \mathbb{N}$ tel que, pour tout entier $n \geqslant n_k$, $N_n \geqslant k$. On admet que la suite (n_k) peut-être choisie croissante. On utilisera cette propriété à la question suivante.
   (b) Montrer par récurrence :
   $$\forall k \geqslant 1, \quad \forall n \geqslant n_k, \quad \forall 1 \leqslant i_1 < i_2 < \cdots < i_k \leqslant n, \quad \mathbb{P}\left(\bigcap_{j=1}^{k} \left[T_{i_j} = 1\right]\right) = \prod_{j=0}^{k-1} \frac{N_n - j}{n - j}$$
   (c) En déduire que :
   $$\forall k \geqslant 1, \quad \forall 1 \leqslant i_1 < i_2 < \cdots < i_k, \quad \mathbb{P}\left(\bigcap_{j=1}^{k} \left[T_{i_j} = 1\right]\right) \underset{n \to +\infty}{\sim} \prod_{j=1}^{k} \mathbb{P}\left(\left[T_{i_j} = 1\right]\right)$$
   L'équivalent précédent permet d'affirmer que les v.a.r. T_i sont « asymptotiquement indépendantes ». Ainsi, pour n grand, on pourra considérer qu'elles sont indépendantes pour simplifier la présentation. C'est cette approche qui est choisie dans la partie suivante.

Partie B : Graphes aléatoires d'Erdös-Renyi

Un graphe aléatoire non orienté G est la donnée :

- d'un ensemble (fini) $S = \{s_1, \dots, s_n\}$ de $n \geqslant 2$ sommets,
- d'une famille de v.a.r. mutuellement indépendantes $(T_{i,j})_{1 \leqslant i < j \leqslant n}$ qui suivent toutes la même loi de Bernoulli de paramètre p (où $p \in]0, 1[$).

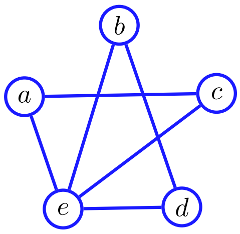
Les arêtes d'un tel graphe sont exactement les paires de sommets $\{s_i, s_j\}$ avec $i < j$ telles que $T_{i,j}$ prend la valeur 1.

On introduit :

- N_n la v.a.r. égale au nombre d'arêtes du graphe,
- pour tout $k \in [\![1, n]\!]$, D_k la v.a.r. égale au degré du sommet s_k,

- pour tout $k \in [\![1, n]\!]$, X_k la v.a.r. égale à 1 si le sommet s_k est isolé et 0 sinon (on rappelle que s_k est isolé si il n'a aucun voisin),
- Z_n la v.a.r. égale au nombre de sommets isolés du graphe.

Voici un exemple de graphe aléatoire avec $S = \{a, b, c, d, e\}$ et $p = 0,4$:



Sur cet exemple, la v.a.r. N_n prend la valeur 6, la v.a.r. D_1 prend la valeur 2, la v.a.r. X_1 prend la valeur 0 et la v.a.r. Z_n prend la valeur 0.

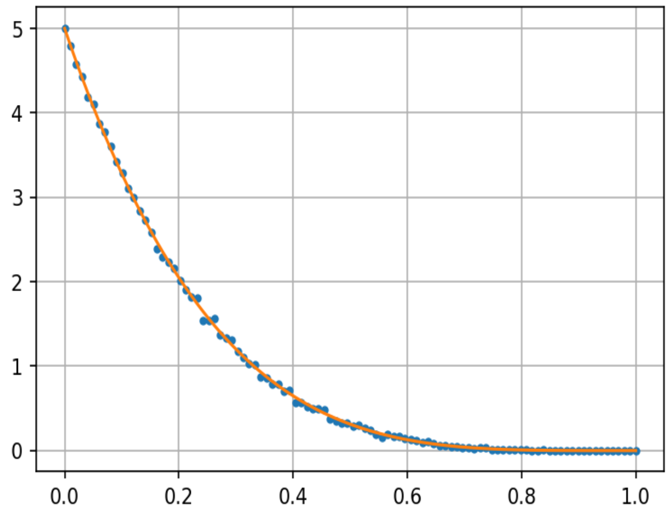
On peut considérer qu'un graphe aléatoire est un modèle très simplifié de réseau social à un instant donné.

5. Ecrire la matrice d'adjacence M du graphe donné en exemple. Quelle propriété possède la matrice M ?
6. Simulation informatique. On importera les bibliothèques `numpy as np` et `numpy.random as rd` si besoin.
   (a) Ecrire une fonction **Python** `listAdj(S,p)` qui génère la liste des listes d'adjacence d'un tel graphe aléatoire ayant `S` pour liste de sommets.
   On donne à titre d'exemple la liste des listes d'adjacence du graphe dessiné en exemple :
   `[['c', 'e'], ['d', 'e'], ['a', 'e'], ['b', 'e'], ['a', 'b', 'c', 'd']]`
   (b) Ecrire une fonction **Python** `simulZ(lst)` qui renvoie le nombre de sommets isolés d'un graphe donné par sa liste de listes d'adjacence `lst`.
7. Etude de N_n.
   (a) Quel est le nombre minimal et le nombre maximal d'arêtes de G ? En déduire $N_n(\Omega)$.
   (b) Exprimer N_n à l'aide des v.a.r. $T_{i,j}$ pour $i < j$. En déduire la loi de N_n.
8. Etude de D_k.
   (a) Soit $k \in [\![1, n]\!]$. Déterminer la loi de D_k.
   (b) Soient $1 \leqslant k < \ell \leqslant n$. Montrer que :
   $$\text{Cov}(D_k, D_\ell) = p(1-p)$$
   (c) Soient $1 \leqslant k < \ell \leqslant n$. Les v.a.r. D_k et D_ℓ sont-elles indépendantes ?
9. Etude de X_k. Calculer $\mathbb{P}([X_k = 1])$. En déduire la loi de X_k.
10. Etude de Z_n.
    (a) On considère la fonction **Python** suivante :

```
def Mystere(S,p):
    return np.mean([simulZ(listAdj(S,p)) for k in range(1000)])
```

    Que renvoie-t-elle ?
    (b) On exécute le code suivant et on reproduit ci-dessous le graphique obtenu. Que peut-on conjecturer ?

```
S = 'abcde'
x = np.linspace(0,1,100)
y = [Mystere(S,p) for p in x]
w = [len(S)*(1-p)**(len(S)-1) for p in x]
plt.grid()
plt.plot(x,y,'.')
plt.plot(x,w)
plt.show()
```



(c) Montrer que : $Z_n = \sum_{k=1}^{n} X_k$. En déduire que la conjecture précédente est vraie.

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

(e) Justifier que, pour tout $1 \leqslant i < j \leqslant n$, $\mathbb{P}([X_i = 1] \cap [X_j = 1]) = (1-p)^{2n-3}$.

(f) En déduire que : $\mathbb{E}(Z_n^2) = n(1-p)^{n-1} + n(n-1)(1-p)^{2n-3}$.

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

11. On souhaite estimer l'influence de la valeur de c sur le nombre de sommets isolés. On exécute le script suivant :

```
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 simulZ(listAdj(range(1,n+1),c*np.log(n)/n)) == 0:
            s = s+1
    res.append(s/200)
print(res)
```

et on obtient le résultat `[0.0, 0.0, 0.0, 0.91, 0.975, 0.99]` après de longues minutes.
Quelle conjecture sur la valeur d'une probabilité peut-on faire lorsque $c < 1$ et $c > 1$? Justifier.

12. (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 $a > 0$, alors

$$\mathbb{P}([X \geqslant a]) \leqslant \frac{\mathbb{E}(X)}{a}$$

Si $c > 1$, en déduire la limite de $\mathbb{P}([Z_n \geqslant 1])$ puis de $\mathbb{P}([Z_n = 0])$ lorsque $n \to +\infty$.

(c) En utilisant l'inégalité de Bienaymé-Tchébichev, montrer que $\mathbb{P}([Z_n = 0]) \leqslant \dfrac{\mathbb{V}(Z_n)}{(\mathbb{E}(Z_n))^2}$.

En déduire que, si $c < 1$, alors $\lim_{n\to+\infty} \mathbb{P}([Z_n = 0]) = 0$.

(d) La conjecture faite en question 11 était-elle correcte ?