



PRÉPASIMPLE

prepasimple.fr

HEC / ESSEC

Mathématiques — Mathématiques appliquées

2024

Mathématiques appliquées

Mathématiques appliquées — Prépa ECG

Annales de concours · Corrigé

Partie I - Nombre aléatoire de triangles

Toutes les variables de cette partie prennent un nombre fini de valeurs. Leurs espérances, variances et covariances existent donc. Toutes les sommes utilisées sont finies.

1. Les éléments de $\mathcal{T}$ sont les parties à trois éléments d'un ensemble à n éléments. Ainsi

$$r = \binom{n}{3} = \frac{n(n-1)(n-2)}{6}.$$

On peut aussi retrouver ce calcul en choisissant successivement trois sommets distincts : les $n(n-1)(n-2)$ listes obtenues décrivent chacune des parties dans ses six ordres possibles.

2.a) Les trois variables $T_{u,v}$, $T_{v,w}$ et $T_{u,w}$ valent 0 ou 1. Leur produit vaut 1 exactement lorsque les trois arêtes du triangle sont présentes. Il vaut 0 sinon. C'est précisément la définition de Y_k , donc $Y_k = T_{u,v}T_{v,w}T_{u,w}$.

2.b) Les trois arêtes sont distinctes ; les trois variables associées sont donc indépendantes par hypothèse. Par conséquent,

$$P(Y_k = 1) = P(T_{u,v} = 1, T_{v,w} = 1, T_{u,w} = 1) = p^3.$$

Comme Y_k est à valeurs dans $\{0, 1\}$, elle suit la loi de Bernoulli de paramètre p^3 ; en particulier $E(Y_k) = p^3$ et $V(Y_k) = p^3(1 - p^3)$.

2.c) Pour chaque réalisation, chaque triangle présent contribue exactement 1 à la somme des indicatrices, les autres contribuent 0. Ainsi

$$Z_n = \sum_{k=1}^r Y_k.$$

La linéarité de l'espérance, qui ne demande aucune indépendance, donne

$$E(Z_n) = \sum_{k=1}^r E(Y_k) = rp^3 = \binom{n}{3} p^3.$$

3.a) Méthode 1 — Bilinéarité de la covariance.

Pour toute variable ayant un moment d'ordre deux, $V(U) = \text{Cov}(U, U)$. En utilisant la linéarité de la covariance dans chacun de ses deux arguments,

$$V(Z_n) = \text{Cov} \left(\sum_{i=1}^r Y_i, \sum_{j=1}^r Y_j \right) = \sum_{i=1}^r \sum_{j=1}^r \text{Cov}(Y_i, Y_j).$$

Méthode 2 — Développement du second moment.

La formule $V(U) = E(U^2) - E(U)^2$ et la distributivité donnent

$$V(Z_n) = \sum_{i,j=1}^r E(Y_i Y_j) - \sum_{i,j=1}^r E(Y_i) E(Y_j) = \sum_{i,j=1}^r \text{Cov}(Y_i, Y_j).$$

La première méthode exploite directement les propriétés de la covariance ; la seconde explique l'origine des termes croisés à partir du second moment.

3.b) Si $(i, j) \in \mathcal{F}$, les deux triplets ont au plus un sommet commun. Ils n'ont donc aucune arête commune. Les variables de Bernoulli d'arêtes utilisées dans Y_i et celles utilisées dans Y_j constituent deux groupes disjoints d'une famille indépendante. Par le lemme des coalitions, Y_i et Y_j sont indépendantes, donc leur covariance est nulle.

Les couples d'indices se répartissent en trois ensembles disjoints : la diagonale $i = j$, $\mathcal{E}$ et $\mathcal{F}$. Sur la diagonale, $\text{Cov}(Y_i, Y_i) = V(Y_i)$. D'où

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

3.c) Pour chaque i , $V(Y_i) = p^3(1 - p^3)$. Pour $(i, j) \in \mathcal{E}$, la formule de covariance donne

$$\text{Cov}(Y_i, Y_j) = E(Y_i Y_j) - E(Y_i) E(Y_j) = E(Y_i Y_j) - p^6.$$

Il y a a_n couples dans $\mathcal{E}$. En sommant,

$$V(Z_n) = rp^3(1 - p^3) + \sum_{(i,j) \in \mathcal{E}} E(Y_i Y_j) - a_n p^6.$$

4. Lorsque $i \equiv j$, les deux triangles distincts partagent exactement une arête. Leur réunion contient donc cinq arêtes distinctes. Le produit $Y_i Y_j$ est l'indicatrice de la présence de ces cinq arêtes ; l'arête commune n'est comptée qu'une fois puisque $T_{u,v}^2 = T_{u,v}$. L'indépendance des cinq variables d'arêtes donne $E(Y_i Y_j) = p^5$.

En sommant sur les a_n couples, $\Delta_n = a_n p^5$. La formule précédente devient

$$V(Z_n) = \binom{n}{3}(p^3 - p^6) + a_n(p^5 - p^6).$$

5.a) Question adaptée au programme actuel : l'indication explicite les choix successifs, sans faire appel à un chapitre général de dénombrement.

Il y a $\binom{n}{2}$ choix pour la paire non ordonnée $\{u, v\}$. Il reste $n - 2$ choix pour w , puis $n - 3$ choix pour y . Les rôles de w et y sont distingués. Le nombre demandé vaut

$$\binom{n}{2}(n - 2)(n - 3) = \frac{n(n - 1)(n - 2)(n - 3)}{2}.$$

Pour $n = 3$, aucun choix de quatre sommets distincts n'est possible ; la formule donne bien 0.

5.b) Méthode 1 — Utiliser le comptage précédent.

Au triplet $(\{u, v\}, w, y)$ on associe le couple ordonné de triangles $(\{u, v, w\}, \{u, v, y\})$. Ces triangles ont exactement la paire $\{u, v\}$ en commun. Réciproquement, un couple ordonné de triangles distincts ayant deux sommets communs permet de retrouver uniquement cette paire, puis le sommet w du premier triangle et le sommet y du second. La correspondance est donc bijective. Ainsi

$$a_n = \binom{n}{2}(n - 2)(n - 3) = \frac{n(n - 1)(n - 2)(n - 3)}{2}.$$

Méthode 2 — Regrouper par ensemble de quatre sommets.

Pour $n \geq 4$, choisissons les quatre sommets qui forment la réunion des deux triangles : $\binom{n}{4}$ possibilités. Parmi ces quatre sommets, il y a six paires possibles pour l'arête commune. Les deux sommets restants peuvent jouer les rôles ordonnés de w et y dans deux ordres. Ainsi $a_n = 12\binom{n}{4}$, ce qui redonne la formule annoncée. Pour $n = 3$, $a_3 = 0$ directement.

La première méthode compte par l'arête commune ; la seconde par la réunion des deux triangles, ce qui explique le coefficient 12.

Partie 2 - Étude informatique

On suppose, conformément au modèle, que les listes d'adjacence décrivent un graphe simple non orienté : pas de boucle, pas de voisin répété et présence réciproque des voisins.

6. Les deux autres sommets d'un triangle contenant s doivent être deux voisins distincts de s , et ils doivent être reliés entre eux. On parcourt les indices $i < j$ des voisins ; chaque paire est examinée une seule fois.

Python

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

La condition teste l'arête entre les deux voisins. Avec zéro ou un voisin, la boucle intérieure n'effectue aucun comptage, et la fonction retourne correctement 0.

7. On compte les triangles contenant le dernier sommet, puis on supprime ce sommet. Chaque triangle est compté exactement lors de la suppression de son sommet de plus grand numéro. Aucun des autres triangles n'est supprimé à cette étape. Le nombre de sommets décroît d'une unité à chaque passage.

Voici l'ensemble des trois fonctions, sans import nécessaire :

Python

```
def supprimeDer(L):
    s = len(L) - 1
    L.pop()
    for a in L:
        if s in a:
            a.remove(s)

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

def nbTriangles(L):
    cpt = 0
    while len(L) >= 3:
        s = len(L) - 1
        cpt += triangle2s(s, L)
        supprimeDer(L)
    return cpt
```

Quand il reste moins de trois sommets, aucun triangle ne subsiste. La fonction retourne donc le nombre total demandé. Elle modifie la liste L reçue : cette conséquence de l'utilisation de `supprimeDer` doit être prise en compte si l'on veut réutiliser le graphe.

Pour conserver l'entrée, on peut appeler `nbTriangles([a[:] for a in L])`, qui copie chaque liste de voisins.

Sur le graphe de l'énoncé, les triangles sont $\{0, 2, 4\}$, $\{1, 2, 4\}$ et $\{1, 3, 4\}$, ce qui donne bien 3.

8. Pour chaque itération, `graphe(n, 1/n)` construit un graphe avec n sommets et une probabilité d'arête $p = 1/n$. `cpt` compte les graphes sans triangle parmi les 1000 graphes générés. La sortie est leur fréquence : elle appartient à $\{0, 0,001, \dots, 1\}$ et fournit une estimation par simulation de $P(Z_n = 0)$ pour $p = 1/n$.

La fonction est :

Python

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

Elle utilise `nbTriangles` ci-dessus et le générateur `graphe` supposé fourni dans le sujet. Pour des générations indépendantes, les indicatrices « aucun triangle » sont indépendantes de même loi de Bernoulli, de paramètre $q_n = P(Z_n = 0)$. La fréquence a alors pour espérance q_n et pour variance $q_n(1 - q_n)/1000$. Une simulation ne démontre pas une identité sur q_n .

Les fonctions de comptage ont été exécutées sur tous les graphes simples de 3 à 6 sommets et confrontées à un comptage direct des triplets. La fonction de simulation a aussi été exécutée avec un générateur réalisant les tirages d'arêtes indépendants du préambule ; sa sortie dépend des tirages.

Partie 3 - Inégalité de Harris

Les variables de cette partie sont finies et les fonctions sont définies sur des ensembles finis : toutes les espérances employées existent, sans condition supplémentaire de convergence. Le transfert d'ordre quelconque est explicitement admis dans le sujet ; Harris est ici un résultat à démontrer, et non un théorème de cours supposé connu.

9.a) Si $x \geq y$, les deux différences $f(x) - f(y)$ et $g(x) - g(y)$ sont positives ou nulles. Si $x \leq y$, elles sont toutes deux négatives ou nulles. Dans les deux cas, leur produit est positif ou nul.

9.b) En développant l'inégalité de 9.a, on obtient, pour chaque $x \in X(\Omega)$,

$$f(x)g(x) + f(y)g(y) \geq f(x)g(y) + f(y)g(x).$$

On multiplie par $P(X = x) \geq 0$ et on somme en x . Le théorème de transfert et

$$\sum_x P(X = x) = 1$$

donnent

$$E(f(X)g(X)) + f(y)g(y) \geq g(y)E(f(X)) + f(y)E(g(X)).$$

9.c) Méthode 1 — Moyenner à nouveau l'inégalité précédente.

On multiplie l'inégalité de 9.b par $P(X = y)$ et on somme en y . Le membre de gauche vaut $2E(f(X)g(X))$. Le membre de droite vaut $2E(f(X))E(g(X))$. En divisant par 2, on obtient (H_1) .

Méthode 2 — Une identité de covariance.

Notons $p_x = P(X = x)$. En développant les quatre termes de la somme finie et en

utilisant

$$\sum_x p_x = 1,$$

on vérifie

$$E(f(X)g(X)) - E(f(X))E(g(X)) = \frac{1}{2} \sum_{x,y \in X(\Omega)} (f(x) - f(y))(g(x) - g(y))p_x p_y.$$

En effet, les termes $f(x)g(x)$ et $f(y)g(y)$ apportent chacun $E(f(X)g(X))$, tandis que les deux termes croisés apportent chacun $-E(f(X))E(g(X))$. Tous les termes de la somme sont positifs ou nuls par 9.a : la covariance est donc positive ou nulle. Cela démontre aussi (H_1) .

La première méthode suit les moyennes successives demandées ; la seconde met en évidence le lien entre les variations dans le même sens et le signe de la covariance.

10.a) Pour alléger les écritures, posons $D_\ell = X_\ell(\Omega)$ et $\mathbf{z} = (z_1, \dots, z_k)$. Le transfert d'ordre $k+1$ et l'indépendance donnent

$$E(Y_{k+1}Z_{k+1}) = \sum_{x \in D_{k+1}} \sum_{\mathbf{z} \in D_1 \times \dots \times D_k} f(\mathbf{z}, x)g(\mathbf{z}, x) \left(\prod_{\ell=1}^k P(X_\ell = z_\ell) \right) P(X_{k+1} = x).$$

L'indépendance des k premières variables permet d'identifier le produit intérieur à leur probabilité jointe. Le transfert d'ordre k transforme donc la somme intérieure en $E(f(X_1, \dots, X_k, x)g(X_1, \dots, X_k, x))$. D'où exactement

$$E(Y_{k+1}Z_{k+1}) = \sum_{x \in D_{k+1}} E(f(X_1, \dots, X_k, x)g(X_1, \dots, X_k, x))P(X_{k+1} = x).$$

10.b) Pour x fixé dans D_{k+1} , les fonctions de k variables $\mathbf{z} \mapsto f(\mathbf{z}, x)$ et $\mathbf{z} \mapsto g(\mathbf{z}, x)$ sont k -croissantes. Les k premières variables sont indépendantes et finies. L'hypothèse de récurrence (H_k) s'applique donc et donne

$$E(f(X_1, \dots, X_k, x)g(X_1, \dots, X_k, x)) \geq u(x)v(x).$$

10.c) Si $x \leq y$ appartiennent à D_{k+1} , la croissance de f par rapport à sa dernière variable donne, pour chaque réalisation,

$$f(X_1, \dots, X_k, x) \leq f(X_1, \dots, X_k, y).$$

La croissance de l'espérance entraîne $u(x) \leq u(y)$. Le même raisonnement donne $v(x) \leq v(y)$. Ainsi u et v sont croissantes.

On multiplie l'inégalité de 10.b par $P(X_{k+1} = x)$ puis on somme. Avec 10.a et le transfert en une variable,

$$E(Y_{k+1}Z_{k+1}) \geq \sum_x u(x)v(x)P(X_{k+1} = x) = E(u(X_{k+1})v(X_{k+1})).$$

10.d) La propriété (H_1) , appliquée à X_{k+1} et aux fonctions croissantes u et v , fournit

$$E(u(X_{k+1})v(X_{k+1})) \geq E(u(X_{k+1}))E(v(X_{k+1})).$$

Le même développement par transfert que celui de 10.a, appliqué seulement à f , donne

$$E(u(X_{k+1})) = \sum_x \sum_{\mathbf{z}} f(\mathbf{z}, x) \left(\prod_{\ell=1}^k P(X_{\ell} = z_{\ell}) \right) P(X_{k+1} = x) = E(Y_{k+1}).$$

De même, $E(v(X_{k+1})) = E(Z_{k+1})$. On obtient $E(Y_{k+1}Z_{k+1}) \geq E(Y_{k+1})E(Z_{k+1})$: c'est (H_{k+1}) .

La propriété est vraie au rang 1 et héréditaire. Elle est donc vraie pour tout entier $k \geq 1$.

10.e) Si f et g sont k -décroissantes, $-f$ et $-g$ sont k -croissantes. En appliquant (H_k) à ces deux fonctions,

$$E((-Y_k)(-Z_k)) \geq E(-Y_k)E(-Z_k),$$

soit $E(Y_k Z_k) \geq E(Y_k)E(Z_k)$. Le sens est conservé.

Si f est k -croissante et g k -décroissante, on applique (H_k) à f et $-g$:

$$-E(Y_k Z_k) \geq -E(Y_k)E(Z_k), \text{ donc}$$

$$E(Y_k Z_k) \leq E(Y_k)E(Z_k).$$

Le même résultat vaut en échangeant les rôles. Il s'agit d'une inégalité large ; l'égalité reste possible, par exemple si une fonction est constante.

Partie 4 - Inégalité de Janson et application

- 11.** Si toutes les variables d'arêtes valent 0, le graphe n'a aucune arête et ne peut contenir aucun triangle. Cela prouve l'inclusion. Il y a $\binom{n}{2}$ arêtes possibles et leurs variables sont indépendantes. D'où

$$P(Z_n = 0) \geq P\left(\bigcap_{u < v} [T_{u,v} = 0]\right) = \prod_{u < v} (1 - p) = (1 - p)^{\binom{n}{2}} > 0,$$

puisque $0 < p < 1$.

- 12.** Comme $Y_i \in \{0, 1\}$, la variable $1 - Y_i$ est l'indicatrice de $[Y_i = 0]$; son espérance vaut donc la probabilité de cet événement.

Les Y_k sont positifs ou nuls. Leur somme $Z_{n,i}$ est nulle exactement lorsqu'ils sont tous nuls pour $k \leq i$. Le produit

$$\prod_{k=1}^i (1 - Y_k)$$

est donc l'indicatrice de $[Z_{n,i} = 0]$. En prenant les espérances,

$$P(Y_i = 0) = E(1 - Y_i) \text{ et}$$

$$P(Z_{n,i} = 0) = E\left(\prod_{k=1}^i (1 - Y_k)\right).$$

13.a) Ordonnons les $m = \binom{n}{2}$ arêtes possibles. Sur $\{0, 1\}^m$, Y_k est le produit des trois coordonnées des arêtes de t_k . Augmenter une coordonnée de 0 à 1 ne peut diminuer ce produit : la fonction est m -croissante, même si elle ne dépend pas des autres coordonnées.

Par conséquent, $1 - Y_i$ est m -décroissante et à valeurs dans $\{0, 1\}$. Lorsque l'on augmente une coordonnée, chacun des facteurs positifs ou nuls du produit

$$\prod_{k=1}^{i-1} (1 - Y_k)$$

diminue ou reste constant. Leur produit diminue ou reste constant aussi. Il est donc m -décroissant. La non-négativité des facteurs est essentielle pour cette dernière affirmation.

13.b) Les variables d'arêtes sont indépendantes. La version décroissante de Harris établie en 10.e s'applique aux deux fonctions de 13.a. Avec 12,

$$P(Z_{n,i} = 0) = E \left((1 - Y_i) \prod_{k=1}^{i-1} (1 - Y_k) \right) \geq P(Y_i = 0) P(Z_{n,i-1} = 0).$$

En itérant de $i = 2$ à r et en utilisant $P(Z_{n,1} = 0) = P(Y_1 = 0)$, on obtient

$$P(Z_n = 0) \geq \prod_{k=1}^r P(Y_k = 0) = (1 - p^3)^r = (1 - p^3)^{\binom{n}{3}}.$$

Pour $n = 3, r = 1$: l'itération est vide et le résultat est une égalité.

14. Méthode 1 — La récurrence demandée.

Pour deux événements, la formule d'une réunion donne

$$P(B_1 \cup B_2) = P(B_1) + P(B_2) - P(B_1 \cap B_2) \leq P(B_1) + P(B_2).$$

Supposons l'inégalité vraie pour k événements. En appliquant le cas de deux événements à $\bigcup_{i=1}^k B_i$ et B_{k+1} , puis l'hypothèse de récurrence,

$$P\left(\bigcup_{i=1}^{k+1} B_i\right) \leq P\left(\bigcup_{i=1}^k B_i\right) + P(B_{k+1}) \leq \sum_{i=1}^{k+1} P(B_i).$$

La propriété est donc vraie pour tout $k \geq 2$.

Méthode 2 — Comparer des indicatrices.

Notons $\mathbf{1}_B$ l'indicatrice d'un événement B . Pour chaque réalisation, si aucun B_i ne se produit, les deux membres de l'inégalité suivante valent 0 ; si au moins un se produit, le membre de droite est au moins 1. Ainsi

$$\mathbf{1}_{\bigcup_{i=1}^k B_i} \leq \sum_{i=1}^k \mathbf{1}_{B_i}.$$

Toutes ces variables sont finies. La croissance et la linéarité de l'espérance donnent immédiatement l'inégalité de Boole. Cette seconde méthode illustre l'utilité des indicatrices déjà employées pour compter les triangles.

La définition de E_A placée après cette question est une définition par somme finie sous une nouvelle probabilité, et ne nécessite aucune théorie générale de l'espérance conditionnelle. Pour y utiliser Harris, on vérifiera en 20.b l'indépendance sous cette probabilité.

15. Posons $D = A \cap B \cap C$. Les deux hypothèses garantissent aussi $P(C) > 0$. Par les définitions,

$$P_C(A)P_{A \cap C}(B) = \frac{P(A \cap C)}{P(C)} \frac{P(D)}{P(A \cap C)} = \frac{P(D)}{P(C)}.$$

Or $0 < P(B \cap C) \leq P(C)$. Comme $P(D) \geq 0$,

$$P_{B \cap C}(A) = \frac{P(D)}{P(B \cap C)} \geq \frac{P(D)}{P(C)},$$

ce qui est l'inégalité demandée, y compris lorsque $P(D) = 0$.

16.a) Pour $j \in J_i$, le triangle t_j ne partage aucune arête avec t_i . L'événement C_i dépend donc seulement de variables d'arêtes extérieures aux trois arêtes de t_i , même si les triangles de J_i peuvent partager entre eux des arêtes. A_i dépend seulement des trois variables d'arêtes de t_i . Par le lemme des coalitions, A_i et C_i sont indépendants; $\overline{A_i}$ et C_i le sont aussi.

On applique 15 avec $A = \overline{A_i}$, $B = B_i$ et $C = C_i$. Les conditionnements sont autorisés par le sujet. Il vient

$$P_{B_i \cap C_i}(\overline{A_i}) \geq P_{C_i}(\overline{A_i})P_{\overline{A_i} \cap C_i}(B_i) = P(\overline{A_i})P_{\overline{A_i} \cap C_i}(B_i).$$

16.b) Le complément de B_i est $\bigcup_{j \in I_i} \overline{A_j}$. Boole, sous la probabilité $Q = P_{\overline{A_i} \cap C_i}$, donne

$$Q(B_i) = 1 - Q\left(\bigcup_{j \in I_i} \overline{A_j}\right) \geq 1 - \sum_{j \in I_i} Q(\overline{A_j}).$$

Si I_i est vide, $B_i = \Omega$ et la formule donne $1 \geq 1$; s'il contient un seul élément, elle est une égalité. Sinon, on utilise 14.

16.c) En utilisant provisoirement (1) dans la majoration de la somme de 16.b,

$$P_{\overline{A_i} \cap C_i}(B_i) \geq 1 - \sum_{j \in I_i} P_{\overline{A_i}}(\overline{A_j}).$$

On multiplie par $P(\overline{A_i}) = p^3 > 0$ et on utilise 16.a. Enfin, les probabilités de A_i et $\overline{A_i}$

sous $P_{B_i \cap C_i}$ ont pour somme 1. Ainsi

$$P_{B_i \cap C_i}(A_i) \leq 1 - P(\overline{A_i}) \left(1 - \sum_{j \in I_i} P_{\overline{A_i}}(\overline{A_j}) \right).$$

La justification de (1) sera donnée en 20.d ; aucune indépendance sous un conditionnement n'a été supposée à ce stade.

16.d) Méthode 1 — Une tangente à l'exponentielle.

La fonction $t \mapsto e^t$ est convexe sur $\mathbb{R}$, car sa dérivée seconde est positive. Son graphe est donc au-dessus de sa tangente en 0 : $e^t \geq 1 + t$. Pour $t = -x$, cela donne $1 - x \leq e^{-x}$.

Méthode 2 — Étudier une différence.

La fonction $h(x) = e^{-x} - 1 + x$ est dérivable et $h'(x) = 1 - e^{-x}$. Cette dérivée est négative sur $] -\infty, 0[$ et positive sur $]0, +\infty[$. Ainsi h atteint son minimum en 0, où $h(0) = 0$. On retrouve $1 - x \leq e^{-x}$ pour tout réel x .

Les deux méthodes reconnaissent la même inégalité usuelle, soit par la convexité, soit par un tableau de variations.

Pour l'appliquer à 16.c, posons

$$x = P(\overline{A_i}) - \sum_{j \in I_i} P(\overline{A_i}) P_{\overline{A_i}}(\overline{A_j}) = P(\overline{A_i}) - \sum_{j \in I_i} P(\overline{A_j} \cap \overline{A_i}).$$

La majoration de 16.c est $1 - x$, donc

$$P_{B_i \cap C_i}(A_i) \leq \exp \left(-P(\overline{A_i}) + \sum_{j \in I_i} P(\overline{A_j} \cap \overline{A_i}) \right).$$

17.a) Z_n est une somme d'indicateurs positives ou nulles. L'événement $[Z_n = 0]$ est donc $\bigcap_{i=1}^r A_i$. La formule des probabilités composées donne

$$P\left(\bigcap_{i=1}^r A_i\right) = P(A_1) \prod_{i=2}^r P_{\bigcap_{j=1}^{i-1} A_j}(A_i).$$

Chaque événement de conditionnement contient l'événement « aucune arête », de probabilité strictement positive d'après 11. Comme $\bigcap_{j=1}^{i-1} A_j = B_i \cap C_i$, on obtient la formule demandée. Pour $r = 1$, le produit vide vaut 1.

17.b) Pour le premier facteur, $P(A_1) = 1 - P(\overline{A_1}) \leq e^{-P(\overline{A_1})}$. Pour les suivants, on utilise 16.d. Tous les facteurs et leurs majorants sont positifs ou nuls ; on peut multiplier les inégalités. D'où

$$P(Z_n = 0) \leq \exp\left(-\sum_{i=1}^r P(\overline{A_i}) + \sum_{i=2}^r \sum_{j \in I_i} P(\overline{A_j} \cap \overline{A_i})\right).$$

Comme $\overline{A_i} = [Y_i = 1]$, $P(\overline{A_i}) = E(Y_i)$ et $P(\overline{A_j} \cap \overline{A_i}) = E(Y_i Y_j)$. La première somme vaut donc $E(Z_n)$.

La seconde somme porte sur les couples de triangles partageant une arête, avec $j < i$. Dans $\mathcal{E}$, les deux couples (i, j) et (j, i) figurent ensemble, avec la même valeur de $E(Y_i Y_j)$. Il n'y a pas de couple diagonal. La double somme vaut donc $\Delta_n/2$. Finalement,

$$P(Z_n = 0) \leq \exp\left(-E(Z_n) + \frac{\Delta_n}{2}\right).$$

17.c) En remplaçant $E(Z_n)$ et Δ_n par les expressions de 2.c et 4, puis en utilisant la borne inférieure de 13.b, on obtient

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

18.a) On écrit exactement

$$\binom{n}{3} \left(\frac{c}{n}\right)^3 = \frac{c^3}{6} \left(1 - \frac{1}{n}\right) \left(1 - \frac{2}{n}\right) \rightarrow \frac{c^3}{6},$$

et

$$\frac{a_n}{2} \left(\frac{c}{n}\right)^5 = \frac{c^5}{4n} \left(1 - \frac{1}{n}\right) \left(1 - \frac{2}{n}\right) \left(1 - \frac{3}{n}\right) \rightarrow 0.$$

En soustrayant ces deux suites convergentes, l'expression demandée tend donc vers $-c^3/6$.

18.b) Le logarithme est défini dès que $n > c$, ce qui suffit pour l'étude de la limite.

Méthode 1 — Équivalent issu de la dérivabilité du logarithme.

Posons $z_n = c^3/n^3$. Alors $z_n \rightarrow 0^+$ et $\ln(1 - z_n)/(-z_n) \rightarrow 1$, car $z \mapsto \ln(1 - z)$ est dérivable en 0, de valeur 0 et de dérivée -1 . On écrit le produit exact

$$\binom{n}{3} \ln(1 - z_n) = -\binom{n}{3} z_n \frac{\ln(1 - z_n)}{-z_n}.$$

Le premier facteur tend vers $-c^3/6$ par 18.a et le second vers 1. La limite vaut donc $-c^3/6$.

Méthode 2 — Encadrement logarithmique.

Pour $0 < z < 1$, on a $\ln(1 - z) = -\int_0^z \frac{dt}{1-t}$. Sur $[0, z]$, $1 \leq (1-t)^{-1} \leq (1-z)^{-1}$. En intégrant et en changeant le signe,

$$-\frac{z}{1-z} \leq \ln(1 - z) \leq -z.$$

En multipliant par $\binom{n}{3} > 0$ et en prenant $z = z_n$, les deux bornes tendent vers $-c^3/6$, puisque $\binom{n}{3} z_n \rightarrow c^3/6$ et $1 - z_n \rightarrow 1$. Le théorème d'encadrement donne la limite.

La première méthode est courte lorsqu'on reconnaît une dérivée en 0; la seconde fournit en plus des bornes valables pour chaque $n > c$.

18.c) Pour $n \geq 3$ et $n > c$, le paramètre $p = c/n$ appartient bien à $]0, 1[$. La borne inférieure de 17.c s'écrit

$$\exp\left(\binom{n}{3} \ln\left(1 - \frac{c^3}{n^3}\right)\right).$$

Par 18.b et la continuité de l'exponentielle, elle tend vers $e^{-c^3/6}$. La borne supérieure de 17.c tend vers la même valeur par 18.a. Le théorème d'encadrement donne donc

$$\lim_{n \rightarrow +\infty} P(Z_n = 0) = e^{-c^3/6} \text{ pour } p = c/n.$$

La loi du graphe dépend ici de n ; on n'a utilisé aucune règle générale de convergence de variables aléatoires, seulement l'encadrement de probabilités réelles établi plus haut.

19. Dans fonctionMystere(100), $p = 1/100$, donc $c = 1$. La probabilité limite vaut $e^{-1/6}$. Avec l'approximation suggérée,

$$e^{-1/6} \simeq 1 - \frac{1}{6} + \frac{1}{72} = \frac{61}{72} \simeq 0,84722.$$

La valeur plus précise est $e^{-1/6} \simeq 0,84648$. La fréquence observée 0,849 est proche de ces valeurs; elle est donc cohérente avec le résultat.

Il faut distinguer les deux approximations : la question 18 concerne une limite quand n tend vers l'infini, tandis que 0,849 est une fréquence sur 1000 simulations avec $n = 100$. À titre d'échelle, pour une probabilité proche de 0,8465 et des tirages indépendants, l'écart-type de cette fréquence est proche de $\sqrt{0,8465(1 - 0,8465)/1000} \simeq 0,0114$. La proximité observée n'est donc pas surprenante et ne constitue pas une preuve du résultat asymptotique.

20.a) Question adaptée au programme actuel : l'hypothèse $P(A) > 0$ est explicitée afin que toutes les divisions définissant P_A soient licites. On ne suppose aucune théorie générale de l'espérance conditionnelle.

Notons $D_x = \bigcap_{\ell=1}^m [X_\ell = x_\ell]$. Les variables étant à valeurs dans $\{0, 1\}$, $A = \bigcap_{i \in I} [X_i = 1]$. Si

$$\prod_{i \in I} x_i \neq 1,$$

alors $D_x \cap A = \emptyset$ et $P_A(D_x) = 0$.

Si

$$\prod_{i \in I} x_i = 1,$$

tous les x_i avec $i \in I$ valent 1, donc $D_{\mathbf{x}} \subset A$. L'indépendance fournit

$$P(A) = \prod_{i \in I} P(X_i = 1) > 0,$$

et

$$P(D_{\mathbf{x}}) = \left(\prod_{i \in I} P(X_i = 1) \right) \left(\prod_{i \in J} P(X_i = x_i) \right).$$

Le quotient $P(D_{\mathbf{x}})/P(A)$ est donc

$$\prod_{i \in J} P(X_i = x_i),$$

ce qui donne la formule demandée. Si I est vide, $A = \Omega$ et la formule redevient simplement l'indépendance initiale. Si J est vide, le produit correspondant vaut 1 et toutes les coordonnées sont fixées à 1 sous P_A .

20.b) Si $i \in I$, la variable X_i est constante égale à 1 sous P_A : $P_A(X_i = x_i) = 1$ si $x_i = 1$, et 0 sinon.

Si $j \in J$, X_j est indépendante des variables dont dépend A , donc

$$P_A(X_j = x_j) = P(X_j = x_j).$$

Ainsi le produit des m probabilités marginales sous P_A est nul si une coordonnée de I

vaut 0, et vaut

$$\prod_{j \in J} P(X_j = x_j)$$

sinon. C'est exactement la probabilité jointe obtenue en 20.a. La caractérisation de l'indépendance d'une famille finie donne donc l'indépendance de $X_1, \dots, X_m$ sous P_A . Les coordonnées fixées sont des Bernoulli de paramètre 1, ce qui ne gêne pas cette indépendance.

20.c) Fixons $i \geq 2$ et $j < i$. Pour alléger les écritures, posons $Q = P_{\overline{A_i}}$. Les événements de conditionnement ont une probabilité strictement positive, comme admis dans le sujet. Par définition et simplification des quotients,

$$P_{\overline{A_i} \cap C_i}(\overline{A_j}) = \frac{P(\overline{A_j} \cap \overline{A_i} \cap C_i)}{P(\overline{A_i} \cap C_i)} = \frac{Q(\overline{A_j} \cap C_i)}{Q(C_i)}.$$

Or $1_{\overline{A_j}} = Y_j$ et

$$1_{C_i} = \prod_{k \in J_i} (1 - Y_k).$$

Le numérateur est donc l'espérance, sous Q , du produit de ces deux indicatrices. Cela donne

$$P_{\overline{A_i} \cap C_i}(\overline{A_j}) = \frac{E_{\overline{A_i}} \left(Y_j \prod_{k \in J_i} (1 - Y_k) \right)}{P_{\overline{A_i}}(C_i)}.$$

Cette identité reste valable si J_i est vide : le produit est alors 1 et $C_i = \Omega$.

20.d) L'événement $\overline{A_i}$ est l'événement « les trois arêtes de t_i sont présentes » ; il s'écrit comme le produit des trois variables de ces arêtes égal à 1. Sa probabilité est $p^3 > 0$. On applique donc 20.b à la famille des $m = \binom{n}{2}$ variables d'arêtes, avec I l'ensemble des trois indices concernés. Comme $n \geq 3$, $m \geq 3$, ce qui respecte l'hypothèse $m > 2$ de 20.a.

Les variables d'arêtes restent indépendantes sous $Q = P_{\overline{A_i}}$; trois d'entre elles y sont simplement constantes égales à 1. La fonction définissant Y_j est croissante par rapport à chacune des coordonnées. Celle définissant

$$W = \prod_{k \in J_i} (1 - Y_k)$$

est décroissante, d'après le même argument qu'en 13.a, y compris sur l'ensemble des valeurs des coordonnées après conditionnement.

La version de Harris avec sens opposés, démontrée en 10.e, s'applique donc à ces variables indépendantes sous Q et donne

$$E_{\overline{A_i}}(Y_j W) \leq E_{\overline{A_i}}(Y_j) E_{\overline{A_i}}(W).$$

Comme $Y_j = \mathbf{1}_{\overline{A_j}}$ et $W = \mathbf{1}_{C_i}$, cela signifie

$$E_{\overline{A_i}}(Y_j W) \leq P_{\overline{A_i}}(\overline{A_j}) P_{\overline{A_i}}(C_i).$$

On divise par $P_{\overline{A_i}}(C_i) > 0$ et on utilise 20.c. On obtient

$$P_{\overline{A_i} \cap C_i}(\overline{A_j}) \leq P_{\overline{A_i}}(\overline{A_j}).$$

C'est exactement (1), ce qui achève la justification de la majoration de Janson sans supposer l'indépendance sous un conditionnement arbitraire.

Relevé des adaptations

- Préambule de la partie I, figure précédant 1 : la figure est retranscrite en SVG dans l'énoncé à partir de l'original, avec les valeurs mesurées. Les positions graphiques ont été relevées dans l'original; les sept arêtes et les cinq étiquettes sont conservées. Il ne s'agit pas d'un graphe à coordonnées mathématiques imposées.
- 5.a : ajout d'une indication sur les choix successifs de la paire commune puis des deux sommets ordonnés, afin de guider le comptage sans supposer une théorie générale de dénombrement.

- 14, préambule suivant la question : précision que E_A est défini uniquement par une somme finie sous P_A , et que Harris y nécessite l'indépendance des variables sous cette probabilité. La preuve de cette indépendance particulière est conservée en 20.b.
- 20.a : ajout explicite de $P(A) > 0$, nécessaire à la définition de P_A . Cette condition est vérifiée par $P(\overline{A_i}) = p^3 > 0$ dans l'application de 20.d.