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.

   Démonstration. Si E est de cardinal n, alors l'ensemble des parties de E est de cardinal 2^n. De plus, il y a $\binom{n}{N}$ 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).

   Démonstration. Soit $i \in [\![1, n]\!]$. Tout d'abord, $T_i(\Omega) = \{0, 1\}$ donc T_i suit une loi de Bernoulli. Notons p_i le paramètre de cette loi. Par équiprobabilité, chaque élément e_j ($j \in [\![1, n]\!]$) joue le même rôle que les autres et donc les p_j sont deux à deux égaux. On note p le paramètre commun de cette loi de Bernoulli. □

   (b) Que vaut $\sum_{i=1}^{n} T_i$?

   Démonstration. Soit $\omega \in \Omega$. L'entier $\sum_{i=1}^{n} T_i(\omega)$ est égal au nombre d'éléments de E qui appartiennent à la partie ω. Or, cette partie ω possède N éléments par définition de Ω, d'où $\sum_{i=1}^{n} T_i(\omega) = N$. Ceci étant vrai pour tout $\omega \in \Omega$, il vient que :

   $$\boxed{\sum_{i=1}^{n} T_i = N}$$

   □

   (c) En déduire la valeur de p.

   Démonstration. La variable aléatoire $\sum_{i=1}^{n} T_i$ admet une espérance comme somme de variables aléatoires qui en admettent une et, par linéarité de l'espérance,

   $$\mathbb{E}\left(\sum_{i=1}^{n} T_i\right) = \sum_{i=1}^{n} \mathbb{E}(T_i) = \sum_{i=1}^{n} p = np$$

   D'autre part,

   $$\mathbb{E}\left(\sum_{i=1}^{n} T_i\right) = \mathbb{E}(N) = N$$

   D'où

   $$\boxed{p = \frac{N}{n}}$$

   □

3. (a) Cas $N = 0$. Que vaut T_i pour $i \in [\![1, n]\!]$? Les v.a.r. T_i sont-elles indépendantes ?

   Démonstration. Si $N = 0$, alors les variables aléatoires T_i sont toutes constantes égales à 0. Dans ce cas, elles sont indépendantes (toute variable aléatoire constante est indépendante de toute autre variable aléatoire). ☐

   (b) Cas $N = n$. Que vaut T_i pour $i \in [\![1, n]\!]$? Les v.a.r. T_i sont-elles indépendantes ?

   Démonstration. Si $N = n$, alors les variables aléatoires T_i sont toutes constantes égales à 1. Dans ce cas, elles sont indépendantes. ☐

   (c) Cas $1 \leqslant N \leqslant n - 1$. Les v.a.r. T_i sont-elles indépendantes ?

   Démonstration. Tout d'abord, puisque $N < n$:

$$\mathbb{P}\left(\bigcap_{i=1}^{n} [T_i = 1]\right) = \mathbb{P}(\varnothing) = 0$$

   Ensuite,

$$\prod_{i=1}^{n} \mathbb{P}([T_i = 1]) = \prod_{i=1}^{n} \frac{N}{n} = \left(\frac{N}{n}\right)^n$$

   et puisque $N > 0$:

$$\prod_{i=1}^{n} \mathbb{P}([T_i = 1]) \neq 0$$

   Donc les variables aléatoires T_i ne sont pas mutuellement indépendantes. ☐

4. On considère dans cette question que N dépend de n (on utilisera la notation N_n) et que $\lim\limits_{n \to +\infty} N_n = +\infty$.

   (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.

   Démonstration. Par hypothèse : $\lim\limits_{n \to +\infty} N_n = +\infty$. Ainsi, par définition de la limite :

$$\forall A > 0, \exists n_0 \in \mathbb{N}, \forall n \geqslant n_0, N_n \geqslant A$$

   En particulier, en prenant $A = k \geqslant 1$ un entier, il existe un entier n_k tel que, pour tout entier $n \geqslant n_k$, $N_n \geqslant k$. ☐

   (b) Montrer par récurrence :

$$\forall k \geqslant 1, \quad \forall n \geqslant n_k, \quad \forall 1 \leqslant i_1 < i_2 < \cdots < i_k \leqslant n, \quad \mathbb{P}\left(\bigcap_{j=1}^{k} \left[T_{i_j} = 1\right]\right) = \prod_{j=0}^{k-1} \frac{N_n - j}{n - j}$$

   Démonstration. Montrons par récurrence : $\forall k \in \mathbb{N}^*$, $\mathcal{P}(k)$

   où $\mathcal{P}(k)$: « $\forall n \geqslant n_k, \quad \forall 1 \leqslant i_1 < i_2 < \cdots < i_k \leqslant n, \quad \mathbb{P}\left(\bigcap_{j=1}^{k} \left[T_{i_j} = 1\right]\right) = \prod_{j=0}^{k-1} \frac{N_n - j}{n - j}$ »

   <u>Initialisation</u> : soit $n \geqslant n_1$. Soit $1 \leqslant i_1 \leqslant n$.

   D'une part, $\mathbb{P}(\left[T_{i_1} = 1\right]) = \dfrac{N_n}{n}$ (cf question 2c).

   D'autre part, $\displaystyle\prod_{j=0}^{1-1} \frac{N_n - j}{n - j} = \frac{N_n}{n}$.

   D'où $\mathcal{P}(1)$.

   <u>Hérédité</u> : soit $k \geqslant 1$. Supposons $\mathcal{P}(k)$. Montrons $\mathcal{P}(k+1)$.

   Soit $n \geqslant n_{k+1}$. Soit $1 \leqslant i_1 < i_2 < \cdots < i_{k+1} \leqslant n$.

   Par hypothèse de récurrence :

$$\mathbb{P}\left(\bigcap_{j=1}^{k} \left[T_{i_j} = 1\right]\right) = \prod_{j=0}^{k-1} \frac{N_n - j}{n - j}$$

De plus, $n \geqslant n_{k+1}$ donc $n \geqslant n_k$ car la suite (n_k) est croissante. D'où $N_n \geqslant k > k-1$ et donc

$$\mathbb{P}\left(\bigcap_{j=1}^{k}\left[T_{i_j}=1\right]\right) \neq 0$$

On en déduit que l'on peut utiliser la formule des probabilités composées de la manière suivante :

$$\mathbb{P}\left(\bigcap_{j=1}^{k+1}\left[T_{i_j}=1\right]\right) = \mathbb{P}\left(\bigcap_{j=1}^{k}\left[T_{i_j}=1\right]\right)\mathbb{P}_{\bigcap_{j=1}^{k}\left[T_{i_j}=1\right]}\left(\left[T_{i_{k+1}}=1\right]\right)$$

Supposons l'événement $\bigcap_{j=1}^{k}\left[T_{i_j}=1\right]$ réalisé. Alors k élements de E ont déjà été choisis et il reste à en choisir $N-k$ parmi les $n-k$ éléments restants de E pour construire une partie à N éléments de E. En utilisant à nouveau la question 2c avec ces paramètres, il vient que

$$\mathbb{P}_{\bigcap_{j=1}^{k}\left[T_{i_j}=1\right]}\left(\left[T_{i_{k+1}}=1\right]\right) = \frac{N_n-k}{n-k}$$

Ainsi,

$$\mathbb{P}\left(\bigcap_{j=1}^{k+1}\left[T_{i_j}=1\right]\right) = \left(\prod_{j=0}^{k-1}\frac{N_n-j}{n-j}\right)\frac{N_n-k}{n-k} = \prod_{j=0}^{k}\frac{N_n-j}{n-j}$$

D'où $\mathcal{P}(k+1)$. □

(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.

Démonstration. Soit $k \geqslant 1$. Soit $1 \leqslant i_1 < i_2 < \cdots < i_k$. Soit $n \geqslant \max(n_k, i_k)$. D'après la question 4b, on a

$$\mathbb{P}\left(\bigcap_{j=1}^{k}\left[T_{i_j}=1\right]\right) = \prod_{j=0}^{k-1}\frac{N_n-j}{n-j} \underset{n\to+\infty}{\sim} \prod_{j=0}^{k-1}\frac{N_n}{n} = \left(\frac{N_n}{n}\right)^k = \prod_{j=1}^{k}\mathbb{P}\left(\left[T_{i_j}=1\right]\right)$$

□

Partie B : Graphes aléatoires d'Erdös-Renyi

Un graphe aléatoire non orienté G est la donnée :

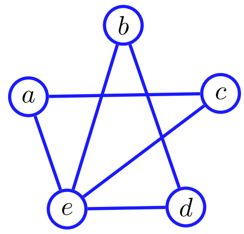
- d'un ensemble (fini) $S = \{s_1, \ldots, s_n\}$ de $n \geqslant 2$ sommets,
- d'une famille de v.a.r. mutuellement indépendantes $(T_{i,j})_{1\leqslant i<j\leqslant n}$ qui suivent toutes la même loi de Bernoulli de paramètre p (où $p \in]0,1[$).

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 ?

Démonstration. Par définition de la matrice d'adjacence :

$$M = \begin{pmatrix} 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 & 0 \end{pmatrix}$$

Cette matrice est symétrique, ce qui est cohérent avec le fait que le graphe soit non orienté. $\square$

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']]`

Démonstration. On propose la fonction **Python** qui suit :

```
def listAdj(S,p):
    n = len(S)
    L = [[] for k in S]
    for i in range(n-1):
        for j in range(i+1,n):
            if rd.random() < p:
                L[i].append(S[j])
                L[j].append(S[i])
    return L
```

$\square$

(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`.

Démonstration. On propose la fonction **Python** qui suit :

```
def simulZ(lst):
    cpt = 0
    n = len(lst)
    for k in range(n):
        if len(lst[k]) == 0:
            cpt = cpt + 1
    return cpt
```

$\square$

7. Etude de N_n.
   (a) Quel est le nombre minimal et le nombre maximal d'arêtes de G ? En déduire $N_n(\Omega)$.

   Démonstration. Il y a au minimum 0 arêtes et au maximum $\binom{n}{2} = \dfrac{n(n-1)}{2}$ arêtes (il y en a autant que de parties à 2 éléments de l'ensemble des sommets de G). Ainsi :

$$\boxed{N_n(\Omega) = \left[\!\left[0, \frac{n(n-1)}{2}\right]\!\right]}$$

   □

   (b) Exprimer N_n à l'aide des v.a.r. $T_{i,j}$ pour $i < j$. En déduire la loi de N_n.

   Démonstration. Puisque N_n est égale au nombre d'arêtes dans le graphe, que ces arêtes sont forcément entre deux sommets s_i et s_j avec $i < j$ et que $T_{i,j}$ prend la valeur 1 si il y a une arête entre s_i et s_j et la valeur 0 sinon, on en déduit :

$$\boxed{N_n = \sum_{1\leqslant i<j\leqslant n} T_{i,j}}$$

   De plus,
   - Les variables aléatoires $T_{i,j}$ sont mutuellement indépendantes
   - Les variables aléatoires $T_{i,j}$ suivent toutes la même loi de Bernoulli de paramètre p
   - Il y a $\binom{n}{2} = \dfrac{n(n-1)}{2}$ couples (i,j) vérifiant $1 \leqslant i < j \leqslant n$

   Ainsi : $\boxed{N_n \hookrightarrow \mathcal{B}\left(\dfrac{n(n-1)}{2}, p\right)}$. □

8. Etude de D_k.
   (a) Soit $k \in [\![1, n]\!]$. Déterminer la loi de D_k.

   Démonstration. Soit $k \in [\![1, n]\!]$. La variable aléatoire D_k est égale au nombre de voisins du sommet s_k. On en déduit :

$$D_k = \sum_{\substack{1\leqslant i\leqslant n\\ i<k}} T_{i,k} + \sum_{\substack{1\leqslant j\leqslant n\\ j>k}} T_{k,j}$$

   Ainsi, D_k est une somme de $n-1$ variables aléatoires indépendantes qui suivent toutes la même loi de Bernoulli de paramètre p. Il vient alors : $\boxed{D_k \hookrightarrow \mathcal{B}(n-1, p)}$. □

   (b) Soient $1 \leqslant k < \ell \leqslant n$. Montrer que :

$$\mathrm{Cov}(D_k, D_\ell) = p(1-p)$$

   Démonstration. Soient $1 \leqslant k < \ell \leqslant n$. Par bilinéarité de la covariance :

$$\begin{aligned}
\mathrm{Cov}(D_k, D_\ell) &= \mathrm{Cov}\left(\sum_{\substack{1\leqslant i\leqslant n\\ i<k}} T_{i,k} + \sum_{\substack{1\leqslant j\leqslant n\\ j>k}} T_{k,j}, \sum_{\substack{1\leqslant m\leqslant n\\ m<\ell}} T_{m,\ell} + \sum_{\substack{1\leqslant r\leqslant n\\ r>\ell}} T_{\ell,r}\right)\\
&= \mathrm{Cov}\left(\sum_{\substack{1\leqslant i\leqslant n\\ i<k}} T_{i,k}, \sum_{\substack{1\leqslant m\leqslant n\\ m<\ell}} T_{m,\ell}\right) + \mathrm{Cov}\left(\sum_{\substack{1\leqslant i\leqslant n\\ i<k}} T_{i,k}, \sum_{\substack{1\leqslant r\leqslant n\\ r>\ell}} T_{\ell,r}\right)\\
&\quad + \mathrm{Cov}\left(\sum_{\substack{1\leqslant j\leqslant n\\ j>k}} T_{k,j}, \sum_{\substack{1\leqslant m\leqslant n\\ m<\ell}} T_{m,\ell}\right) + \mathrm{Cov}\left(\sum_{\substack{1\leqslant j\leqslant n\\ j>k}} T_{k,j}, \sum_{\substack{1\leqslant r\leqslant n\\ r>\ell}} T_{\ell,r}\right)\\
&= \sum_{\substack{1\leqslant i\leqslant n\\ i<k}} \sum_{\substack{1\leqslant m\leqslant n\\ m<\ell}} \mathrm{Cov}\left(T_{i,k}, T_{m,\ell}\right) + \sum_{\substack{1\leqslant i\leqslant n\\ i<k}} \sum_{\substack{1\leqslant r\leqslant n\\ r>\ell}} \mathrm{Cov}\left(T_{i,k}, T_{\ell,r}\right)\\
&\quad + \sum_{\substack{1\leqslant j\leqslant n\\ j>k}} \sum_{\substack{1\leqslant m\leqslant n\\ m<\ell}} \mathrm{Cov}\left(T_{k,j}, T_{m,\ell}\right) + \sum_{\substack{1\leqslant j\leqslant n\\ j>k}} \sum_{\substack{1\leqslant r\leqslant n\\ r>\ell}} \mathrm{Cov}\left(T_{k,j}, T_{\ell,r}\right)
\end{aligned}$$

   De plus :

- Dans la première somme (double), pour tout $(i,m) \in [\![1,n]\!]^2$, les variables aléatoires $T_{i,k}$ et $T_{m,\ell}$ sont indépendantes (car $k \neq \ell$) donc $\mathrm{Cov}\left(T_{i,k}, T_{m,\ell}\right) = 0$.
- Dans la dernière somme, pour tout $(j,r) \in [\![1,n]\!]^2$, les variables aléatoires $T_{k,j}$ et $T_{\ell,r}$ sont indépendantes (car $k \neq \ell$) donc $\mathrm{Cov}\left(T_{k,j}, T_{\ell,r}\right) = 0$.

Ainsi :

$$\mathrm{Cov}(D_k, D_\ell) = \sum_{\substack{1\leqslant i\leqslant n \\ i<k}} \sum_{\substack{1\leqslant r\leqslant n \\ r>\ell}} \mathrm{Cov}\left(T_{i,k}, T_{\ell,r}\right) + \sum_{\substack{1\leqslant j\leqslant n \\ j>k}} \sum_{\substack{1\leqslant m\leqslant n \\ m<\ell}} \mathrm{Cov}\left(T_{k,j}, T_{m,\ell}\right)$$

De plus,

- Dans la première somme, on a toujours $i < k < \ell$ et donc les variables aléatoires $T_{i,k}$ et $T_{\ell,r}$ sont indépendantes. Ainsi :

$$\sum_{\substack{1\leqslant i\leqslant n \\ i<k}} \sum_{\substack{1\leqslant r\leqslant n \\ r>\ell}} \mathrm{Cov}\left(T_{i,k}, T_{\ell,r}\right) = 0$$

- Dans la deuxième somme, tous les termes sont nuls sauf celui correspondant aux indices $j = \ell$ et $m = k$. Ainsi :

$$\sum_{\substack{1\leqslant j\leqslant n \\ j>k}} \sum_{\substack{1\leqslant m\leqslant n \\ m<\ell}} \mathrm{Cov}\left(T_{k,j}, T_{m,\ell}\right) = \mathrm{Cov}\left(T_{k,\ell}, T_{k,\ell}\right) = \mathbb{V}(T_{k,\ell}) = p(1-p)$$

D'où

$$\boxed{\mathrm{Cov}(D_k, D_\ell) = p(1-p)}$$

☐

(c) Soient $1 \leqslant k < \ell \leqslant n$. Les v.a.r. D_k et D_ℓ sont-elles indépendantes ?

Démonstration. Soient $1 \leqslant k < \ell \leqslant n$. D'après la question précédente : $\mathrm{Cov}(D_k, D_\ell) = p(1-p) \neq 0$ donc les variables aléatoires D_k et D_ℓ ne sont pas indépendantes. ☐

9. Etude de X_k. Calculer $\mathbb{P}([X_k = 1])$. En déduire la loi de X_k.

Démonstration. Soit $k \in [\![1,n]\!]$. Tout d'abord, $X_k(\Omega) = \{0,1\}$ donc X_k suit une loi de Bernoulli. De plus :

$$\begin{aligned} \mathbb{P}([X_k = 1]) &= \mathbb{P}([D_k = 0]) \\ &= (1-p)^{n-1} && \textit{(car } D_k \hookrightarrow \mathcal{B}(n-1,p)\textit{)} \end{aligned}$$

Ainsi : $\boxed{X_k \hookrightarrow \mathcal{B}\left((1-p)^{n-1}\right)}$. ☐

10. Etude de Z_n.

(a) On considère la fonction **Python** suivante :

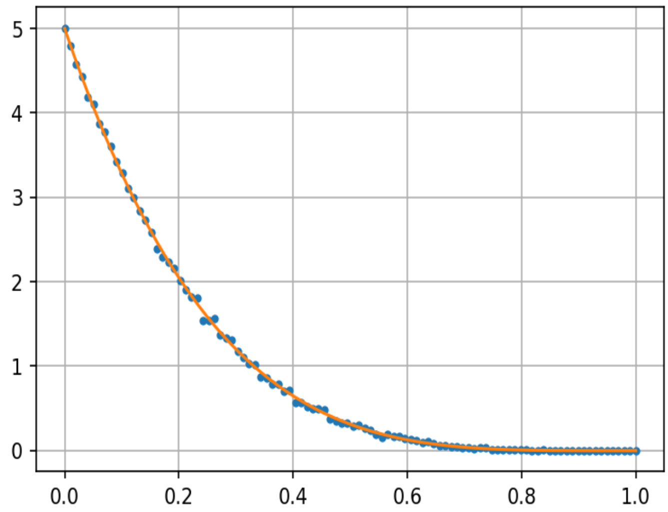
```
def Mystere(S,p):
    return np.mean([simulZ(listAdj(S,p)) for k in range(1000)])
```

Que renvoie-t-elle ?

Démonstration. L'appel `Mystere(S,p)` renvoie la moyenne empirique de 1000 simulations de Z_n. D'après la loi faible des grands nombres (Z_n est finie donc admet une variance), il s'agit d'une approximation de $\mathbb{E}(Z_n)$. ☐

(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()
```



Démonstration. D'après ce graphique, $\mathbb{E}(Z_n)$ semble coïncider avec $n(1-p)^{n-1}$ (tout du moins pour $n = 5$ puisque dans ce script le graphe contient 5 sommets). En effet, les valeurs approchées de $\mathbb{E}(Z_n)$ calculées grâce à l'appel `Mystere(S,p)` pour différentes valeurs de p se superposent à la courbe de la fonction $p \mapsto n(1-p)^{n-1}$. □

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

Démonstration. Soit $\omega \in \Omega$. Soit $k \in [\![0, n]\!]$. Rappelons que X_i prend la valeur 1 si s_i est isolé et la valeur 0 sinon. Ainsi :

$$\begin{aligned} \sum_{i=1}^{n} X_i(\omega) = k &\iff \mathrm{Card}(\{i \in [\![1, n]\!] \mid X_i(\omega) = 1\}) = k \\ &\iff \text{exactement } k \text{ sommets sont isolés} \\ &\iff Z_n(\omega) = k \end{aligned}$$

Ceci étant vrai pour tout $\omega \in \Omega$:

$$\boxed{Z_n = \sum_{k=1}^{n} X_k}$$

On en déduit que Z_n admet une espérance comme somme de variables aléatoires admettant une espérance et, par linéarité :

$$\mathbb{E}(Z_n) = \sum_{k=1}^{n} \mathbb{E}(X_k) = \sum_{k=1}^{n} (1-p)^{n-1} = n(1-p)^{n-1}$$

La conjecture faite à la question précédente est vérifiée. □

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

Démonstration.

$$\begin{aligned}
Z_n^2 &= \left(\sum_{k=1}^{n} X_k\right)^2 \\
&= \left(\sum_{k=1}^{n} X_k\right)\left(\sum_{j=1}^{n} X_j\right) \\
&= \sum_{1\leqslant k,j\leqslant n} X_kX_j \\
&= \sum_{\substack{1\leqslant k,j\leqslant n\\ j=k}} X_kX_j + \sum_{\substack{1\leqslant k,j\leqslant n\\ j\neq k}} X_kX_j \\
&= \sum_{k=1}^{n} X_k^2 + \sum_{\substack{1\leqslant k,j\leqslant n\\ j<k}} X_kX_j + \sum_{\substack{1\leqslant k,j\leqslant n\\ j>k}} X_kX_j
\end{aligned}$$

Or,

- $X_k(\Omega) = \{0,1\}$ donc $X_k^2 = X_k$.
- $\displaystyle\sum_{\substack{1\leqslant k,j\leqslant n\\ j>k}} X_kX_j = \sum_{\substack{1\leqslant k,j\leqslant n\\ k<j}} X_jX_k = \sum_{\substack{1\leqslant k,j\leqslant n\\ j<k}} X_kX_j$ (en renommant les indices)

D'où

$$\boxed{Z_n^2 = \sum_{k=1}^{n} X_k + 2\sum_{1\leqslant i<j\leqslant n} X_iX_j}$$

□

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

Démonstration. Soient $1 \leqslant i < j \leqslant n$.
Tout d'abord, d'après la question 9 :

$$\mathbb{P}([X_i = 1]) = (1-p)^{n-1} \neq 0$$

Ainsi :

$$\mathbb{P}([X_i = 1] \cap [X_j = 1]) = \mathbb{P}([X_i = 1])\mathbb{P}_{[X_i=1]}([X_j = 1])$$

Si l'événement $[X_i = 1]$ est réalisé, alors le sommet s_i est isolé et donc, en particulier, on sait qu'il n'y a pas d'arête reliant s_i et s_j. Dans ce cas, tout se passe comme si on considérait un graphe aléatoire à $n-1$ sommet (on a retiré le sommet s_i). Donc la loi conditionnelle de X_j sachant $[X_i = 1]$ dans un graphe à n sommets coïncide avec la loi de X_j dans un graphe à $n-1$ sommets. D'où :

$$\mathbb{P}_{[X_i=1]}([X_j = 1]) = (1-p)^{n-2}$$

Finalement :

$$\boxed{\mathbb{P}([X_i = 1] \cap [X_j = 1]) = (1-p)^{2n-3}}$$

□

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

Démonstration. Les variables aléatoires Z_n^2, X_k et X_iX_j sont toutes finies donc admettent chacune une espérance. Par linéarité :

$$\mathbb{E}(Z_n^2) = \sum_{k=1}^{n} \mathbb{E}(X_k) + 2\sum_{1\leqslant i<j\leqslant n} \mathbb{E}(X_iX_j)$$

On sait déjà que $X_k \hookrightarrow \mathcal{B}\Big((1-p)^{n-1}\Big)$ et, d'après la question précédente, $X_iX_j \hookrightarrow \mathcal{B}\Big((1-p)^{2n-3}\Big)$. D'où :

$$\begin{aligned}
\mathbb{E}(Z_n^2) &= \sum_{k=1}^{n}(1-p)^{n-1} + 2\sum_{1\leqslant i<j\leqslant n}(1-p)^{2n-3} \\
&= n(1-p)^{n-1} + 2\frac{n(n-1)}{2}(1-p)^{2n-3} \\
&= n(1-p)^{n-1} + n(n-1)(1-p)^{2n-3}
\end{aligned}$$

□

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.

Démonstration. D'après la loi faible des grands nombres, ce script renvoie une liste contenant des approximations de la probabilité $\mathbb{P}([Z_n = 0])$ pour un graphe aléatoire à $n = 1000$ sommets et pour différentes valeurs du paramètre c. On observe deux choses :

- Lorsque $c = 0,3$ ou $c = 0,5$ ou $c = 0,7$, $\mathbb{P}([Z_n = 0]) = 0$.
- Lorsque $c = 1,3$ ou $c = 1,5$ ou $c = 1,7$, $\mathbb{P}([Z_n = 0])$ est proche de 1.

On émet alors la conjecture suivante :

- Si $c < 1$, alors $\mathbb{P}([Z_n = 0])$ est proche de 0.
- Si $c > 1$, alors $\mathbb{P}([Z_n = 0])$ est proche de 1.

□

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

Démonstration. Tout d'abord, par croissances comparées :

$$p_n = c\frac{\ln(n)}{n} \underset{n\to+\infty}{\longrightarrow} 0$$

D'où

$$\frac{(1-p_n)^n}{(1-p_n)^{n-1}} = 1 - p_n \underset{n\to+\infty}{\longrightarrow} 1$$

et donc

$$\boxed{(1-p_n)^{n-1} \underset{n\to+\infty}{\sim} (1-p_n)^n}$$

Montrons pour finir que $(1-p_n)^n \underset{n\to+\infty}{\sim} n^{-c}$. On a

$$(1-p_n)^n = \mathrm{e}^{n\ln(1-p_n)}$$

Rappelons le DL à l'ordre 2 de $\ln(1+u)$ en 0 :

$$\ln(1+u) = u - \frac{1}{2}u^2 + \underset{u\to 0}{o}(u^2)$$

Ainsi :

$$\begin{aligned}\ln(1-p_n) &= -p_n - \frac{1}{2}p_n^2 + \underset{n\to+\infty}{o}(p_n^2)\\ &= -c\frac{\ln(n)}{n} - \frac{1}{2}c^2\frac{\ln(n)^2}{n^2} + \underset{n\to+\infty}{o}\left(\frac{\ln(n)^2}{n^2}\right)\end{aligned}$$

et donc

$$n\ln(1-p_n) = -c\ln(n) - \frac{1}{2}c^2\frac{\ln(n)^2}{n} + \underset{n\to+\infty}{o}\left(\frac{\ln(n)^2}{n}\right)$$

ce qui donne :

$$\begin{aligned}(1-p_n)^n &= \mathrm{e}^{-c\ln(n)}\mathrm{e}^{-\frac{1}{2}c^2\frac{\ln(n)^2}{n} + \underset{n\to+\infty}{o}\left(\frac{\ln(n)^2}{n}\right)}\\ &= n^{-c}\mathrm{e}^{-\frac{1}{2}c^2\frac{\ln(n)^2}{n} + \underset{n\to+\infty}{o}\left(\frac{\ln(n)^2}{n}\right)}\end{aligned}$$

Or, par croissances comparées :

$$-\frac{1}{2}c^2\frac{\ln(n)^2}{n} + \underset{n\to+\infty}{o}\left(\frac{\ln(n)^2}{n}\right) \underset{n\to+\infty}{\longrightarrow} 0$$

et par continuité de l'exponentielle en 0 :

$$\mathrm{e}^{-\frac{1}{2}c^2\frac{\ln(n)^2}{n} + \underset{n\to+\infty}{o}\left(\frac{\ln(n)^2}{n}\right)} \underset{n\to+\infty}{\longrightarrow} 1$$

On obtient bien l'équivalent voulu :

$$\boxed{(1-p_n)^{n-1} \underset{n\to+\infty}{\sim} (1-p_n)^n \underset{n\to+\infty}{\sim} n^{-c}}$$

□

(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 $n \to +\infty$.

Démonstration. La variable aléatoire Z_n est positive et admet une espérance donc, en appliquant l'inégalité de Markov avec $a = 1$, on obtient :

$$0 \leqslant \mathbb{P}([Z_n \geqslant 1]) \leqslant \mathbb{E}(Z_n) = n(1-p_n)^{n-1} \underset{n\to+\infty}{\sim} nn^{-c} = n^{1-c}$$

Or, on suppose dans cette question que $c > 1$, donc $1 - c < 0$. On en déduit :

$$\mathbb{E}(Z_n) \underset{n\to+\infty}{\longrightarrow} 0$$

Par théorème d'encadrement :

$$\boxed{\mathbb{P}([Z_n \geqslant 1]) \underset{n\to+\infty}{\longrightarrow} 0}$$

De plus, $Z_n(\Omega) = [\![0, n]\!]$ donc

$$\mathbb{P}([Z_n = 0]) + \mathbb{P}([Z_n \geqslant 1]) = 1$$

ce qui permet de conclure :

$$\boxed{\mathbb{P}([Z_n = 0]) \underset{n\to+\infty}{\longrightarrow} 1}$$

□

(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 $\lim\limits_{n\to+\infty} \mathbb{P}([Z_n = 0]) = 0$.

Démonstration. La variable aléatoire Z_n admet un moment d'ordre 2 donc une variance. D'après l'inégalité de Bienaymé-Tchébichev :

$$\forall \varepsilon > 0,\ \mathbb{P}\left([|Z_n - \mathbb{E}(Z_n)| \geqslant \varepsilon]\right) \leqslant \frac{\mathbb{V}(Z_n)}{\varepsilon^2}$$

En particulier, pour $\varepsilon = \mathbb{E}(Z_n) = n(1-p_n)^{n-1} > 0$:

$$\mathbb{P}\left([|Z_n - \mathbb{E}(Z_n)| \geqslant \mathbb{E}(Z_n)]\right) \leqslant \frac{\mathbb{V}(Z_n)}{\mathbb{E}(Z_n)^2}$$

De plus, on remarque :

$$[Z_n = 0] \subset [|Z_n - \mathbb{E}(Z_n)| \geqslant \mathbb{E}(Z_n)]$$

Ainsi, par croissance de $\mathbb{P}$:

$$\boxed{0 \leqslant \mathbb{P}([Z_n = 0]) \leqslant \frac{\mathbb{V}(Z_n)}{\mathbb{E}(Z_n)^2}}$$

Or,

$$\begin{aligned}
\frac{\mathbb{V}(Z_n)}{\mathbb{E}(Z_n)^2} &= \frac{\mathbb{E}(Z_n^2) - \mathbb{E}(Z_n)^2}{\mathbb{E}(Z_n)^2} && \textit{(par Koenig-Huygens)} \\
&= \frac{\mathbb{E}(Z_n^2)}{\mathbb{E}(Z_n)^2} - 1 \\
&= \frac{n(1-p)^{n-1} + n(n-1)(1-p)^{2n-3}}{\left(n(1-p_n)^{n-1}\right)^2} - 1 \\
&= \frac{1}{n(1-p_n)^{n-1}} + \frac{n-1}{n(1-p_n)} - 1
\end{aligned}$$

et $n(1-p_n)^{n-1} \underset{n\to+\infty}{\sim} n^{1-c} \underset{n\to+\infty}{\longrightarrow} +\infty$ (car $c < 1$). D'où :

$$\lim_{n\to+\infty} \frac{\mathbb{V}(Z_n)}{\mathbb{E}(Z_n)^2} = 0$$

Finalement, par théorème d'encadrement :

$$\boxed{\lim_{n\to+\infty} \mathbb{P}([Z_n = 0]) = 0}$$

□

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

Démonstration. Oui, la conjecture faite en question 11 était correcte. □