Sujets de révisions
Problèmes du TOP3

HEC/ESSEC I 2024 - nombre de triangles dans un graphe aléatoire, inégalité de Harris, inégalité de Janson, inégalité de Boole, espérance conditionnelle

Dans tout le sujet on considère un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$, toutes les variables aléatoires qui interviennent dans la suite sont définies sur cet espace.

Soit n un entier supérieur ou égal à 3 et p un réel appartenant à $]0,1[$.

Pour générer des graphes non orientés de manière aléatoire, on se donne :

- $S = [\![0, n-1]\!]$, les sommets du graphe ;
- pour toute paire de sommets $\{u,v\}$ avec $u < v$, $T_{u,v}$ une variable aléatoire de Bernoulli de paramètre p.
  Les variables aléatoires $T_{u,v}$ pour $\{u,v\}$ décrivant les paires de sommets avec $u < v$, sont supposées indépendantes ;
- les arêtes d'un graphe G ainsi généré sont les paires $\{u,v\}$ telles que $T_{u,v} = 1$ si $u < v$ ou $T_{v,u} = 1$ si $v < u$.

Dans tout le problème, par convention, une somme portant sur un ensemble d'indices vide vaut 0, un produit vaut 1, une intersection vaut Ω, une réunion vaut $\varnothing$.

Partie 1 - Nombre aléatoire de triangles

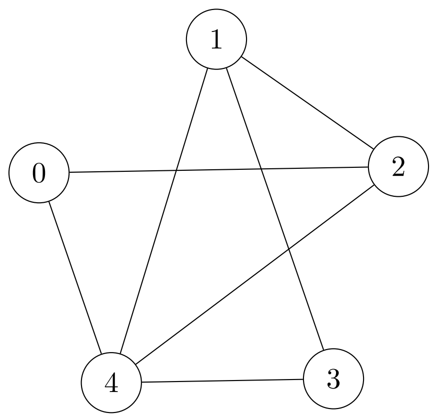
On note $\mathcal{T}$ l'ensemble des parties $\{u,v,w\}$ à trois éléments de l'ensemble des sommets, r le nombre de ses éléments et on pose

$$\mathcal{T} = \{t_1, \dots, t_r\}$$

Étant donné $t = \{u,v,w\}$, un élément de $\mathcal{T}$, on dit que t est un triangle dans un graphe G généré aléatoirement si $\{u,v\}$, $\{v,w\}$ et $\{w,u\}$ sont des arêtes de G.

Pour tout $k \in [\![1, r]\!]$, on note Y_k la variable aléatoire de Bernoulli associée à l'événement « t_k est un triangle de G » et Z_n la variable aléatoire égale au nombre de triangles de G.

Par exemple si $n = 5$ et le graphe G est représenté ainsi,



alors $Z_5 = 3$.

1. Quelle est la valeur de r en fonction de n ?

2. ***a)*** Soit $k \in [\![1, r]\!]$. Posons $t_k = \{u, v, w\}$ avec $u < v < w$. Montrer que $Y_k = T_{u,v}T_{v,w}T_{u,w}$.

b) En déduire que, pour tout $k \in [\![1, r]\!]$, Y_k suit la loi de Bernoulli de paramètre p^3.

c) Justifier que $Z_n = \sum\limits_{k=1}^{r} Y_k$. En déduire que $\mathbb{E}(Z_n) = \dbinom{n}{3} p^3$.

► On s'intéresse à la variance de Z_n.
Si i et j appartiennent à $[\![1, r]\!]$ et sont différents, on note $i \equiv j$ lorsque t_i et t_j ont exactement deux éléments en commun et $i \not\equiv j$ dans le cas contraire.
On note $\mathcal{E}$ l'ensemble des couples (i, j) tels que $i \equiv j$, et $\mathcal{F}$ l'ensemble des couples (i, j) tels que $i \neq j$ et $i \not\equiv j$.
On désigne par a_n le nombre d'éléments de $\mathcal{E}$.

3. ***a)*** Montrer que :

$$\mathbb{V}(Z_n) = \mathrm{Cov}\left(\sum_{i=1}^{r} Y_i, \sum_{j=1}^{r} Y_j\right) = \sum_{(i,j)\in[\![1,r]\!]^2} \mathrm{Cov}(Y_i, Y_j)$$

b) Montrer que si $(i, j) \in \mathcal{F}$, Y_i et Y_j sont indépendantes. En déduire que :

$$\mathbb{V}(Z_n) = \sum_{i=1}^{r} \mathbb{V}(Y_i) + \sum_{(i,j)\in\mathcal{E}} \mathrm{Cov}(Y_i, Y_j)$$

c) En conclure que :

$$\mathbb{V}(Z_n) = rp^3(1 - p^3) + \left(\sum_{(i,j)\in\mathcal{E}} \mathbb{E}(Y_iY_j)\right) - a_n p^6$$

► On note $\Delta_n = \sum\limits_{(i,j)\in\mathcal{E}} \mathbb{E}(Y_iY_j)$

4. Montrer que si $i \equiv j$, $\mathbb{E}(Y_iY_j) = p^5$ et en déduire que $\Delta_n = a_n p^5$.
En conclure que : $\mathbb{V}(Z_n) = \dbinom{n}{3}(p^3 - p^6) + a_n(p^5 - p^6)$.

5. Calcul de a_n.

a) Déterminer le nombre de triplets $(\{u, v\}, w, y)$ où u, v, w, y sont quatre éléments distincts de l'ensemble $[\![0, n-1]\!]$.

b) En déduire que $a_n = \dfrac{n(n-1)(n-2)(n-3)}{2}$.

Partie 2 - Étude informatique

On se donne un graphe G généré par le procédé décrit dans le préambule.

On définit la fonction `supprimeDer(L)` qui si `L` est la liste des listes d'adjacence du graphe G dont les sommets sont $0, 1, \ldots, n-1$, modifie `L` afin qu'elle devienne la liste des listes d'adjacence du graphe G', dont les sommets sont $0, 1, \ldots, n-2$, obtenu en supprimant dans G le sommet $n-1$ et les arêtes contenant ce sommet.

```
def supprimeDer(L):
    s = len(L) - 1
    L.pop() # supprime le dernier élément de la liste L
    for a in L:
        if s in a:
            a.remove(s) # supprime s dans la liste a
```

6. Compléter la fonction suivante pour qu'elle retourne le nombre de triangles dont un des sommets est le sommet `s` dans le graphe G dont la liste des listes d'adjacence est `L` :

```
def triangle2s(s, L):
    cpt = 0
    adj = L[s]
    for i in range(len(adj)):
        for j in range(..., len(adj)) :
            if ... in L[...]:
                cpt += 1
    return cpt
```

7. Écrire une fonction `nbTriangles(L)`, utilisant les deux fonctions précédentes, qui retourne le nombre de triangles du graphe G dont la liste des listes d'adjacence est représentée par L.

8. On suppose que la fonction `graphe(n,p)` génère un graphe aléatoire suivant les hypothèses décrites dans le préambule.

Expliquer ce que retourne la fonction suivante :

```
def fonctionMystere(n):
    cpt = 0
    for i in range(1000):
        L = graphe(n, 1/n)
        if nbTriangles(L) == 0:
            cpt += 1
    return cpt / 1000
```

Partie 3 - Inégalité de Harris

k désigne un entier naturel non nul.

- Soit f une fonction définie sur une partie $\mathcal{D}$ de $\mathbb{R}^k$ à valeurs dans $\mathbb{R}$.
  Si $k \geqslant 2$, on dit que f est k-croissante sur $\mathcal{D}$ si, pour tout $(x_1, \ldots, x_k)$ élément de $\mathcal{D}$ et $i \in [\![1, k]\!]$, $t \mapsto f(x_1, \ldots, x_{i-1}, t, x_{i+1}, \ldots, x_k)$ est croissante sur son ensemble de définition.
  Si $k = 1$, une fonction 1-croissante sur $\mathcal{D}$ est simplement une fonction croissante sur $\mathcal{D}$.
- On définit de même la notion de fonction k-décroissante.
- On considère $X_1, \ldots, X_k$ des variables aléatoires finies.
  On admet le résultat suivant (théorème de transfert d'ordre k) :
  Si f est une fonction définie sur $X_1(\Omega) \times \cdots \times X_k(\Omega)$ et $Y_k = f(X_1, \ldots, X_k)$ alors

$$\mathbb{E}(Y_k) = \sum_{(x_1,\ldots,x_k) \in X_1(\Omega) \times \cdots \times X_k(\Omega)} f(x_1, \ldots, x_k) \mathbb{P}\left([X_1 = x_1] \cap \cdots \cap [X_k = x_k]\right)$$

On note (H_k) la propriété suivante :

Si $X_1, \ldots, X_k$ sont des variables aléatoires finies indépendantes, f et g deux fonctions définies sur $X_1(\Omega) \times \cdots \times X_k(\Omega)$ et k-croissantes sur cet ensemble, et si l'on pose $Y_k = f(X_1, \ldots, X_k)$ et $Z_k = g(X_1, \ldots, X_k)$, alors :

$$\mathbb{E}(Y_k Z_k) \geqslant \mathbb{E}(Y_k)\mathbb{E}(Z_k) \quad \text{(inégalité de Harris)}$$

9. Dans cette question $k = 1$, on pose $X = X_1$ une variable aléatoire finie, f et g sont deux fonctions croissantes sur $X(\Omega)$.

a) Montrer que pour tout $(x, y) \in (X(\Omega))^2$, $(f(x) - f(y))(g(x) - g(y)) \geqslant 0$.

b) Montrer que pour tout $y \in X(\Omega)$,

$$\mathbb{E}(f(X)g(X)) + f(y)g(y) \geqslant g(y)\mathbb{E}(f(X)) + f(y)\mathbb{E}(g(X))$$

c) En déduire que (H_1) est vraie.

10. On suppose que (H_k) est vraie pour un certain k et on considère $X_1, \ldots, X_{k+1}$ des variables aléatoires finies indépendantes, f et g deux fonctions définies sur $X_1(\Omega) \times \cdots \times X_{k+1}(\Omega)$ et $(k+1)$-croissantes. On pose $Y_{k+1} = f(X_1, \ldots, X_{k+1})$ et $Z_{k+1} = g(X_1, \ldots, X_{k+1})$.

a) A l'aide des théorèmes de transfert d'ordre $k + 1$ et k, montrer que :

$$\mathbb{E}(Y_{k+1}Z_{k+1}) = \sum_{x \in X_{k+1}(\Omega)} \mathbb{E}(f(X_1, \ldots, X_k, x)g(X_1, \ldots, X_k, x))\mathbb{P}([X_{k+1} = x])$$

b) Justifier que pour tout $x \in X_{k+1}(\Omega)$:

$$\mathbb{E}(f(X_1, \ldots, X_k, x)g(X_1, \ldots, X_k, x)) \geqslant \mathbb{E}(f(X_1, \ldots, X_k, x))\mathbb{E}(g(X_1, \ldots, X_k, x))$$

c) On pose pour tout $x \in X_{k+1}(\Omega)$, $u(x) = \mathbb{E}(f(X_1, \ldots, X_k, x))$ et $v(x) = \mathbb{E}(g(X_1, \ldots, X_k, x))$. Montrer que u et v sont croissantes sur $X_{k+1}(\Omega)$ et $\mathbb{E}(Y_{k+1}Z_{k+1}) \geqslant \mathbb{E}(u(X_{k+1})v(X_{k+1}))$.

d) En conclure que (H_{k+1}) est vraie. Conclure.

e) La propriété (H_k) reste-t-elle vraie si f et g sont k-décroissantes ? Justifier votre réponse. Que se passe-t-il si l'une est k-croissante et l'autre k-décroissante ?

Partie 4 - Inégalité de Janson et application

On reprend les notations de la partie 1.

De plus, pour tout $i \in [\![1, r]\!]$, on pose $Z_{n,i} = \sum_{k=1}^{i} Y_k$. On remarquera que $Z_{n,r} = Z_n$.

Dans cette partie on établit un encadrement de $\mathbb{P}([Z_n = 0])$.

11. Justifier que $\bigcap_{0 \leqslant u < v \leqslant n-1} [T_{u,v} = 0] \subset [Z_n = 0]$. En déduire que $\mathbb{P}([Z_n = 0]) \geqslant (1 - p)^{\binom{n}{2}} > 0$.

12. Montrer que pour tout $i \in [\![1, r]\!]$, $\mathbb{P}([Y_i = 0]) = \mathbb{E}(1 - Y_i)$ et $\mathbb{P}([Z_{n,i} = 0]) = \mathbb{E}\left(\prod_{k=1}^{i}(1 - Y_k)\right)$.

13. ***a)*** On pose $m = \binom{n}{2}$. Justifier brièvement que, pour tout $k \in [\![1, r]\!]$, Y_k s'exprime comme une fonction m-croissante sur $\{0, 1\}^m$ des variables aléatoires $T_{u,v}$ pour $u < v$ éléments de $[\![0, n-1]\!]$.

En déduire que, pour tout $i \in [\![2, r]\!]$, $1 - Y_i$ puis $\prod_{k=1}^{i-1}(1 - Y_k)$ s'expriment comme des fonctions m-décroissantes des variables aléatoires $T_{u,v}$ pour $u < v$ éléments de $[\![0, n - 1]\!]$.

b) En conclure que, pour tout $i \in [\![2, r]\!]$, $\mathbb{P}([Z_{n,i} = 0]) \geqslant \mathbb{P}([Z_{n,i-1} = 0])\mathbb{P}([Y_i = 0])$ puis que :

$$\mathbb{P}([Z_n = 0]) \geqslant \prod_{k=1}^{r} \mathbb{P}([Y_k = 0]) \qquad \text{puis} \qquad \mathbb{P}([Z_n = 0]) \geqslant (1 - p^3)^{\binom{n}{3}}$$

14. *Inégalité de Boole.* Montrer par récurrence sur $k \geqslant 2$ que si $B_1, \ldots, B_k$ sont des événements, on a :

$$\mathbb{P}\left(\bigcup_{i=1}^{k} B_i\right) \leqslant \sum_{i=1}^{k} \mathbb{P}(B_i)$$

- Si A est un événement de probabilité non nulle, on rappelle que la probabilité conditionnelle sachant A est notée $\mathbb{P}_A$. On admet qu'elle possède les mêmes propriétés que la probabilité $\mathbb{P}$.
  En particulier l'inégalité de Boole est vérifiée par $\mathbb{P}_A$.
  De plus si X est une variable finie, on note $\mathbb{E}_A(X)$ l'espérance de X pour la probabilité $\mathbb{P}_A$ ce qui signifie que :
  $$\mathbb{E}_A(X) = \sum_{x \in X(\Omega)} x\mathbb{P}_A([X = x])$$
  Cette espérance conditionnelle possède les mêmes propriétés que l'espérance en particulier l'inégalité de Harris vue dans la partie 3.

15. Soit A, B et C trois événements tels que $\mathbb{P}(B \cap C) \neq 0$ et $\mathbb{P}(A \cap C) \neq 0$.
Montrer que $\mathbb{P}_{B\cap C}(A) \geqslant \mathbb{P}_C(A)\mathbb{P}_{A\cap C}(B)$.

- On admet que les probabilités conditionnelles qui interviennent dans la suite sont bien définies.

16. Pour tout $i \in [\![1, r]\!]$, on pose $A_i = [Y_i = 0]$.
On note aussi $I_i = \{j \in [\![1, i-1]\!] \mid j \equiv i\}$ et $J_i = \{j \in [\![1, i-1]\!] \mid j \not\equiv i\}$.
Soit $i \geqslant 2$, on définit $B_i = \bigcap_{j \in I_i} A_j$ et $C_i = \bigcap_{j \in J_i} A_j$, ainsi on a : $B_i \cap C_i = \bigcap_{j=1}^{i-1} A_j$.

a) Justifier que A_i et C_i sont indépendants. En déduire que $\mathbb{P}_{B_i\cap C_i}(\overline{A_i}) \geqslant \mathbb{P}(\overline{A_i})\mathbb{P}_{\overline{A_i}\cap C_i}(B_i)$.

b) Établir que $\mathbb{P}_{\overline{A_i}\cap C_i}(B_i) \geqslant 1 - \sum_{j \in I_i} \mathbb{P}_{\overline{A_i}\cap C_i}(\overline{A_j})$.

c) **On admet provisoirement que pour** $j \in [\![1, i-1]\!]$:
$$\mathbb{P}_{\overline{A_i}\cap C_i}(\overline{A_j}) \leqslant \mathbb{P}_{\overline{A_i}}(\overline{A_j}) \qquad (1)$$
En déduire que, $\mathbb{P}_{B_i\cap C_i}(A_i) \leqslant 1 - \mathbb{P}(\overline{A_i}) \left(1 - \sum_{j\in I_i} \mathbb{P}_{\overline{A_i}}(\overline{A_j})\right)$.

d) Justifier que pour tout $x \in \mathbb{R}$, $1 - x \leqslant \exp(-x)$ et en déduire que :
$$\mathbb{P}_{B_i\cap C_i}(A_i) \leqslant \exp\left(-\mathbb{P}(\overline{A_i}) + \sum_{j\in I_i} \mathbb{P}(\overline{A_j} \cap \overline{A_i})\right)$$

17. On rappelle que $\Delta_n = \sum_{(i,j)\in\mathcal{E}} \mathbb{E}(Y_iY_j)$ où $\mathcal{E}$ a été défini dans la partie 1 à la suite de la question ***2***.

a) Montrer que $\mathbb{P}([Z_n = 0]) = \mathbb{P}\left(\bigcap_{i=1}^{r} A_i\right) = \mathbb{P}(A_1) \prod_{i=2}^{r} \mathbb{P}_{B_i\cap C_i}(A_i)$.

b) En conclure que :
$$\mathbb{P}([Z_n = 0]) \leqslant \exp\left(-\mathbb{E}(Z_n) + \frac{\Delta_n}{2}\right) \quad \text{(inégalité de Janson)}$$

c) En déduire l'encadrement :
$$(1 - p^3)^{\binom{n}{3}} \leqslant \mathbb{P}([Z_n = 0]) \leqslant \exp\left(-\binom{n}{3}p^3 + \frac{a_n}{2}p^5\right)$$

18. Soit c un réel strictement positif.

a) Montrer que $\lim_{n\to+\infty} -\binom{n}{3}\left(\frac{c}{n}\right)^3 + \frac{a_n}{2}\left(\frac{c}{n}\right)^5 = -\frac{c^3}{6}$.

b) Établir que $\lim_{n\to+\infty} \binom{n}{3} \ln\left(1 - \frac{c^3}{n^3}\right) = -\frac{c^3}{6}$.

c) On suppose que $n > c$ et $p = \dfrac{c}{n}$. En déduire la limite de $\mathbb{P}([Z_n = 0])$ quand $n \to +\infty$.

19. On reprend les notations de la partie 2. L'exécution de l'instruction `fonctionMystere(100)` affiche dans la console **Python** `0.849`. Est-ce cohérent avec le résultat de la question précédente si on considère que pour x assez petit, e^{-x} est proche de $1 - x + \dfrac{x^2}{2}$?

20. *Démonstration de (1)* - Soit m un entier plus grand que 2. On considère $X_1, \ldots, X_m$ des variables de Bernoulli indépendantes et I un sous ensemble de $[\![1, m]\!]$. On note J le complémentaire de I dans $[\![1, m]\!]$. On note A l'événement $\left[\prod_{i \in I} X_i = 1\right]$.

a) Montrer que, pour tout $(x_1, \ldots, x_m) \in \{0, 1\}^m$,

$$\mathbb{P}_A([X_1 = x_1] \cap \cdots \cap [X_m = x_m]) = \begin{cases} \prod_{i \in J} \mathbb{P}([X_i = x_i]) & \text{si } \prod_{i \in I} x_i = 1 \\ 0 & \text{sinon.} \end{cases}$$

b) En déduire que les variables aléatoires $X_1, \ldots, X_m$ sont indépendantes pour la probabilité conditionnelle $\mathbb{P}_A$.

► Soit $i \geqslant 2$, on reprend les notations de la question ***16***.

c) Montrer que pour $j \in [\![1, i-1]\!]$, $\mathbb{P}_{\overline{A_i} \cap C_i}(\overline{A_j}) = \dfrac{\mathbb{E}_{\overline{A_i}}\left(Y_j \prod_{k \in J_i} (1 - Y_k)\right)}{\mathbb{P}_{\overline{A_i}}(C_i)}$.

d) En utilisant l'inégalité de Harris, montrer que pour $j \in [\![1, i-1]\!]$:

$$\mathbb{P}_{\overline{A_i} \cap C_i}(\overline{A_j}) \leqslant \mathbb{P}_{\overline{A_i}}(\overline{A_j})$$

ESSEC II 2024 - énergie d'une matrice, trace d'une matrice, produit de Kronecker, réduction des matrices symétriques, graphe complet

On s'intéresse dans ce problème à l'énergie d'un graphe qui est définie à partir de l'énergie de sa matrice d'adjacence.

L'énergie d'un graphe a été introduite en 1978 par Ivan Gutman. Ce n'est qu'à partir des années 2000 que des recherches approfondies sur cette notion ont été entreprises.

Aujourd'hui plus de deux articles par semaine sont publiés sur l'énergie des graphes avec de nombreuses applications dans divers domaines scientifiques.

Les parties 2 et 3 sont indépendantes et un bref aide-mémoire Python se trouve en page 5.

Dans tout le problème n est un entier supérieur ou égal à 2.

On rappelle que la notation $\sum\limits_{1\leqslant i,j\leqslant n}$ est analogue à la notation $\sum\limits_{(i,j)\in[\![1,n]\!]^2}$.

Partie 1 - Énergie et trace d'une matrice

1. Soit A une matrice carrée symétrique appartenant à $\mathscr{M}_n(\mathbb{R})$. Justifier l'existence d'une matrice carrée inversible P appartenant à $\mathscr{M}_n(\mathbb{R})$ telle que $P^{-1}AP$ soit une matrice diagonale. Que peut-on dire des éléments diagonaux de $P^{-1}AP$?

► *En notant $\lambda_1, \ldots, \lambda_n$ les éléments diagonaux de $P^{-1}AP$, on pose alors $\mathcal{E}(A) = \sum\limits_{k=1}^{n} |\lambda_k|$, on nomme ce réel positif l'énergie de A.*
On admet que cette somme ne dépend pas du choix de P.

2. Montrer que $\mathcal{E}(A) = 3$ si $A = \begin{pmatrix} 1 & -1 & -1 \\ -1 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix}$.

3. Ecrire une fonction **Python** `energie(A)` qui renvoie l'énergie de la matrice symétrique représentée par le tableau `numpy A`.

► *Si $A = (a_{i,j})_{1\leqslant i,j\leqslant n}$ est une matrice carrée appartenant à $\mathscr{M}_n(\mathbb{R})$, on définit la trace de A, notée $\mathrm{tr}(A)$ par :*

$$\mathrm{tr}(A) = \sum_{i=1}^{n} a_{i,i}$$

4. Notons $A = (a_{i,j})_{1\leqslant i,j\leqslant n}$ et $B = (b_{i,j})_{1\leqslant i,j\leqslant n}$.

a) Montrer que : $\mathrm{tr}(AB) = \sum\limits_{i=1}^{n}\left(\sum\limits_{j=1}^{n} a_{i,j}b_{j,i}\right) = \sum\limits_{1\leqslant i,j\leqslant n} a_{i,j}b_{j,i}$.

b) En déduire que : $\mathrm{tr}(AB) = \mathrm{tr}(BA)$ et $\mathrm{tr}({}^tAA) = \sum\limits_{1\leqslant i,j\leqslant n} a_{j,i}^2$.

Que peut-on dire de A si $\mathrm{tr}({}^tAA) = 0$?

c) Si A et B sont semblables, montrer que $\mathrm{tr}(A) = \mathrm{tr}(B)$.

5. Dans cette question $A = (a_{i,j})_{1\leqslant i,j\leqslant n}$ est symétrique et on utilise les notations de la question *1*.

a) Montrer que $|\mathrm{tr}(A)| \leqslant \mathcal{E}(A)$.

b) Justifier que $\mathrm{tr}(A^2) = \sum\limits_{k=1}^{n} \lambda_k^2$.

6. Dans la console **Python**, on obtient :

```
> > > energie(3*np.eye(3)-np.ones([3,3]))
6.0
```

a) Déterminer quelle est la matrice A associée au tableau `3*np.eye(3)-np.ones([3,3])`.

b) En calculant $\operatorname{tr}(A)$, $\operatorname{tr}(A^2)$ et en déterminant une valeur propre évidente de A, expliciter son spectre et retrouver son énergie.

Partie 2 - Produit de Kronecker $2 \times n$ de matrices symétriques

- Soit $U = \begin{pmatrix} u & v \\ v & w \end{pmatrix}$ une matrice carrée appartenant à $\mathscr{M}_2(\mathbb{R})$ symétrique et A une matrice carrée appartenant à $\mathscr{M}_n(\mathbb{R})$ symétrique.
  On définit $U * A$ la matrice carrée appartenant à $\mathscr{M}_{2n}(\mathbb{R})$ que l'on peut naturellement représenter ainsi $\begin{pmatrix} uA & vA \\ vA & wA \end{pmatrix}$ (écriture par blocs).

- Par exemple, si $U = \begin{pmatrix} 2 & 1 \\ 1 & 0 \end{pmatrix}$ et $A = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}$, alors $U * A = \begin{pmatrix} 2A & A \\ A & 0_3 \end{pmatrix}$ par blocs,
  d'où $U * A = \begin{pmatrix} 0 & 2 & 2 & 0 & 1 & 1 \\ 2 & 0 & 0 & 1 & 0 & 0 \\ 2 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$ (0_3 désigne la matrice nulle de $\mathscr{M}_3(\mathbb{R})$).

- Si X est une matrice colonne appartenant à $\mathscr{M}_{2n,1}(\mathbb{R})$, $X = \begin{pmatrix} x_1 \\ \vdots \\ x_{2n} \end{pmatrix}$, on écrira $X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$ avec $X_1 = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ et $X_2 = \begin{pmatrix} x_{n+1} \\ \vdots \\ x_{2n} \end{pmatrix}$.
  On admet qu'alors la matrice colonne $(U * A)X$ de $\mathscr{M}_{2n,1}(\mathbb{R})$ est égale à $\begin{pmatrix} uAX_1 + vAX_2 \\ vAX_1 + wAX_2 \end{pmatrix}$.

7. ***a)*** Écrire une fonction **Python** `prod2K(u,v,w,A)` qui étant donné $U = \begin{pmatrix} u & v \\ v & w \end{pmatrix}$ et A, représentée par un tableau `numpy`, renvoie $U * A$ sous la forme d'un tableau `numpy`.

b) Compléter le code suivant s'affichant dans la console **Python** :

```
> > > prod2K(...,-1,...,...)
array([[-2., 1., 2., -1.],
       [1., -2., -1., 2.],
       [2., -1., -4., 2.],
       [-1., 2., 2., -4.]])
```

8. Soit $\begin{pmatrix} a \\ b \end{pmatrix}$ un vecteur propre de U pour la valeur propre λ et X un vecteur propre de A pour μ. On pose $Y = \begin{pmatrix} aX \\ bX \end{pmatrix}$.

a) Établir l'égalité : $(U * A)Y = \mu \begin{pmatrix} (ua + vb)X \\ (va + wb)X \end{pmatrix}$.

b) Montrer que Y est un vecteur propre de $U * A$ et préciser pour quelle valeur propre.

9. ***a)*** Justifier l'existence d'une base $\left(\begin{pmatrix} a \\ b \end{pmatrix}, \begin{pmatrix} c \\ d \end{pmatrix} \right)$ de $\mathscr{M}_{2,1}(\mathbb{R})$ et de $(X_1, \ldots, X_n)$ une base de $\mathscr{M}_{n,1}(\mathbb{R})$, formée de vecteurs propres de U et A respectivement.

- *On note γ_1 et γ_2 les valeurs propres associées à $\begin{pmatrix} a \\ b \end{pmatrix}$ et $\begin{pmatrix} c \\ d \end{pmatrix}$ respectivement et $\mu_1, \dots, \mu_n$ celles associées à $X_1, \dots, X_n$ respectivement.*

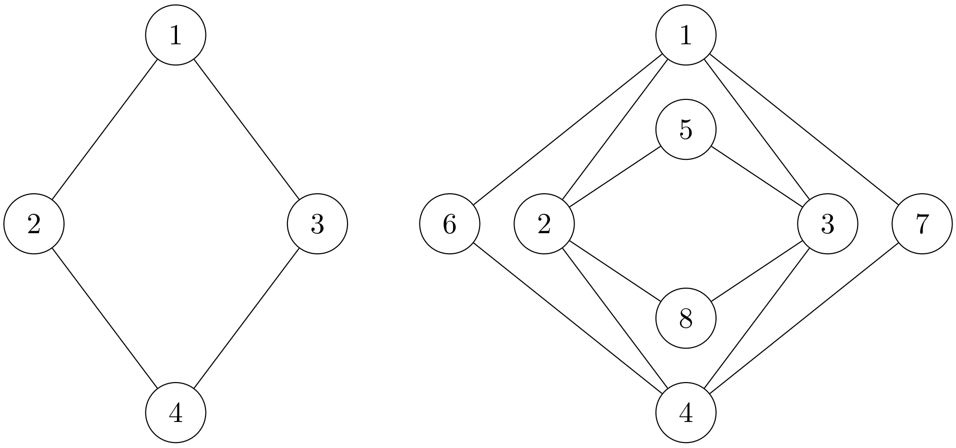
  On pose, pour tout $i \in [\![1, n]\!]$, $Y_i = \begin{pmatrix} aX_i \\ bX_i \end{pmatrix}$ et $Z_i = \begin{pmatrix} cX_i \\ dX_i \end{pmatrix}$.

b) Montrer que la famille $(Y_1, Z_1, \dots, Y_n, Z_n)$ est libre.

c) En déduire que $U * A$ est semblable à la matrice diagonale dont les éléments diagonaux sont $\gamma_1\mu_1, \gamma_2\mu_1, \dots, \gamma_1\mu_n, \gamma_2\mu_n$.

d) En conclure que $\mathcal{E}(U * A) = \mathcal{E}(U) \times \mathcal{E}(A)$.

10. *Un exemple* - On considère un graphe G_n, non orienté et sans boucle, dont les sommets sont $1, \dots, n$ et l'ensemble des arêtes est noté $\mathcal{A}_n$. On définit le graphe G_{2n} dont les sommets sont $1, \dots, 2n$ et les arêtes sont celles de G_n, ainsi que, pour toute arête $\{i, j\} \in \mathcal{A}_n$, les arêtes $\{i+n, j\}$ et $\{i, j+n\}$. Voici un exemple de représentation de G_4 et G_8 :



On note A_n et A_{2n} les matrices d'adjacence de G_n et G_{2n}.

a) Déterminer U telle que $A_{2n} = U * A_n$.

b) En déduire que $\mathcal{E}(A_{2n}) = \sqrt{5}\ \mathcal{E}(A_n)$.

Partie 3 - Encadrement de l'énergie d'une matrice d'adjacence

Soit m, n et p des entiers tels que $m \geqslant 1$, $n \geqslant p \geqslant 2$.

On suppose dans cette partie que $A = (a_{i,j})_{1 \leqslant i,j \leqslant n}$ est la matrice d'adjacence d'un graphe $G(A)$, non orienté sans boucle, à n sommets $1, \dots, n$, m arêtes et $n - p$ sommets isolés, c'est-à-dire de degré 0.

On note $(E_1, \dots, E_n)$ la base canonique de $\mathscr{M}_{n,1}(\mathbb{R})$ et I l'ensemble des sommets non isolés de $G(A)$.

11. ***a)*** Montrer que $\displaystyle\sum_{1 \leqslant i,j \leqslant n} a_{i,j} = \sum_{(i,j) \in I^2} a_{i,j} = 2m$.

b) Établir que : $1 \leqslant \dfrac{2m}{p} \leqslant p - 1$.

12. ***a)*** Justifier qu'il existe une matrice carrée inversible P appartenant à $\mathscr{M}_n(\mathbb{R})$ telle que $P^{-1}AP$ soit une matrice diagonale différente de la matrice nulle.

- Dans la suite on note D cette matrice diagonale.

b) En déduire que $\mathrm{tr}(D) = 0$, $\mathrm{tr}(D^2) = \displaystyle\sum_{1 \leqslant i,j \leqslant n} a_{i,j}^2 = 2m$.

- *On suppose dans la suite que P est telle que les éléments diagonaux de D, $\lambda_1, \dots, \lambda_n$ vérifient $|\lambda_1| \geqslant \cdots \geqslant |\lambda_n|$. On pose $\theta = \max\limits_{1 \leqslant k \leqslant n} \lambda_k$.*

13. ***a)*** Soit k un sommet isolé. Montrer que $E_k \in \ker(A)$.
En déduire que $\dim(\ker(A)) \geqslant n - p$ puis que, si $p < n$, $\lambda_{p+1} = \cdots = \lambda_n = 0$.

b) Montrer que $\sum\limits_{k=1}^{p} \lambda_k^2 = 2m$ puis que $0 < \theta < \sqrt{2m}$.

c) Soit $r \in \mathbb{N}^*$ et $x_1, \ldots, x_r$ des réels. Montrer que : $\sum\limits_{1 \leqslant i,j \leqslant r} 2x_i x_j \leqslant \sum\limits_{1 \leqslant i,j \leqslant r} (x_i^2 + x_j^2)$.

En déduire que $\left(\sum\limits_{i=1}^{r} x_i\right)^2 \leqslant r \left(\sum\limits_{i=1}^{r} x_i^2\right)$.

d) En conclure que $\mathcal{E}(A) \leqslant \theta + \sqrt{(p-1)(2m - \theta^2)}$.

14. On admet qu'on peut choisir la matrice P de la question ***12.a)*** de sorte que $P^{-1} = {}^tP$.
On pose alors $Q = {}^tP$.

Soit $X = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ une matrice colonne appartenant à $\mathscr{M}_{n,1}(\mathbb{R})$ et $Y = QX = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$.

- *On admet que si M est une matrice carrée appartenant à $\mathscr{M}_n(\mathbb{R})$ et Z une matrice colonne appartenant à $\mathscr{M}_{n,1}(\mathbb{R})$ alors ${}^t(MZ) = {}^tZ{}^tM$.*
- *Si U et V sont deux matrices colonnes appartenant à $\mathscr{M}_{n,1}(\mathbb{R})$, ${}^tUV$ est une matrice carrée appartenant à $\mathscr{M}_{1,1}(\mathbb{R})$, on l'identifie à son unique coefficient. Donc ${}^tUV \in \mathbb{R}$.*

a) Montrer que $Q^{-1} = {}^tQ$ puis que $A = {}^tQDQ$ et ${}^tYY = {}^tXX$.

b) Montrer que ${}^tXAX = {}^tYDY = \sum\limits_{k=1}^{n} \lambda_k y_k^2$.

c) En remarquant que $\sum\limits_{k=1}^{n} y_k^2 = {}^tYY$, en déduire que ${}^tXAX \leqslant \theta\, {}^tXX$.

15. Soit U la matrice colonne appartenant à $\mathscr{M}_{n,1}(\mathbb{R})$ dont le $i$ème coefficient vaut 1 si i n'est pas isolé et 0 sinon.

a) Montrer que ${}^tUAU = \sum\limits_{(i,j)\in I^2} a_{i,j} = 2m$. En déduire que $\dfrac{2m}{p} \leqslant \theta$.

b) Établir que : $\dfrac{1}{\sqrt{p}} \leqslant \dfrac{\sqrt{2m}}{p} \leqslant \dfrac{\theta}{\sqrt{2m}} \leqslant 1$.

16. ***a)*** Étudier la fonction $F : x \mapsto x + \sqrt{(p-1)(1-x^2)}$ sur $[0,1]$.

b) En déduire que $\mathcal{E}(A) \leqslant \sqrt{2m} F\left(\dfrac{\sqrt{2m}}{p}\right)$, c'est-à-dire que :

$$\mathcal{E}(A) \leqslant \frac{2m}{p} + \frac{1}{p}\sqrt{(p-1)2m(p^2 - 2m)} \qquad (1)$$

17. On suppose dans cette question que A est la matrice d'adjacence d'un graphe complet de sommets $1, \ldots, n$ donc $m = \dfrac{n(n-1)}{2}$ et $p = n$.

a) Représenter la matrice A.

b) Montrer que -1 est une valeur propre de A et que le sous-espace propre associé est de dimension $n-1$.

c) Établir aussi que $n-1$ est une valeur propre de A. En déduire que $\mathcal{E}(A) = 2(n-1)$ et que l'inégalité (1) est alors une égalité.

- *On note $\alpha = \min\limits_{1 \leqslant k \leqslant p} |\lambda_k|$ et $\beta = \max\limits_{1 \leqslant k \leqslant p} |\lambda_k|$. On note aussi d le degré maximal des sommets du graphe $G(A)$.*

18. Écrire une fonction **Python** `degMax(A)` qui renvoie le maximum des degrés des sommets du graphe $G(A)$, celui-ci étant donné par sa matrice d'adjacence sous la forme du tableau `numpy A`.

19. ***a)*** Montrer que pour tout $j \in \{1, \dots, p\}$, $|\lambda_j|\,(\alpha+\beta) \geqslant \lambda_j^2 + \alpha\beta$.

b) En déduire que $(\alpha+\beta)\mathcal{E}(A) \geqslant \mathrm{tr}(A^2) + p\alpha\beta$, puis que $\mathcal{E}(A) \geqslant \dfrac{2m + p\alpha\beta}{\alpha+\beta}$.

c) Soit a et b deux réels strictement positifs tels que $a \leqslant b^2$. Étudier les variations de $\varphi : t \mapsto \dfrac{a+bt}{b+t}$ sur $\mathbb{R}^+$.

d) Montrer que $2m \leqslant p\beta^2$. En déduire que $\mathcal{E}(A) \geqslant \dfrac{2m}{\beta}$.

20. Soit $X = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ un vecteur propre pour une valeur propre λ de A telle que $|\lambda| = \beta$.

a) Montrer que pour tout $i \in \{1, \dots, n\}$, $\beta\,|x_i| \leqslant d \max\limits_{1 \leqslant j \leqslant n} |x_j|$.

b) En conclure que $\mathcal{E}(A) \geqslant \dfrac{2m}{d} \geqslant \dfrac{2m}{p-1} \quad (2)$.

c) Montrer que l'égalité dans (2) est réalisée pour la matrice A carrée appartenant à $\mathscr{M}_n(\mathbb{R})$ associée au graphe dont l'unique arête est $\{1, 2\}$.

Aide-mémoire Python

On suppose que l'on a exécuté `import numpy as np, numpy.linalg as al` en début de session.

- Si `T` est un tableau `numpy`, `np.shape(T)` renvoie le nombre de lignes et le nombre de colonnes de T, dans cet ordre, sous la forme d'un couple.
- `np.zeros([p,q])` crée un tableau `numpy` à p lignes et q colonnes ne contenant que des 0.
- `np.ones([p,q])` crée un tableau `numpy` à p lignes et q colonnes ne contenant que des 1.
- `np.eye(p)` crée un tableau `numpy` à p lignes et p colonnes ne contenant que des 1 sur sa diagonale et des 0 ailleurs.
- La fonction `al.eigvals`, appliquée à un tableau `numpy` représentant une matrice symétrique A, renvoie le tableau des coefficients d'une matrice diagonale semblable à A.

HEC 2010 - loi exponentielle, loi géométrique, loi de la somme, de la différence, de la valeur absolue, du $\max$, du $\min$, covariance, coefficient de corrélation linéaire, estimation, loi de Gumbel

- Toutes les variables aléatoires qui interviennent dans ce problème sont supposées définies sur un espace probabilisé $(\Omega, \mathscr{A}, \mathbb{P})$.
- Sous réserve d'existence, on note $\mathbb{E}(X)$ et $\mathbb{V}(X)$ respectivement l'espérance et la variance d'une variable aléatoire X, et $\mathrm{Cov}(X, Y)$ la covariance de deux v.a.r. X et Y.
- Dans les parties I et III, la fonction de répartition et une densité d'une variable aléatoire X à densité sont notées respectivement F_X et f_X.
- **On admet** que les formules donnant l'espérance et la variance d'une somme de variables aléatoires discrètes, ainsi que la définition et les propriétés de la covariance et du coefficient de corrélation linéaire de deux variables aléatoires discrètes, s'appliquent au cas de variables aléatoires à densité.
- Pour n entier supérieur ou égal à 2, on dit que les variables aléatoires à densité X_1, X_2, ..., X_n sont indépendantes si pour tout n-uplet $(x_1, x_2, \ldots, x_n)$ de réels, les événements $[X_1 \leqslant x_1]$, $[X_2 \leqslant x_2]$, ..., $[X_n \leqslant x_n]$ sont indépendants.
- L'objet du problème est double. D'une part, montrer certaines analogies entre les lois géométriques et exponentielles, d'autre part mettre en évidence quelques propriétés asymptotiques de variables aléatoires issues de la loi exponentielle.
  La partie II est indépendante de la partie I.
  La partie III est indépendante de la partie II et largement indépendante de la partie I.

Partie I. Loi exponentielle

1. ***a)*** Rappeler la valeur de $\displaystyle\int_0^{+\infty} \mathrm{e}^{-t}\, dt$.

Établir pour tout n de $\mathbb{N}^*$ la convergence de l'intégrale $\displaystyle\int_0^{+\infty} t^n \mathrm{e}^{-t}\, dt$.

On pose alors $I_0 = \displaystyle\int_0^{+\infty} \mathrm{e}^{-t}\, dt$ et pour tout n de $\mathbb{N}^*$ $I_n = \displaystyle\int_0^{+\infty} t^n \mathrm{e}^{-t}\, dt$.

b) Soit n un entier de $\mathbb{N}^*$. À l'aide d'une intégration par parties, établir une relation de récurrence entre I_n et I_{n-1}. En déduire la valeur de I_n en fonction de n.

Soit λ un réel strictement positif. Soient X_1 et X_2 deux variables indépendantes de même loi exponentielle de paramètre λ (d'espérance $\frac{1}{\lambda}$).

On pose : $Y = X_1 - X_2$, $T = \max(X_1, X_2)$ et $Z = \min(X_1, X_2)$.

2. Justifier les relations $T + Z = X_1 + X_2$ et $T - Z = |X_1 - X_2| = |Y|$.

3. ***a)*** Rappeler sans démonstration les valeurs de $\mathbb{V}(X_1)$ et de $\mathbb{P}\big([X_1 \leqslant x]\big)$, pour tout réel x.

b) Calculer $\mathbb{E}(X_1 + X_2)$, $\mathbb{V}(X_1 + X_2)$, $\mathbb{E}(Y)$, $\mathbb{V}(Y)$.

4. Déterminer pour tout réel z, $F_Z(z)$ et $f_Z(z)$. Reconnaître la loi de Z et en déduire $\mathbb{E}(Z)$ et $\mathbb{V}(Z)$.

5. ***a)*** Montrer que pour tout réel t, on a : $F_T(t) = \begin{cases} \left(1 - \mathrm{e}^{-\lambda t}\right)^2 & \text{si } t \geqslant 0 \\ 0 & \text{si } t < 0 \end{cases}$.

Exprimer pour tout réel t, $f_T(t)$.

b) Justifier l'existence de $\mathbb{E}(T)$ et $\mathbb{V}(T)$. Montrer que $\mathbb{E}(T) = \dfrac{3}{2\lambda}$ et $\mathbb{V}(T) = \dfrac{5}{4\lambda^2}$.
(on pourra utiliser des changements de variables affine)

6. On note r le coefficient de corrélation linéaire de Z et T. Montrer que $r = \dfrac{1}{\sqrt{5}}$.

7. ***a)*** Préciser $Y(\Omega)$ et $|Y|(\Omega)$.

b) Déterminer une densité de la variable aléatoire $-X_2$.

c) Montrer que pour tout réel y, l'intégrale $\displaystyle\int_{-\infty}^{+\infty} f_{X_1}(t)\, f_{-X_2}(y-t)\, dt$ est convergente et qu'elle vaut $\dfrac{\lambda}{2}\, \mathrm{e}^{-\lambda\,|y|}$ (on distinguera les deux cas : $y \geqslant 0$ et $y < 0$).

d) Établir que la fonction $y \mapsto \dfrac{\lambda}{2}\mathrm{e}^{-\lambda\,|y|}$ est une densité de probabilité sur $\mathbb{R}$; on admet que c'est une densité de la variable aléatoire Y.

e) Déterminer pour tout y réel, $f_{|Y|}(y)$. Reconnaître la loi de $|Y| = T - Z$.

Partie II. Loi géométrique

Soit p un réel de $]0,1[$ et $q = 1-p$. Soit X_1 et X_2 deux variables indépendantes de même loi géométrique de paramètre p (d'espérance $\frac{1}{p}$).
On pose : $Y = X_1 - X_2$, $T = \max(X_1, X_2)$ et $Z = \min(X_1, X_2)$.
On rappelle que $T + Z = X_1 + X_2$ et $T - Z = |X_1 - X_2| = |Y|$.

8. ***a)*** Rappeler sans démonstration les valeurs respectives de $\mathbb{V}(X_1)$ et de $\mathbb{P}\big([X_1 \leqslant k]\big)$, pour tout k de $X_1(\Omega)$.

b) Calculer $\mathbb{E}(X_1+X_2)$, $\mathbb{V}(X_1+X_2)$, $\mathbb{E}(X_1-X_2)$, $\mathbb{V}(X_1-X_2)$.

c) Établir la relation : $\mathbb{P}\big([X_1 = X_2]\big) = \dfrac{p}{1+q}$.

9. ***a)*** Montrer que Z suit la loi géométrique de paramètre $1-q^2$. En déduire $\mathbb{E}(Z)$, $\mathbb{V}(Z)$ et $\mathbb{E}(T)$.

b) Soit k un entier de $\mathbb{N}^*$. Justifier l'égalité : $[Z = k] \cup [T = k] = [X_1 = k] \cup [X_2 = k]$.
En déduire la relation suivante : $\mathbb{P}\big([T = k]\big) = 2\ \mathbb{P}\big([X_1 = k]\big) - \mathbb{P}\big([Z = k]\big)$.

c) Établir la formule : $\mathbb{V}(T) = \dfrac{q\ \left(2q^2+q+2\right)}{\left(1-q^2\right)^2}$.

10. ***a)*** Préciser $(T-Z)(\Omega)$.
Exprimer pour tout j de $\mathbb{N}^*$, l'événement $[Z = j] \cap [Z = T]$ en fonction des événements $[X_1 = j]$ et $[X_2 = j]$. En déduire pour tout j de $\mathbb{N}^*$, l'expression de $\mathbb{P}\big([Z = j] \cap [Z = T]\big)$.

b) Montrer que pour tout couple (j,l) de $(\mathbb{N}^*)^2$, on a : $\mathbb{P}\big([Z = j] \cap [T - Z = l]\big) = 2\ p^2\ q^{2j+l-2}$.

c) Montrer que pour tout k de $\mathbb{Z}$, $\mathbb{P}\big([X_1 - X_2 = k]\big) = \dfrac{pq^{|k|}}{1+q}$.
(on distinguera trois cas : $k = 0$, $k > 0$ et $k < 0$)

d) En déduire la loi de la variable aléatoire $|X_1 - X_2|$.

e) Établir à l'aide des questions précédentes que les variables Z et $T - Z$ sont indépendantes.

11. ***a)*** À l'aide du résultat de la question ***3.e***, calculer $\mathrm{Cov}(Z, T)$.
Les variables Z et T sont-elles indépendantes ?

b) Calculer en fonction de q, le coefficient de corrélation linéaire ρ de Z et T.

c) Déterminer la loi de probabilité du couple (Z, T).

d) Déterminer pour tout j de $\mathbb{N}^*$, la loi de probabilité conditionnelle de T sachant l'événement $[Z = j]$.

e) Soit j un élément de $\mathbb{N}^*$. On suppose qu'il existe une variable aléatoire D_j à valeur dans $\mathbb{N}^*$, dont la loi de probabilité est la loi conditionnelle de T sachant l'événement $[Z = j]$.
Calculer $\mathbb{E}\left(D_j\right)$.

Partie III. Convergences

Dans les questions 12 à 15, λ désigne un paramètre réel strictement positif, inconnu.
Pour n élément de $\mathbb{N}^*$, on considère un n-échantillon $(X_1, X_2, \ldots, X_n)$ de variables aléatoires à valeurs strictement positives, indépendantes, de même loi exponentielle de paramètre λ.
On pose pour tout n de $\mathbb{N}^*$: $S_n = \sum\limits_{k=1}^{n} X_k$ et $J_n = \lambda S_n$.

12. Calculer pour tout n de $\mathbb{N}^*$, $\mathbb{E}(S_n)$, $\mathbb{V}(S_n)$, $\mathbb{E}(J_n)$ et $\mathbb{V}(J_n)$.

13. On admet qu'une densité f_{J_n} de J_n est donnée par $f_{J_n}(x) = \begin{cases} \dfrac{\mathrm{e}^{-x}x^{n-1}}{(n-1)!} & \text{si } x > 0 \\ 0 & \text{si } x \leqslant 0 \end{cases}$.

a) À l'aide du théorème de transfert, établir pour tout n supérieur ou égal à 3, l'existence de $\mathbb{E}\left(\dfrac{1}{J_n}\right)$ et de $\mathbb{E}\left(\dfrac{1}{J_n^2}\right)$, et donner leur valeurs respectives.

b) On pose pour tout n supérieur ou égal à 3 : $\widehat{\lambda_n} = \dfrac{n}{S_n}$. Justifier que $\widehat{\lambda_n}$ est un estimateur de λ. Est-il sans biais ? Calculer la limite, lorsque n tend vers $+\infty$, du risque quadratique associé à $\widehat{\lambda_n}$ en λ.

14. Dans cette question, on veut déterminer un intervalle de confiance du paramètre λ au risque α.
On note Φ la fonction de répartition de la loi normale centrée réduite, et u_α le réel strictement positif tel que $\Phi(u_\alpha) = 1 - \dfrac{\alpha}{2}$.

a) Énoncer le théorème de la limite centrée. En déduire que la variable aléatoire N_n définie par $N_n = \lambda \dfrac{S_n}{\sqrt{n}} - \sqrt{n}$ converge en loi vers la loi normale centrée réduite.

b) En déduire que pour n assez grand, on a approximativement : $\mathbb{P}\big([-u_\alpha \leqslant N_n \leqslant u_\alpha]\big) = 1 - \alpha$.

c) Montrer que pour n assez grand, l'intervalle $\left[\left(1 - \dfrac{u_\alpha}{\sqrt{n}}\right)\widehat{\lambda_n}, \left(1 + \dfrac{u_\alpha}{\sqrt{n}}\right)\widehat{\lambda_n}\right]$ est un intervalle de confiance de λ au risque α. On note λ_0 la réalisation de $\widehat{\lambda_n}$ sur le n-échantillon.

15. Avec le n-échantillon $(X_1, X_2, \ldots, X_n)$, on construit un nouvel intervalle de confiance de λ au risque β $(\beta \neq \alpha)$, tel que la longueur de cet intervalle soit k $(k > 1)$ fois plus petite que celle obtenue avec le risque α.

a) Justifier l'existence de la fonction réciproque Φ^{-1} de Φ.
Quel est le domaine de définition de Φ^{-1} ?

b) Établir l'égalité $\beta = 2\Phi\left(\dfrac{1}{k}\Phi^{-1}\left(\dfrac{\alpha}{2}\right)\right)$.
En déduire que $\beta > \alpha$. Ce dernier résultat était-il prévisible ?

Dans les questions 16 à 18, on suppose que $\lambda = 1$.

16. On pose pour tout n de $\mathbb{N}^*$: $T_n = \max(X_1, X_2, \ldots, X_n)$.
Pour tout n de $\mathbb{N}^*$, pour tout réel x positif ou nul, on pose :

$$g_n(x) = \int_0^x F_{T_n}(t)\ dt \quad \text{et} \quad h_n(x) = \int_0^x t\, f_{T_n}(t)\ dt$$

a) Exprimer $h_n(x)$ en fonction de $F_{T_n}(x)$ et $g_n(x)$.

b) Déterminer pour tout réel t, l'expression de $F_{T_n}(t)$ en fonction de t.
Établir pour tout n supérieur ou égal à 2, la relation : $g_{n-1}(x) - g_n(x) = \dfrac{1}{n}\ F_{T_n}(x)$.

c) En déduire que pour tout n de $\mathbb{N}^*$, pour tout réel x positif ou nul, l'expression de $g_n(x)$ en fonction de x, $F_{T_1}(x)$, $F_{T_2}(x)$, $\ldots$, $F_{T_n}(x)$.

d) Montrer que $F_{T_n}(x) - 1$ est équivalent à $-n\mathrm{e}^{-x}$, lorsque x tend vers $+\infty$.

e) Déduire des questions ***c)*** et ***d)*** l'existence de $\mathbb{E}(T_n)$ et montrer que $\mathbb{E}(T_n) = \sum\limits_{k=1}^{n} \dfrac{1}{k}$.

17. On veut étudier dans cette question la convergence en loi de la suite de variables aléatoires $(G_n)_{n\geqslant 1}$ définie par : pour tout n de $\mathbb{N}^*$, $G_n = T_n - \mathbb{E}(T_n)$.
On pose pour tout n de $\mathbb{N}^*$: $\gamma_n = -\ln(n) + \mathbb{E}(T_n)$ et on admet sans démonstration que la suite $(\gamma_n)_{n\geqslant 1}$ est convergente ; on note γ sa limite.

a) Montrer que pour tout x réel et n assez grand, on a : $F_{G_n}(x) = \left(1 - \dfrac{1}{n}\ \mathrm{e}^{-(x+\gamma_n)}\right)^n$.

b) En déduire que pour tout x réel, on a : $\lim\limits_{n\to+\infty} F_{G_n}(x) = \mathrm{e}^{-\mathrm{e}^{-(x+\gamma)}}$.

c) Montrer que la fonction $F_G : \mathbb{R} \to \mathbb{R}$ définie par $F_G : x \mapsto \mathrm{e}^{-\mathrm{e}^{-(x+\gamma)}}$ est la fonction de répartition d'une variable aléatoire G à densité. Conclure.

18. ***a)*** Soit X une variable aléatoire à densité de fonction de répartition F_X strictement croissante. Déterminer la loi de la variable aléatoire Y définie par $Y = F_X(X)$.

b) Écrire une fonction **Python** d'en-tête `Gumbel` qui permet de simuler la variable aléatoire G.
On supposera que la constante γ est définie en langage **Python** par une constante `gamma`.
On rappelle que la fonction **Python** `rd.random()` permet de simuler la loi uniforme sur $]0, 1[$.

HEC 2019 - loi de Rademacher, loi binomiale, loi uniforme, loi normale, loi de Poisson, convergence en loi, fonctions génératrice des moments, des cumulants et des probabilités

Dans ce problème, on définit et on étudie les fonctions génératrices des cumulants de variables aléatoires discrètes ou à densité.
Les cumulants d'ordre 3 *et* 4 *permettent de définir des paramètres d'asymétrie et d'aplatissement qui viennent compléter la description usuelle d'une loi de probabilité par son espérance (paramètre de position) et sa variance (paramètre de dispersion) ; ces cumulants sont notamment utilisés pour l'évaluation des risques financiers.*

Dans tout le problème :

- on note $(\Omega, \mathscr{A}, \mathbb{P})$ un espace probabilisé et toutes les variables aléatoires introduites dans l'énoncé sont des variables aléatoires réelles définies sur $(\Omega, \mathscr{A})$;
- sous réserve d'existence, l'espérance et la variance d'une variable aléatoire X sont respectivement notées $\mathbb{E}(X)$ et $\mathbb{V}(X)$;
- pour tout variable aléatoire X et pour tout réel t pour lesquels la variable aléatoire $\mathrm{e}^{t\,X}$ admet une espérance, on pose :
$$M_X(t) = \mathbb{E}\left(\mathrm{e}^{t\,X}\right) \quad \text{et} \quad K_X(t) = \ln\left(M_X(t)\right);$$
(les fonctions M_X et K_X sont respectivement appelées la *fonction génératrice des moments* et la *fonction génératrice des cumulants* de X)
- lorsque, pour un entier $p \in \mathbb{N}^*$, la fonction K_X est de classe $\mathcal{C}^p$ sur un intervalle ouvert contenant l'origine, on appelle *cumulant d'ordre p de X*, noté $Q_p(X)$, la valeur de la dérivée $p^{\text{ème}}$ de K_X en 0 :
$$Q_p(X) \; = \; K_X^{(p)}(0)$$

Partie I. Fonction génératrice des moments de variables aléatoires discrètes

Dans toute cette partie :

- on note n un entier supérieur ou égal à 2 ;
- toutes les variables aléatoires considérées sont discrètes à valeurs entières ;
- on note S une variable aléatoire à valeurs dans $\{-1, 1\}$ dont la loi est donnée par :
$$\mathbb{P}([S=-1]) \; = \; \mathbb{P}([S=+1]) \; = \; \frac{1}{2}$$

1. Soit X une variable aléatoire à valeurs dans $[\![-n, n]\!]$.

a) Pour tout $t \in \mathbb{R}$, écrire $M_X(t)$ sous la forme d'une somme et en déduire que la fonction M_X est de classe $\mathcal{C}^\infty$ sur $\mathbb{R}$.

b) Justifier pour tout $p \in \mathbb{N}^*$, l'égalité : $M_X^{(p)}(0) = \mathbb{E}(X^p)$.

c) Soit Y une variable aléatoire à valeurs dans $[\![-n, n]\!]$ dont la fonction génératrice des moments M_Y est la même que celle de X.
On note G_X et G_Y les deux polynômes définis par :
$$\forall x \in \mathbb{R}, \quad \left\{ \begin{array}{rcl} G_X(x) & = & \sum\limits_{k=0}^{2n} \mathbb{P}([X=k-n])\, x^k \\ G_Y(x) & = & \sum\limits_{k=0}^{2n} \mathbb{P}([Y=k-n])\, x^k \end{array} \right.$$

(i) Vérifier pour tout $t \in \mathbb{R}$, l'égalité : $G_X(\mathrm{e}^t) = \mathrm{e}^{nt}\, M_X(t)$.

(ii) Justifier la relation : $\forall t \in \mathbb{R},\ G_X(e^t) = G_Y(e^t)$.

(iii) En déduire que la variable aléatoire Y suit la même loi que X.

2. Dans cette question, on note X_2 une variable aléatoire qui suit la loi binomiale $\mathcal{B}\left(2, \dfrac{1}{2}\right)$.

On suppose que les variables aléatoires X_2 et S sont indépendantes et on pose $Y_2 = S\,X_2$.

a) (i) Préciser l'ensemble des valeurs possibles de la variable aléatoire Y_2.

(ii) Calculer les probabilités $\mathbb{P}([Y_2 = y])$ attachées aux diverses valeurs possibles y de Y_2.

b) Vérifier que la variable aléatoire $X_2 - (S+1)$ suit la même loi que Y_2.

3. Le script **Python** suivant permet d'effectuer des simulations de la variable aléatoire Y_2 définie dans la question précédente.

```
1  n = 10
2  X = rd.binomial(2,0.5,[n,2])
3  B = rd.binomial(1,0.5,[n,2])
4  S = 2*B - np.ones([n,2])
5  Z1 = [S[:,0]*X[:,0], X[:,0] - S[:,0] - np.ones(n)]
6  Z2 = [S[:,0]*X[:,0], X[:,1] - S[:,1] - np.ones(n)]
```

a) Que contiennent les variables `X` et `S` après l'exécution des quatre premières instructions ?

b) Expliquer pourquoi, après l'exécution des six instructions, chacun des coefficients des matrices `Z1` et `Z2` contient une simulation de la variable aléatoire Y_2.

c) On modifie la première ligne du script précédent en affectant à `n` une valeur beaucoup plus grande que 10 (par exemple, 100000) et en lui adjoignant les deux instructions 7 et 8 suivantes :

```
7  p1 = len(np.argwhere(Z1[0] == Z1[1])) / n
8  p2 = len(np.argwhere(Z2[0] == Z2[1])) / n
```

Quelles valeurs numériques approchées la loi faible des grands nombres permet-elle de fournir pour `p1` et `p2` après l'exécution des huit lignes du nouveau script ?

Dans le langage **Python**, la fonction `len` fournit la « longueur » d'un vecteur, d'une liste ou d'une matrice carrée et la fonction `np.argwhere` calcule les positions des coefficients d'une matrice pour lesquels une propriété est vraie, comme l'illustre le script suivant :

```
--> A = np.array([1,2,0,4])
--> B = np.array([2,2,4,3])
--> len(A)
    ans = 4.
--> np.argwhere(A < B)
    = [[0]
       [2]]
    # car 1 < 2 et 0 < 4, alors que 2 ⩾ 2 et 4 ⩾ 3
```

4. Dans cette question, on note X_n une variable aléatoire qui suit la loi binomiale $\mathcal{B}\left(n, \dfrac{1}{2}\right)$.

On suppose que les variables aléatoires X_n et S sont indépendantes et on pose $Y_n = S\,X_n$.

a) Justifier que la fonction M_{X_n} est définie sur $\mathbb{R}$ et calculer $M_{X_n}(t)$ pour tout $t \in \mathbb{R}$.

b) Montrer que la fonction M_{Y_n} est donnée par : $\forall t \in \mathbb{R},\ M_{Y_n}(t) = \dfrac{1}{2^{n+1}}\left((1+e^t)^n + (1+e^{-t})^n\right)$.

c) En utilisant l'égalité $(1+\mathrm{e}^{-t})^n = \mathrm{e}^{-nt}\,(1+\mathrm{e}^t)^n$, montrer que Y_n suit la même loi que la différence $X_n - H_n$, où H_n est une variable aléatoire indépendante de X_n dont on précisera la loi.

Partie II. Propriétés générales des fonctions génératrices des cumulants et quelques exemples

5. Soit X une variable aléatoire et $\mathcal{D}_X$ le domaine de définition de la fonction K_X.

a) Donner la valeur de $K_X(0)$.

b) Soit $(a,b) \in \mathbb{R}^2$ et $Y = a\,X + b$. Justifier pour tout réel t pour lequel $a\,t$ appartient à $\mathcal{D}_X$, l'égalité :

$$K_Y(t) \;=\; b\,t + K_X(a\,t)$$

c) On suppose ici que les variables aléatoires X et $-X$ suivent la même loi.
Que peut-on dire dans ce cas des cumulants d'ordre impair de la variables aléatoire X ?

6. Soit X et Y deux variables aléatoires indépendantes et $\mathcal{D}_X$ et $\mathcal{D}_Y$ les domaines de définition respectifs des fonctions K_X et K_Y.

a) Monter que pour tout réel t appartenant à la fois à $\mathcal{D}_X$ et $\mathcal{D}_Y$, on a : $K_{X+Y}(t) = K_X(t) + K_Y(t)$.

b) En déduire une relation entre les cumulants des variables aléatoires X, Y et $X+Y$.

7. Soit U une variable aléatoire suivant la loi uniforme sur l'intervalle $[0,1]$.

a) Montrer que la fonction M_U est définie sur $\mathbb{R}$ et donnée par : $\forall t \in \mathbb{R},\ M_U(t) = \begin{cases} \dfrac{\mathrm{e}^t - 1}{t} & \text{si } t \neq 0 \\ 1 & \text{si } t = 0 \end{cases}$.

b) Calculer la dérivée de la fonction M_U en tout point $t \neq 0$.

c) Trouver la limite du quotient $\dfrac{M_U(t) - 1}{t}$ lorsque t tend vers 0.

d) Montrer que la fonction M_U est de classe $\mathcal{C}^1$ sur $\mathbb{R}$.

8. Soit α et β deux réels tels que $\alpha < \beta$.
Dans cette question, on note X une variable aléatoire qui suit la loi uniforme sur l'intervalle $[\alpha, \beta]$.

a) Exprimer K_X en fonction de M_U, où la variable aléatoire U a été définie dans la question ***7.***

b) Justifier que la fonction K_X est de classe $\mathcal{C}^1$ sur $\mathbb{R}$ et établir l'égalité : $Q_1(X) = \mathbb{E}(X)$.

9. Soit un réel $\lambda > 0$ et soit T une variable aléatoire qui suit la loi de Poisson de paramètre λ.

a) Déterminer les fonctions M_T et K_T.

b) En déduire les cumulants de T.

10. Soit Z une variable aléatoire qui suit la loi normale centrée réduite.

a) Justifier pour tout $t \in \mathbb{R}$, la convergence de l'intégrale $\displaystyle\int_{-\infty}^{+\infty} \exp\left(t\,x - \frac{x^2}{2}\right)\,dx$.

b) Montrer que la fonction M_Z est définie sur $\mathbb{R}$ et donnée par : $\forall t \in \mathbb{R},\ M_Z(t) = \exp\left(\dfrac{t^2}{2}\right)$.

c) En déduire la valeur de tous les cumulants d'une variable aléatoire qui suit une loi normale d'espérance $\mu \in \mathbb{R}$ et d'écart-type $\sigma \in \mathbb{R}_+^*$.

11. Soit $(T_n)_{n\in\mathbb{N}^*}$ une suite de variables aléatoires telles que, pour tout $n \in \mathbb{N}^*$, la variable aléatoire T_n suit la loi de Poisson de paramètre n. Pour tout $n \in \mathbb{N}^*$, on pose : $W_n = \dfrac{T_n - n}{\sqrt{n}}$.

a) Justifier la convergence en loi de la suite de variables aléatoires $(W_n)_{n\in\mathbb{N}^*}$ vers une variable aléatoire W.

b) Déterminer la fonction K_{W_n}.

c) Montrer que pour tout $t \in \mathbb{R}$, on a : $\displaystyle\lim_{n\to+\infty} K_{W_n}(t) = K_W(t)$.

Partie III. Cumulant d'ordre 4

Dans cette partie, on considère une variable aléatoire X telle que M_X est de classe $\mathcal{C}^4$ sur un intervalle ouvert I contenant l'origine.
On admet alors que X possède des moments jusqu'à l'ordre 4 qui coïncident avec les dérivées successives de la fonction M_X en 0. Autrement dit, pour tout $k \in [\![1,4]\!]$, on a : $M_X^{(k)}(0) = \mathbb{E}(X^k)$.

De plus, on pose : $\mu_4(X) = \mathbb{E}\left(\big(X - \mathbb{E}(X)\big)^4\right)$.

12. Justifier les égalités : $Q_1(X) = \mathbb{E}(X)$ et $Q_2(X) = \mathbb{V}(X)$.

13. Soit X_1 et X_2 deux variables aléatoires indépendantes et de même loi que X. On pose : $S = X_1 - X_2$.

a) Montrer que la variable aléatoire S possède un moment d'ordre 4 et établir l'égalité :

$$\mathbb{E}(S^4) \;=\; 2\,\mu_4(X) + 6\,\big(\mathbb{V}(X)\big)^2$$

b) Montrer que les fonctions M_S et K_S sont de classe $\mathcal{C}^4$ sur I et que pour tout $t \in I$, on a :

$$M_S^{(4)}(t) \;=\; K_S^{(4)}(t)\,M_S(t) + 3\,K_S^{(3)}(t)\,M_S'(t) + 3\,K_S''(t)\,M_S''(t) + K_S'(t)\,M_S^{(3)}(t)$$

c) En déduire l'égalité : $\mathbb{E}(S^4) \;=\; Q_4(S) + 3\,\big(\mathbb{V}(S)\big)^2$.

14. Justifier que le cumulant d'ordre 4 de X est donné par la relation : $Q_4(X) \;=\; \mu_4(X) - 3\,\big(\mathbb{V}(X)\big)^2$.

HEC 2020 - couple de variables aléatoires à densité, loi exponentielle, loi de Bernoulli, loi binomiale, inégalité de Boole

On s'intéresse dans ce sujet au problème de la *double dépense* de *bitcoins* par un groupe d'individus mal intentionnés.
On rappelle que le bitcoin est une monnaie virtuelle dont l'utilisation pour des transactions est associée à une structure unique appelée *blockchain*, partagée sur le réseau des usagers de cette monnaie et ayant pour but de sécuriser ces transactions.
La modélisation étudiée ne nécessite pas de connaissances particulières sur le *bitcoin* et la *blockchain*.

Partie I - Deux résultats généraux

On démontre dans cette partie deux résultats préliminaires, aux questions ***5.*** et ***6.***. Ces résultats seront utilisés dans la suite du sujet et pourront être admis.

Calcul d'une probabilité

Soient X et Y deux variables aléatoires sur un espace probabilisé, à densité et indépendantes.
On note F_X et F_Y les fonctions de répartition de X et Y.
On suppose que Y est à valeurs positives et possède une densité f_Y dont la restriction à $[0,+\infty[$ est continue sur cet intervalle.
Pour tout $x \in \mathbb{R}_+$, on pose : $H(x) = \mathbb{P}(\,[X \leqslant Y] \cap [Y \leqslant x]\,)$.

1. a) Montrer que H est une fonction croissante sur $\mathbb{R}_+$ qui admet une limite finie en $+\infty$.

b) En utilisant la suite $\big(H(n)\big)_{n\in\mathbb{N}}$, démontrer : $\lim\limits_{x\to+\infty} H(x) = \mathbb{P}(\,[X \leqslant Y]\,)$.
Que vaut $H(0)$?

2. Soit (u,v) un couple de réels positifs tels que : $u < v$.

a) Montrer : $H(v) - H(u) = \mathbb{P}(\,[X \leqslant Y] \cap [u < Y \leqslant v]\,)$. Puis :

$$F_X(u)\ \frac{F_Y(v) - F_Y(u)}{v-u} \ \leqslant\ \frac{H(v)-H(u)}{v-u} \ \leqslant\ F_X(v)\ \frac{F_Y(v)-F_Y(u)}{v-u}$$

b) En déduire que pour tout $x \in \mathbb{R}_+$, H est dérivable en x et : $H'(x) = F_X(x)\, f_Y(x)$.

c) En conclure que pour tout x réel positif : $H(x) = \displaystyle\int_0^x F_X(t)\, f_Y(t)\ dt$.

3. Démontrer : $\mathbb{P}(\,[X \leqslant Y]\,) = \displaystyle\int_0^{+\infty} F_X(t)\, f_Y(t)\ dt$.

4. En utilisant la fonction $K : x \mapsto \mathbb{P}(\,[X < Y] \cap [Y \leqslant x]\,)$, on montrerait de même et nous l'admettrons :

$$\mathbb{P}(\,[X < Y]\,) \ = \ \int_0^{+\infty} F_X(t)\, f_Y(t)\ dt \ = \ \mathbb{P}(\,[X \leqslant Y]\,)$$

Que peut-on en déduire pour $\mathbb{P}(\,[X = Y]\,)$?

5. *Application aux lois exponentielles*
On suppose que U et V sont deux variables aléatoires indépendantes suivant des lois exponentielles de paramètres respectifs λ et μ, réels strictement positifs.
Soit θ un réel positif ou nul.

a) Déterminer la fonction de répartition de la variable aléatoire $X = U - \theta$.

b) En déduire que pour tout $\theta \geqslant 0$:

$$\mathbb{P}(\,[U - \theta \leqslant V]\,) \;=\; 1 - \frac{\mu}{\lambda + \mu}\;\mathrm{e}^{-\lambda\,\theta}$$

Inégalité de Boole

6. On considère $(B_k)_{k \in \mathbb{N}^*}$ une famille d'événements d'un espace probabilisé.

a) Montrer par récurrence sur $n \in \mathbb{N}^*$: $\mathbb{P}\left(\bigcup\limits_{k=1}^{n} B_k\right) \leqslant \sum\limits_{k=1}^{n} \mathbb{P}(B_k)$.

b) On suppose que la série $\sum\limits_{k \geqslant 1} \mathbb{P}(B_k)$ converge. Démontrer :

$$\mathbb{P}\left(\bigcup_{k=1}^{+\infty} B_k\right) \;\leqslant\; \sum_{k=1}^{+\infty} \mathbb{P}(B_k)$$

Partie II - Une compétition entre deux groupes

Dans toute la suite du sujet, on désigne par p un réel de l'intervalle $]0,1[$ et on pose $q = 1 - p$.
On modélise une compétition entre deux groupes d'individus A et B avec les règles suivantes.

- Le groupe A doit résoudre une suite de problèmes $(P_k)_{k \geqslant 1}$ dans l'ordre des indices. Au temps $t = 0$, le groupe commence la résolution du problème P_1, ce qui lui prend un temps représenté par la variable aléatoire X_1. Une fois P_1 résolu, le groupe aborde immédiatement le problème P_2, et on note X_2 le temps consacré à la résolution de P_2 par le groupe A, et ainsi de suite.
  Pour tout $k \in \mathbb{N}^*$, on note X_k la variable aléatoire donnant le temps consacré à la résolution du problème P_k par le groupe A.
- De même, le groupe B doit résoudre dans l'ordre une suite de problèmes $(Q_k)_{k \geqslant 1}$; la résolution du premier problème Q_1 commence au temps $t = 0$ et on note, pour tout $k \in \mathbb{N}^*$, Y_k la variable aléatoire donnant le temps consacré par le groupe B à la résolution du problème Q_k.
- À ce jeu est associé un espace probabilisé $(\Omega, \mathscr{A}, \mathbb{P})$ sur lequel sont définies les suites de variables aléatoires $(X_k)_{k \geqslant 1}$ et $(Y_k)_{k \geqslant 1}$, et on fait les hypothèses suivantes :
  - × pour tout $k \in \mathbb{N}^*$, X_k suit la loi exponentielle de paramètre p, notée $\mathcal{E}\,(p)$, et Y_k suit la loi exponentielle $\mathcal{E}\,(q)$;
  - × pour tout $k \in \mathbb{N}^*$, les variables aléatoires $X_1, \ldots, X_k, Y_1, \ldots, Y_k$ sont indépendantes.
- On établit alors la liste de tous les problèmes résolus *dans l'ordre où ils le sont par les deux groupes*. En cas de simultanéité temporelle de la résolution par les deux groupes d'un de leurs problèmes, on placera d'abord le problème résolu par A dans la liste puis celui résolu par B.
  Pour tout $n \in \mathbb{N}^*$, on note U_n la variable aléatoire de Bernoulli associée à l'événement « le $n^{\text{ème}}$ problème placé dans la liste est un problème résolu par le groupe A ».
  Par exemple, si la liste des cinq premiers problèmes résolus est $(P_1, P_2, Q_1, P_3, Q_2)$, alors $U_1 = 1$, $U_2 = 1$, $U_3 = 0$, $U_4 = 1$ et $U_5 = 0$.
- Pour tout $n \geqslant 0$, on note aussi S_n la variable aléatoire donnant le nombre de problèmes qui ont été résolus par A présents dans la liste des n premiers problèmes résolus. En particulier, S_0 vaut toujours 0.

7. ***a)*** Que représente la variable aléatoire $\sum\limits_{k=1}^{n} X_k$?

b) On suppose que $X_1 = 5$, $X_2 = 2$, $X_3 = 3$, $X_4 = 2$, $Y_1 = 2$, $Y_2 = 2$, $Y_3 = 4$ et $Y_4 = 2$.
Déterminer $U_1, \ldots, U_7$.
Peut-on aussi en déduire la valeur de U_8 ?

c) Compléter le script **Python** suivant pour qu'il simule le jeu et, pour n, p donnés, affiche la liste des valeurs $U_1, U_2, \ldots, U_n$:

```
p = float(input('p = '))
n = int(input('n = '))
q = 1 - p
U = np.zeros(n)
sommeX = rd.exponential(1/p)
sommeY = rd.exponential(1/q)
mini = min(sommeX, sommeY)
for k in range(n):
    if sommeX == ...:
        U[k] = ...
        sommeX = sommeX + rd.exponential(1/p)
    else:
        sommeY = ...
    mini = min(sommeX, sommeY)
...
```

d) Quelle(s) instruction(s) faut-il ajouter pour afficher la valeur de S_n ?

8. *Loi de U_n*
Dans cette question, on démontre par récurrence sur $n \geqslant 1$: $\mathbb{P}(\,[U_n = 1]\,) = p$.

a) Démontrer : $\mathbb{P}(\,[U_1 = 1]\,) = \mathbb{P}(\,[X_1 \leqslant Y_1]\,) = p$.

b) (i) Démontrer, pour tout réel $x < 0$: $\mathbb{P}_{[U_1=1]}(\,[Y_1 - X_1 \leqslant x]\,) = 0$.

(ii) Soit x un réel positif ou nul.
Établir : $\mathbb{P}_{[U_1=1]}(\,[Y_1 - X_1 \leqslant x]\,) = \dfrac{1}{p}\, \mathbb{P}(\,[X_1 \leqslant Y_1 \leqslant X_1 + x]\,)$,
puis calculer $\mathbb{P}_{[U_1=1]}(\,[Y_1 - X_1 \leqslant x]\,)$.

c) On peut interpréter ce résultat en disant que la *loi conditionnelle de* $Y_1 - X_1$ *sachant* $[U_1 = 1]$ est une loi exponentielle. Quelle est son paramètre ?
Par analogie, quelle est la loi conditionnelle de $X_1 - Y_1$ sachant $[U_1 = 0]$? (on n'attend pas une démonstration précise mais un argument de bon sens pour justifier le résultat proposé).

d) On suppose que $n \in \mathbb{N}^*$ et $\mathbb{P}(\,[U_1 = 1]\,) = p$.
Déduire de cette hypothèse et de la question précédente :
$\mathbb{P}_{[U_1=1]}(\,[U_{n+1} = 1]\,) = p$ et $\mathbb{P}_{[U_1=0]}(\,[U_{n+1} = 1]\,) = \ p$.

e) Conclure.

9. On montrerait aussi par récurrence, et nous l'admettrons, que pour tout $n \in \mathbb{N}^*$, les variables aléatoires $U_1, \ldots, U_n$ sont mutuellement indépendantes.
En déduire la loi de S_n.

Soit $r \in \mathbb{N}$, on s'intéresse, dans les questions qui suivent, à la probabilité a_r de l'événement :

A_r : « il existe un $n \geqslant r$ tel que, lorsque n problèmes en tout ont été résolus, le groupe A en a résolu r de plus que le groupe B »

10. ***a)*** Justifier : $a_0 = 1$.

b) Démontrer, pour tout $r \geqslant 1$:

$$\mathbb{P}_{[U_1=1]}(A_r) = \mathbb{P}(A_{r-1}) \quad \text{et} \quad \mathbb{P}_{[U_1=0]}(A_r) = \mathbb{P}(A_{r+1})$$

c) En déduire, pour tout $r \geqslant 1$: $a_{r+1} = \dfrac{1}{q}\ a_r - \dfrac{p}{q}\ a_{r-1}$.

d) En remarquant que $1 - 4pq = (1-2p)^2$, donner une expression de a_r en fonction de p, q, r et de deux constantes que l'on introduira.

11. Le cas $p \geqslant \dfrac{1}{2}$.

Montrer que, dans les cas $p = \dfrac{1}{2}$ et $p > \dfrac{1}{2}$, la suite $(a_r)_{r\in\mathbb{N}}$ est constante et égale à 1.

12. Le cas $p < \dfrac{1}{2}$.

a) Soit k un entier naturel.

(i) Établir : $A_{2k} = \bigcup\limits_{i \geqslant k} [S_{2i} = i + k]$.

(ii) Montrer que pour tout $i \geqslant k$, on a : $\mathbb{P}(\,[S_{2i} = i+k]\,) = \dbinom{2i}{i+k} p^{i+k}\, q^{i-k}$.

(iii) Après avoir donné la valeur de la somme $\sum\limits_{j=0}^{2i} \dbinom{2i}{j}$, démontrer :

$$\forall i \geqslant k,\ \binom{2i}{i+k} \leqslant 4^i$$

(iv) En déduire l'inégalité :

$$\sum_{i=k}^{+\infty} \mathbb{P}(\,[S_{2i} = k+i]\,) \ \leqslant\ \left(\frac{p}{q}\right)^k \ \frac{(4pq)^k}{1-4pq}$$

b) Montrer en utilisant l'inégalité de Boole (voir question ***6.***) que si $p < \dfrac{1}{2}$, alors : $\lim\limits_{k\to+\infty} a_{2k} = 0$.

c) Conclure en utilisant la question ***10.d)***, que si $p < \dfrac{1}{2}$, alors :

$$\forall r \in \mathbb{N},\ a_r = \left(\frac{p}{q}\right)^r$$

On a ainsi établi dans les questions ***11.*** et ***12.*** :

$$\forall r \in \mathbb{N}, \quad a_r = \begin{cases} \left(\dfrac{p}{q}\right)^r & \text{si } p < \dfrac{1}{2} \\[2ex] 1 & \text{si } p \geqslant \dfrac{1}{2} \end{cases}$$

Ce résultat pourra être admis et utilisé dans la suite du sujet.

Partie III - La *blockchain* et la stratégie de la *double dépense*

On utilise, dans cette partie, les notations et résultats de la partie **II**.
Soit n un entier supérieur ou égal à 1.
La *blockchain* est formée d'une suite de blocs, chacun associé à plusieurs transactions. Elle contient l'historique de toutes les transactions effectuées depuis la création du *bitcoin*.
Avant d'être placé dans la *blockchain*, un nouveau bloc doit être validé. Cette validation nécessite la mise en oeuvre d'une grande puissance de calcul pour résoudre un problème dépendant fortement du contenu du bloc et des blocs qui le précèdent.
Les individus qui valident les blocs sont appelés mineurs.
Il est possible qu'à un instant donné, coexistent sur le réseau deux *blockchains*, valides et différentes. Dans ce cas, le réseau choisira celle qui comporte le plus de blocs et l'autre sera abandonnée.
Par prudence, lorsqu'un bloc est validé, il est recommandé d'attendre que $n-1$ blocs le suivant soient aussi validés pour considérer que les transactions incluses dans le bloc soient honnêtes.
Un groupe de mineurs mal intentionnés, noté A, peut essayer de dépenser deux fois les mêmes *bitcoins* en procédant ainsi :

- le groupe A demande la validation de l'achat d'un bien d'un montant de s *bitcoins* qu'il a en sa possession.
- lorsque le bloc K incluant cette transaction est proposé à la validation sur le réseau, A modifie ce bloc en K', qu'il ne diffuse pas, en remplaçant l'achat par une vente des s *bitcoins* en euros à son profit par exemple. Il se met alors à la validation de ce nouveau bloc et crée ainsi une deuxième instance de la *blockchain* qu'il continue à développer sans la diffuser.
- lorsque le groupe B, représentant l'ensemble des autres mineurs du réseau, a validé K ainsi que les $n-1$ blocs suivants, le vendeur du bien considère que la transaction est valide et fournit le bien.
- le groupe A attend alors d'avoir une *blockchain* plus longue que celle de B, qui est publique, pour la diffuser donc invalider la *blockchain* publique et l'achat du bien. Le crédit en *bitcoins* du vendeur du bien est alors annulé.

On reprend et on complète la modélisation de la partie précédente pour déterminer la probabilité que la stratégie de la *double dépense* réussisse et le choix de n pour que cette probabilité soit faible.
Une première phase du jeu, décrit dans la partie **II**, s'achève à l'instant aléatoire t où le problème Q_n est ajouté à la liste des problèmes résolus.
Le groupe de mineurs A est ensuite déclaré vainqueur s'il se trouve un instant $t' \geqslant t$ où le nombre de problèmes résolus par A dans la liste des problèmes résolus depuis le début du jeu, est strictement supérieur au nombre de ceux résolus par B dans cette même liste. On note G_n cet événement.
On détermine, dans cette partie, la probabilité de G_n en fonction de n et de p.

13. On s'intéresse tout d'abord à la loi de la variable aléatoire T_n égale au nombre de problèmes résolus par le groupe A lorsque l'on place Q_n dans la liste des problèmes résolus.

a) Démontrer, pour tout $k \in \mathbb{N}$: $[T_n = k] = [S_{n+k-1} = k] \cap [U_{n+k} = 0]$.

b) En déduire : $\mathbb{P}(\,[T_n = k]\,) = \dbinom{n+k-1}{k} p^k\, q^n$.

14. ***a)*** En utilisant la formule des probabilités totales, établir :

$$\mathbb{P}(G_n) \;=\; \mathbb{P}(\,[T_n \geqslant n+1]\,) + \sum_{k=0}^{n} \mathbb{P}(\,[T_n = k]\,)\, a_{n+1-k}$$

b) Dans le cas où $p \geqslant \dfrac{1}{2}$, en déduire : $\mathbb{P}(G_n) = 1$.

c) De même lorsque $p < \dfrac{1}{2}$, démontrer :

$$\mathbb{P}(G_n) \;=\; 1 - \sum_{k=0}^{n} \binom{n+k-1}{k} \left(p^k\, q^n - p^{n+1}\, q^{k-1}\right)$$

15. *Une meilleure expression de $\mathbb{P}(G_n)$ lorsque $p < \dfrac{1}{2}$*

Pour tout $x \in [0,1]$ et $n \in \mathbb{N}^*$, on pose :

$$u_n(x) \;=\; (1-x)^n \sum_{k=0}^{n} \binom{n+k-1}{k} x^k$$

a) Vérifier que pour tout $n \in \mathbb{N}^*$: $\mathbb{P}(G_n) = 1 - u_n(p) + \dfrac{p}{q}\, u_n(q)$.

b) Pour tout $x \in [0,1]$ et $n \in \mathbb{N}^*$, établir la relation :

$$u_{n+1}(x) \;=\; u_n(x) + (1-x)^n\, x^{n+1} \left(\binom{2n}{n+1} - \binom{2n+1}{n+1} x \right)$$

c) En déduire, pour tout $n \in \mathbb{N}^*$:

$$\mathbb{P}(G_{n+1}) \;=\; \mathbb{P}(G_n) - \left(1 - \frac{p}{q}\right) (pq)^{n+1} \binom{2n+1}{n+1}$$

d) Montrer par récurrence, pour tout $n \in \mathbb{N}^*$:

$$\mathbb{P}(G_n) \;=\; \frac{p}{q} - \left(1 - \frac{p}{q}\right) \sum_{k=1}^{n} \binom{2k-1}{k} (pq)^k$$

16. *Application à la sécurisation des transactions*

Connaissant $p < \dfrac{1}{2}$, on cherche à limiter le risque que la stratégie mise en place par le groupe de mineurs A réussisse.

a) Après avoir établi la formule $\dbinom{n}{k} = \dfrac{n}{k} \dbinom{n-1}{k-1}$ lorsque $k \in [\![1, n]\!]$, écrire une fonction **Python** qui calcule les coefficients binomiaux.

b) Écrire un script **Python** qui détermine n_p, le plus petit entier n tel que $\mathbb{P}(G_n) \leqslant \varepsilon$ pour $p < \dfrac{1}{2}$ et $\varepsilon > 0$ saisis au clavier par l'utilisateur.

NB : Pour $\varepsilon = 10^{-4} = 0,1\%$ et p variant entre 10% et 32%, on obtient pour la représentation de n_p en fonction de p :

HEC 2008 (Exercice) - fonction de deux variables, méthode des moindres carrés, droite de régression linéaire

Étant donné un entier n supérieur ou égal à 2, on considère un nuage de n points du plan, c'est-à-dire un n-uplet $\big((x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\big)$ d'éléments de $\mathbb{R}^2$. On suppose que les réels x_1, x_2, …, x_n (resp. y_1, y_2, …, y_n) ne sont pas tous égaux.
On appelle moyenne arithmétique $\overline{x}$ et écart-type σ_x du n-uplet $x = (x_1, \dots, x_n)$, les réels suivants :

$$\overline{x} \ = \ \frac{1}{n} \sum_{k=1}^{n} x_k \qquad \text{et} \qquad \sigma_x \ = \ \sqrt{\frac{1}{n} \sum_{k=1}^{n} (x_k - \overline{x})^2}$$

On définit de même la moyenne arithmétique $\overline{y}$ et l'écart-type σ_y du n-uplet $y = (y_1, \dots, y_n)$.
La covariance $\text{cov}(x, y)$ et le coefficient de corrélation linéaire $r(x, y)$ du couple (x, y) sont donnés par :

$$\text{cov}(x, y) \ = \ \frac{1}{n} \sum_{k=1}^{n} (x_k - \overline{x})\,(y_k - \overline{y}) \qquad \text{et} \qquad r(x, y) = \frac{\text{cov}(x, y)}{\sigma_x \times \sigma_y}$$

Soit f la fonction définie sur $\mathbb{R}^2$ à valeurs réelles qui, à tout couple (a, b) de $\mathbb{R}^2$, associe le réel $f(a, b)$ tel que :

$$f(a, b) \ = \ \sum_{k=1}^{n} (a\,x_k + b - y_k)^2$$

1. Justifier que f est de classe $\mathcal{C}^2$ sur $\mathbb{R}^2$.

2. ***a)*** Écrire le système d'équations (S) permettant de déterminer les points critiques de f.

b) Résoudre le système (S). En déduire que f admet un unique point critique $(\hat{a}, \hat{b})$ que l'on exprimera en fonction de $\overline{x}$, $\overline{y}$, σ_x^2 et $\text{cov}(x, y)$.

c) Montrer que ce point critique correspond à un minimum local de f.

d) Établir la formule suivante : $f(\hat{a}, \hat{b}) = n\,\sigma_y^2\,\big(1 - r^2(x, y)\big)$.

3. ***a)*** Montrer que l'on a : $|r(x, y)| \leqslant 1$.

b) Que peut-on dire du nuage de points lorsque $|r(x, y)| = 1$?