HEC-ESSEC 2024 – Mathématiques appliquées

La présentation, la lisibilité, l'orthographe, la qualité de la rédaction, la clarté et la précision des raisonnements entreront pour une part importante dans l'appréciation des copies.
Les candidats sont invités à ***encadrer*** *dans la mesure du possible les résultats de leurs calculs.*
Ils ne doivent faire usage d'aucun document. L'utilisation de toute calculatrice et de tout matériel électronique est interdite. Seule l'utilisation d'une règle graduée est autorisée.
Si, au cours de l'épreuve, un candidat repère ce qui lui semble être une erreur d'énoncé, il le signale sur sa copie et poursuit sa composition en expliquant les raisons des initiatives qu'il est amené à prendre.

Dans tout le sujet on considère un espace probabilisé $(\Omega, \mathcal{A}, \mathbf{P})$, 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 génèrer 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 de Bernoulli de paramètre p.
  Les variables $T_{u,v}$, 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 $\emptyset$.

Partie 1 – Nombre aléatoire de triangles

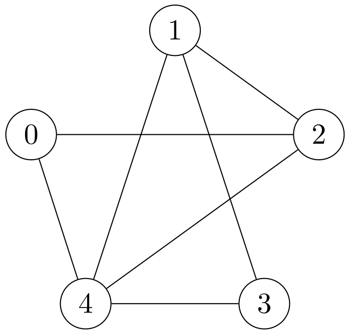
On note $\mathcal{T}$ l'ensemble des parties $\{u, v, w\}$ à trois éléments de l'ensemble des sommets, r le nombre de ses éléments et on pose

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

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 triangle de G.

Par exemple si $n = 5$ et le graphe de 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 $Z_n = \sum_{k=1}^{r} Y_k$. En déduire que $E(Z_n) = \binom{n}{3}p^3$.

▶ 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 :

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

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

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

   (c) En conclure que :

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

▶ On note $\Delta_n = \sum_{(i,j)\in\mathcal{E}} E(Y_iY_j)$.

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

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 – Nombre aléatoire de triangles

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 triangles2s(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 $(x_1, \ldots, x_k)$ élémént 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 $X_1(\Omega) \times \ldots \times X_k(\Omega)$ et $Y_k = f(X_1, \ldots, X_k)$ alors

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

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 $X_1(\Omega) \times \ldots \times X_k(\Omega)$ 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 :

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

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 $y \in X(\Omega)$,
   $$E(f(X)g(X)) + f(y)g(y) \geqslant g(y)E(f(X)) + f(y)E(g(X))$$
   (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 $X_1(\Omega) \times \ldots \times X_{k+1}(\Omega)$ 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) À l'aide des théorèmes de transfert d'ordre $k+1$ et k, montrer que :
    $$E(Y_{k+1}Z_{k+1}) = \sum_{x\in X_{k+1}(\Omega)} E(f(X_1, \ldots, X_k, x)g(X_1, \ldots, X_k, x))\mathbf{P}(X_{k+1} = x)$$
    (b) Justifier que pour tout $x \in X_{k+1}(\Omega)$:
    $$E(f(X_1, \ldots, X_k, x)g(X_1, \ldots, X_k, x)) \geqslant E(f(X_1, \ldots, X_k, x))E(g(X_1, \ldots, X_k, x))$$
    (c) On pose pour tout $x \in X_{k+1}(\Omega)$, $u(x) = E(f(X_1, \ldots, X_k, x))$ et $v(x) = E(g(X_1, \ldots, X_k, x))$.
    Montrer que u et v sont croissantes sur $X_{k+1}(\Omega)$ et $E(Y_{k+1}Z_{k+1}) \geqslant E(u(X_{k+1})v(X_{k+1}))$.

(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 $\mathbf{P}(Z_n = 0)$.

11. Justifier que $\bigcap_{0 \leqslant 0 < v \leqslant n-1} [T_{u,v} = 0] \subset [Z_n = 0]$. En déduire que

$$\mathbf{P}(Z_n = 0) \geqslant (1-p)^{\binom{n}{2}} > 0$$

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

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 $u <$ 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 $i \in [\![2, r]\!]$, $\mathbf{P}(Z_{n,i} = 0) \geqslant \mathbf{P}(Z_{n,i-1} = 0)\mathbf{P}(Y_i = 0)$ puis que :

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

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

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

► Si A est un événement de probabilité non nulle, on rappelle que la probabilité conditionnelle sachant A est notée $\mathbf{P}_A$. On admet qu'elle possède les mêmes propriétés que la probabilité $\mathbf{P}$.
En particulier l'inégalité de Boole est vérifiée par $\mathbf{P}_A$.
De plus si X est une variable finie, on note $E_A(X)$ l'espérance de X pour la probabilité $\mathbf{P}_A$ ce qui signifie que :

$$E_A(X) = \sum_{x \in X(\Omega)} x \mathbf{P}_A(X = x)$$

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 $\mathbf{P}(B \cap C) \neq 0$ et $P(A \cap C) \neq 0$. Montrer que $\mathbf{P}_{B\cap C}(A) \geqslant \mathbf{P}_C(A)\mathbf{P}_{A\cap C}(B)$.

▶ 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 $I_i = \{j \in [\![1, i-1]\!]/j \equiv i\}$ et $J_i = \{j \in [\![1, i-1]\!]/j \not\equiv i\}$.
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

$$P_{B_i\cap C_i}(\overline{A_i}) \geqslant \mathbf{P}(\overline{A_i})\mathbf{P}_{\overline{A_i}\cap C_i}(B_i)$$

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

(c) **On admet provisoirement que pour** $j \in [\![1, i-1]\!]$ **:**

$$\mathbf{P}_{\overline{A_i}\cap C_i}(\overline{A_j}) \leqslant \mathbf{P}_{\overline{A_i}}(\overline{A_j}) \tag{1}$$

En déduire que $\mathbf{P}_{B_i\cap C_i}(A_i) \leqslant 1 - \mathbf{P}(\overline{A_i}) \left(1 - \sum_{j\in I_i} \mathbf{P}_{\overline{A_i}}(\overline{A_j})\right)$.

(d) Justifier que pour tout $x \in \mathbb{R}$, $1 - x \leqslant \exp(-x)$ et en déduire que :

$$\mathbf{P}_{B_i\cap C_i}(A_i) \leqslant \exp\left(-\mathbf{P}(\overline{A_i}) + \sum_{j\in I_i} \mathbf{P}(\overline{A_j} \cap \overline{A_i})\right)$$

17. On rappelle que $\Delta_n = \sum_{(i,j)\in\mathcal{E}} E(Y_iY_j)$ où $\mathcal{E}$ a été défini dans la partie 1 à la suite de la question 3.

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

(b) En conclure que :

$$\mathbf{P}(Z_n = 0) \leqslant \exp\left(-E(Z_n) + \frac{\Delta_n}{2}\right) \qquad \text{(inégalité de Janson)}$$

(c) En déduire l'encadrement :

$$(1 - p^3)^{\binom{n}{3}} \leqslant \mathbf{P}(Z_n = 0) \leqslant \exp\left(-\binom{n}{3}p^3 + \frac{a_n}{2}p^5\right)$$

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 que la limite de $\mathbf{P}(Z_n = 0)$ 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, pout tout $(x_1, \ldots, x_m) \in \{0, 1\}^m$,

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

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

► Soit $i \geqslant 2$. On reprend les notations de la question 16.

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

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

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