

# Estimation de la taille d'un graphe par marches aléatoires

Matthieu DINOT, Dorian BOULLY

## Préliminaire : spectre des matrices à coefficients positifs

Pour une matrice réelle  $X$  de taille quelconque (en particulier  $X$  peut être un vecteur de  $\mathbb{R}^N$  ou une matrice  $(N, N)$ ), on note  $X \geq 0$  lorsque tous ses coefficients sont positifs ou nuls, et  $X > 0$  lorsque tous ses coefficients sont strictement positifs. On désigne par  $X(i, j)$  le coefficient d'indice  $(i, j)$  de  $X$ , et si  $X$  est un vecteur de  $\mathbb{R}^N$ , on note  $X(i)$  au lieu de  $X(i, 0)$ .

**Théorème 1** (Perron–Frobenius). *Soit  $A > 0 \in \mathcal{M}_N(\mathbb{R})$ . Alors*

- *$A$  possède une valeur propre réelle  $\rho > 0$  telle que toutes les autres valeurs propres de  $A$  sont de module strictement inférieur à  $\rho$ . On dit que  $\rho$  est dominante*
- *Il existe un vecteur propre  $X > 0$  associé à la valeur propre  $\rho$ .*
- *$\rho$  est une valeur propre simple de  $A$ , c'est à dire que sa multiplicité en tant que racine du polynôme caractéristique de  $A$  vaut 1. En particulier, le sous espace propre associé à  $\rho$  est de dimension 1.*

Ce théorème nous est utile pour étudier le spectre des matrices stochastiques. Une matrice  $A \in \mathcal{M}_N(\mathbb{R})$  est dite stochastique lorsque  $A \geq 0$  et

$$\forall i \in \{1, \dots, N\}, \quad \sum_{j=1}^N A(i, j) = 1.$$

**Corollaire 1.** *Soit  $A$  une matrice stochastique. On suppose qu'il existe un entier  $k$  tel que  $A^k > 0$  ( $A$  est alors dite régulière). Alors 1 est valeur propre simple et dominante de  $A$ .*

**Preuve du théorème 1.** Soit  $\|\cdot\|$  une norme sur  $\mathbb{R}^N$ . Posons  $\mathcal{S} = \{X \in \mathbb{R}^N \mid X \geq 0 \text{ et } \|X\| = 1\}$ , qui est compact car fermé borné de  $\mathbb{R}^N$ , et  $\Lambda = \{\lambda \in \mathbb{R} \mid \exists X \in \mathcal{S}, (A - \lambda I)X \geq 0\}$ . On voit que  $0 \in \Lambda$  et que  $\Lambda$  est majoré par la somme des coefficients de  $A$ . Soit donc  $\rho = \sup \Lambda \in \mathbb{R}$ . Il existe des suites  $(\lambda_n) \in \Lambda^{\mathbb{N}}$  et  $(X_n) \in \mathcal{S}^{\mathbb{N}}$  telles que

$$\lambda_n \xrightarrow{n \rightarrow +\infty} \rho \quad \text{et} \quad \forall n \in \mathbb{N}, (A - \lambda_n I)X_n \geq 0.$$

$\mathcal{S}$  étant compact, on peut supposer quitte à extraire que  $X_n$  converge vers  $X \in \mathcal{S}$ . En passant à la limite, il vient  $(A - \rho I)X \geq 0$ . Supposons que  $(A - \rho I)X \neq 0$ . Soit  $Y = \frac{AX}{\|AX\|} \in \mathcal{S}$ . On a par construction  $(A - \rho I)Y = \frac{1}{\|AX\|} A(A - \rho I)X > 0$ , et donc pour  $\varepsilon > 0$  assez petit,  $(A - (\rho + \varepsilon)I)Y > 0$ , d'où  $\rho + \varepsilon \in \Lambda$ , ce qui est absurde. Ainsi,  $AX = \rho X$ . De plus  $AX > 0$  car  $A > 0$ ,  $X \geq 0$  et  $X \neq 0$ , donc  $\rho > 0$  et  $X > 0$ .  $\rho$  est donc une valeur propre strictement positive de  $A$  et il existe un vecteur propre  $X > 0$  associé à  $\rho$ .

Montrons que  $\rho$  est dominante. Soit  $\lambda \in \mathbb{C}$  une valeur propre de  $A$  distincte de  $\rho$ , et  $Z \in \mathbb{C}^N$  un vecteur propre associé à  $\lambda$ . On a

$$\forall i, \quad \sum_j A(i, j) Z(j) = \lambda Z(i) \quad \text{donc} \quad \forall i, \quad |\lambda| |Z(i)| \leq \sum_j A(i, j) |Z(j)|.$$

Ainsi,  $|\lambda| \in \Lambda$ , donc  $|\lambda| \leq \rho$ . Montrons que l'inégalité est stricte. Pour cela, on suppose par l'absurde que  $|\lambda| = |\rho|$ . En notant  $|Z|$  le vecteur dont chaque composante est  $|Z(i)|$ , on a  $(A - |\lambda| I) |Z| \geq 0$ . Si  $(A - |\lambda| I) |Z| \neq 0$ , alors  $(A - |\lambda| I) A |Z| > 0$  et par suite  $|\lambda| < \rho$ , ce qui est contraire aux hypothèses. Donc

$$\forall i, \quad |\lambda| |Z(i)| = \sum_j A(i, j) |Z(j)|.$$

Mais alors il existe  $X \geq 0$ ,  $X \neq 0$  et  $\theta \in \mathbb{R}$  tels que  $Z = e^{i\theta} X$ . On en déduit que  $AX = \lambda X$ , et donc que  $\lambda$  est réelle, et strictement positive. On a donc  $\lambda = |\lambda| = \rho$ , ce qui est absurde.  $\rho$  est donc dominante.

Montrons enfin que  $\rho$  est simple. On commence par montrer que l'espace propre  $E_\rho$  est de dimension 1. Si ce n'est pas le cas, il existe deux vecteurs propres  $X > 0$  et  $Y \neq 0$  associés à  $\rho$  tels que  $(X, Y)$  est libre. Quitte à changer  $Y$  en  $-Y$ , on peut supposer qu'une de ses composantes est strictement positive. Le réel  $\alpha := \min \{X(i)/Y(i) : 1 \leq i \leq N \text{ et } Y(i) > 0\}$  est bien défini et vérifie  $X - \alpha Y \geq 0$  mais  $X - \alpha Y \not\geq 0$ . Or,  $X - \alpha Y$  est encore un vecteur propre de  $A$  associé à  $\rho$  (il est non nul par liberté de  $(X, y)$ ), et  $A > 0$  donc

$$\rho(X - \alpha Y) = A(X - \alpha Y) > 0,$$

ce qui est en contradiction avec  $X - \alpha Y \not\geq 0$ . Ainsi,  $E_\rho$  est de dimension 1. On améliore maintenant le résultat en montrant que  $\rho$  est simple. Soit  $(X_1 = X, \dots, X_N)$  une base de  $\mathbb{R}^N$  qui complète un vecteur propre  $X > 0$  associé à  $\rho$ . En notant  $P$  la matrice dont la  $i$ -ième colonne est  $X_i$ , on a

$$P^{-1}AP = \left( \begin{array}{c|ccc} \rho & \star & \cdots & \star \\ \hline 0 & & & \\ \vdots & & B & \\ 0 & & & \end{array} \right), \quad \text{avec } B \in \mathcal{M}_{N-1}(\mathbb{R}).$$

Les polynômes caractéristiques  $\chi_A$  et  $\chi_B$  de  $A$  et  $B$  vérifient  $\chi_A = (X - \rho)\chi_B$ . Supposons que  $\rho$  est racine d'ordre au moins 2 de  $\chi_A$ . Alors  $\rho$  est racine de  $\chi_B$ , donc valeur propre de  $B$ . Soit  $Z \in \mathbb{R}^{N-1}$  un vecteur propre de  $B$  associé à  $\rho$ . En posant  $Y = P \begin{pmatrix} 0 \\ Z \end{pmatrix}$ , il vient

$$AY = P \left( \begin{array}{c|ccc} \rho & \star & \cdots & \star \\ \hline 0 & & & \\ \vdots & & B & \\ 0 & & & \end{array} \right) \begin{pmatrix} 0 \\ Z \end{pmatrix} = P \begin{pmatrix} \alpha \\ BZ \end{pmatrix} = P \begin{pmatrix} \alpha \\ \rho Z \end{pmatrix} = \rho Y + \alpha X.$$

Le réel  $\alpha$  est non nul sinon  $(X, Y)$  formerait une famille libre de vecteurs propres de  $A$  associés à  $\rho$ . De plus, une récurrence immédiate donne  $A^k Y = \rho^k Y + k\alpha \rho^{k-1} X$  pour  $k \geq 0$ , et donc

$$A^k |Y| \geq |A^k Y| = |\rho^k Y + k\alpha \rho^{k-1} X| \geq \rho^{k-1} (k|\alpha| |X| - \rho |Y|) = \rho^{k-1} (k|\alpha| |X - \rho |Y||).$$

Comme  $\alpha \neq 0$  et  $X > 0$ , on peut prendre  $k$  suffisamment grand pour que  $k\alpha X - \rho |Y| > \rho Y$ . On obtient  $A^k |Y| > \rho^k |Y|$ . Comme  $A^k > 0$ , on en déduit en appliquant le début de la preuve que  $A^k$  possède une valeur propre strictement supérieure à  $\rho^k$ . C'est absurde car toute valeur propre de  $A^k$  est de la forme  $\lambda^k$  avec  $\lambda$  valeur propre de  $A$  (par trigonalisation de  $A$ ). On a donc montré que  $\rho$  est racine d'ordre 1 de  $\chi_A$ . Cela achève la preuve.  $\square$

**Preuve du corollaire 1.** Remarquons qu'une matrice  $M$  est stochastique si et seulement si le vecteur  $U := (1, \dots, 1)^\top$  vérifie  $MU = U$ . On en déduit facilement qu'un produit de matrices stochastiques est stochastique. De plus, on a :

$$\forall X \in \mathbb{R}^N, \forall i, \quad |(MX)(i)| = \left| \sum_j M(i, j)X(j) \right| \leq \sum_j M(i, j) |X(j)| \leq \sum_j M(i, j) \|X\|_\infty = \|X\|_\infty$$

donc  $\|M\| \leq 1$ . Soit  $k$  tel que  $A^k > 0$ . Comme  $A^k > 0$ , le théorème de Perron–Frobenius s'applique. Le fait que  $\|A^k\| \leq 1$  entraîne  $\rho \leq 1$ , mais on a aussi  $\rho \geq 1$  car  $A^k U = U$ . Donc  $\rho = 1$ . Ainsi, 1 est valeur propre simple et dominante de  $A^k$ . Puis on trigonalise  $A$ . Il existe une matrice inversible  $P$  telle que

$$A = P \begin{pmatrix} 1 & \star & \cdots & \star \\ 0 & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \star \\ 0 & \cdots & 0 & \lambda_N \end{pmatrix} P^{-1} \quad \text{donc} \quad A^k = P \begin{pmatrix} 1 & \star & \cdots & \star \\ 0 & \lambda_2^k & \ddots & \vdots \\ \vdots & \ddots & \ddots & \star \\ 0 & \cdots & 0 & \lambda_N^k \end{pmatrix} P^{-1}.$$

On en déduit que pour  $2 \leq i \leq N$ ,  $|\lambda_i|^k < 1$ , donc  $|\lambda_i| < 1$ . Ainsi, 1 est valeur propre dominante de  $A$  et l'espace propre associé est de dimension 1. En reprenant la fin de la preuve du théorème de Perron–Frobenius, on montre que 1 est en fait simple. D'où le résultat.  $\square$

Il est possible d'appliquer ces résultats à l'étude d'une marche aléatoire sur un graphe  $G = (V, E)$  de taille  $N$  (orienté ou non). On peut définir une telle marche aléatoire à l'aide d'une matrice de transition  $P$  telle que le coefficient  $P(i, j)$  est la probabilité de passer du sommet  $j$  au sommet  $i$ . On impose la contrainte  $P(i, j) \neq 0$  si et seulement si  $(j, i) \in E$ . Ainsi  $P$  représente bien une marche aléatoire sur le graphe  $G$ . On remarque que  $P^\top$  est stochastique et que si  $\pi_t$  représente le vecteur probabilité de présence à l'instant  $t$ , on a

$$\pi_{t+1} = P\pi_t.$$

Regardons maintenant sous quelles conditions il est possible d'appliquer le corollaire 1 à une matrice de transition  $P$ . Remarquons pour cela que  $P^k(i, j)$  est non nul si et seulement si il existe un chemin de longueur  $k$  de  $j$  vers  $i$  dans le graphe  $G$ . En effet, on a

$$P^k(i, j) = \sum_{l_1, \dots, l_{k-1}} P(i, l_1)P(l_1, l_2) \cdots P(l_{k-1}, j).$$

Chaque terme de cette somme est non nul lorsque  $j \rightarrow l_{k-1} \rightarrow \cdots \rightarrow l_1 \rightarrow i$  est un chemin dans le graphe  $G$ . Comme  $P \geq 0$ ,  $P^k(i, j)$  est non nul si et seulement si au moins un terme de la somme est non nul, ce qui permet de conclure. Pour que  $P$  soit régulière, il faut donc que  $G$  soit connexe, mais cela ne suffit pas. Par exemple, si l'on prend un graphe biparti, les chemins de longueur paire ont un départ et une arrivée dans la même « partie » du graphe, donc  $P^k$  n'aura jamais tous ses coefficients non nuls. D'ailleurs, comme indiqué dans l'énoncé (**S1**), si  $P$  représente une marche aléatoire sur un graphe biparti, où le choix du sommet au temps  $t + 1$  se fait uniformément parmi les voisins du sommet au temps  $t$ , alors  $-1$  est dans le spectre de  $P$ . Par contre, une matrice de transition  $P$  sur un graphe connexe tel que tout sommet est relié à lui même est régulière. En effet, comme le graphe est connexe, il existe un entier  $k$  tel que pour tout  $(i, j) \in V^2$ , il existe un chemin de  $j$  vers  $i$  de longueur au plus  $k$ . En ajoutant autant de fois que nécessaire l'arête  $(j, j)$  au début du chemin, on trouve un chemin de longueur exactement  $k$ . Justement, le fait de considérer des marches aléatoires paresseuses revient à ajouter toutes les arêtes  $(j, j)$ ,  $j \in V$  au graphe étudié, ce qui ne change bien entendu pas sa taille et permet de garantir que 1 est valeur propre dominante et simple.

# 1 Partie théorique

Les questions suivantes restent valables dans le cas plus général d'une marche aléatoire sur un graphe  $G$  non orienté connexe quelconque (pas forcément régulier) représentée par une matrice de transition (symétrique)  $P$  dont 1 est valeur propre simple et dominante.

**T1.** Soient  $s \in \{0, \dots, \tau - 1\}$  et  $i \in V$ . On a :

$$\begin{aligned} \mathbb{P}(X_{s+1}^t = i) &= \sum_{j \in V} \mathbb{P}((X_{s+1}^t = i) \cap (X_s^t = j)) \\ &= \sum_{j \in V} \frac{1}{d} \mathbb{1}_E(i, j) \mathbb{P}(X_s^t = j) \\ &= \sum_{j \in V} P_{ij} \mathbb{P}(X_s^t = j). \end{aligned}$$

Ainsi  $(\mathbb{P}(X_{s+1}^t = i))_{i \in V} = P \cdot (X_s^t = i)_{i \in V}$ . Une récurrence immédiate montre alors que

$$\pi = P^\tau \cdot (\delta_{i, i_0})_{i \in V} \quad (1)$$

où  $\delta$  désigne le symbole de Krönecker.

**T2.** Le théorème spectral assure l'existence d'une matrice orthogonale  $O$  telle que

$$P = {}^t O \text{Diag}(1, \lambda_2, \dots, \lambda_N) O.$$

Vu les valeurs propres de  $P$ , on voit que, pour une norme quelconque sur les matrices  $(N, N)$  (en particulier la norme subordonnée à  $\|\cdot\|_2$ ) :

$$P^\tau = {}^t O \Delta^\tau O \xrightarrow{\tau \rightarrow +\infty} P^\infty := {}^t O \text{Diag}(1, 0, \dots, 0) O$$

car la convergence a lieu coefficient par coefficient. La convergence de  $(P^\tau)_{\tau \geq 1}$  pour la norme subordonnée à  $\|\cdot\|_2$  entraîne la convergence simple de la suite d'applications linéaires associées au sens de  $\|\cdot\|_2$  vers la même limite. On peut conclure de deux manières :

— Via un calcul explicite :

$$\pi = P^\tau (\delta_{i, i_0})_{i \in V} \xrightarrow{\tau \rightarrow +\infty} P^\infty (\delta_{i, i_0})_{i \in V} = O_{1i_0} {}^t O {}^t (1, 0, \dots, 0) = O_{1i_0} O^{-1} {}^t (1, 0, \dots, 0).$$

Or  $O^{-1} {}^t (1, 0, \dots, 0)$  est le vecteur propre normalisé de  $P$  de valeur propre 1 (il y en a un seul car la valeur propre 1 était simple). On vérifie facilement qu'il s'agit de  $N^{-1/2} {}^t (1, \dots, 1)$ . Enfin, on a  $O_{1i_0} = N^{-1/2}$  car c'est un coefficient de la première ligne de  $O$ , donc de la première colonne de  ${}^t O = O^{-1}$ , qui n'est autre que la matrice d'une base orthonormée de diagonalisation de  $P$  dans la base canonique. On conclut comme attendu que

$$\lim_{\tau \rightarrow +\infty} \pi = N^{-1/2} {}^t (1, \dots, 1). \quad (2)$$

— En remarquant que  $\pi^\infty := \lim_{\tau \rightarrow +\infty} \pi$  est le vecteur propre de  $P$  associé à la valeur propre 1 et dont la somme des coefficients vaut 1. En effet, pour tout  $\tau$ , la somme des coefficients de  $\pi$  vaut 1, ce qui reste vrai à la limite (en particulier  $\pi^\infty \neq 0$ ), et on a :

$$P\pi^\infty = P \lim_{\tau \rightarrow +\infty} P^\tau (\delta_{i, i_0})_{i \in V} = \lim_{\tau \rightarrow +\infty} P^{\tau+1} (\delta_{i, i_0})_{i \in V} = \pi^\infty.$$

Encore une fois, on trouve la loi uniforme sur  $V$ .

**T3.** Notons

$$\mathcal{A}_m = \{(y_1, \dots, y_m) \in V^m \mid \text{Card}\{y_1, \dots, y_{m-1}\} = \text{Card}\{y_1, \dots, y_m\} = m - (\ell - 1)\},$$

de sorte que

$$(C_{\ell-1} = m) = \bigsqcup_{\mathbf{y} \in \mathcal{A}_m} (\mathbf{Y} = \mathbf{y}).$$

Notons aussi  $\mathcal{B}_n(U)$  l'ensemble des  $n$ -uplets injectifs à valeurs dans  $V \setminus U$ . On a :

$$\begin{aligned} \mathbb{P}((C_\ell - C_{\ell-1} > n) \cap (C_{\ell-1} = m)) &= \sum_{(y_1, \dots, y_n) \in \mathcal{A}_m} \mathbb{P}\left((C_\ell - C_{\ell-1} > n) \cap \bigcap_{1 \leq t \leq m} (Y_t = y_t)\right) \\ &= \sum_{\substack{(y_1, \dots, y_m) \in \mathcal{A}_m \\ (y_{m+1}, \dots, y_{m+n}) \in \mathcal{B}_n(\{y_1, \dots, y_m\})}} \mathbb{P}\left(\bigcap_{1 \leq t \leq m+n} (Y_t = y_t)\right) \\ &= \sum_{\substack{(y_1, \dots, y_m) \in \mathcal{A}_m \\ (y_{m+1}, \dots, y_{m+n}) \in \mathcal{B}_n(\{y_1, \dots, y_m\})}} \prod_{1 \leq t \leq m+n} \mathbb{P}(Y_t = y_t) \end{aligned} \quad (3)$$

$$\begin{aligned} &= \sum_{\substack{(y_1, \dots, y_m) \in \mathcal{A}_m \\ (y_{m+1}, \dots, y_{m+n}) \in \mathcal{B}_n(\{y_1, \dots, y_m\})}} \frac{1}{N^{m+n}} \\ &= \frac{1}{N^{m+n}} \sum_{(y_1, \dots, y_m) \in \mathcal{A}_m} \text{Card } \mathcal{B}_n(\{y_1, \dots, y_m\}) \\ &= \frac{1}{N^{m+n}} \sum_{(y_1, \dots, y_m) \in \mathcal{A}_m} n! \binom{N - (m - (\ell - 1))}{n} \\ &= \frac{(N - m + \ell - 1)(N - m + \ell - 2) \cdots (N - m + \ell - n)}{N^n} \frac{\text{Card } \mathcal{A}_m}{N^m} \\ &= \frac{(N - m + \ell - 1)(N - m + \ell - 2) \cdots (N - m + \ell - n)}{N^n} \mathbb{P}(C_{\ell-1} = m). \end{aligned} \quad (4)$$

On trouve comme attendu :

$$\mathbb{P}((C_\ell - C_{\ell-1} > n) \mid (C_{\ell-1} = m)) = \frac{(N - m + \ell - 1)(N - m + \ell - 2) \cdots (N - m + \ell - n)}{N^n}. \quad (5)$$

On aurait pu démontrer ce fait de manière moins formelle, les idées essentielles étant que

- les  $Y_t$  sont i.i.d. et suivent une loi uniforme sur  $V$  ;
- le cardinal de  $\mathcal{B}_n(U)$  ne dépend que de  $n$  et du cardinal de  $U$ , mais pas des valeurs de ses éléments.

Si l'on ne fait pas l'approximation de remplacer les variables  $Y_t$  par des variables uniformément distribuées sur  $V$ , il n'y a pas égalité entre les lignes (3) et (4). Cependant, la question **T2** montre que la ligne (4) est la limite de la ligne (3) lorsque  $\tau$  tend vers l'infini. Ainsi, pour être tout à fait précis, on a montré que

$$\lim_{\tau \rightarrow +\infty} \mathbb{P}((C_\ell - C_{\ell-1} > n) \mid (C_{\ell-1} = m)) = \frac{(N - m + \ell - 1)(N - m + \ell - 2) \cdots (N - m + \ell - n)}{N^n}.$$

L'approximation est donc raisonnable jusqu'ici.

**T4.** Nous allons légèrement améliorer le résultat de la première limite pour pouvoir en déduire la seconde. Soient  $a, b$  des réels strictement positifs et  $(a_N)_{N \geq 1}, (b_N)_{N \geq 1}$  des suites d'entiers telles que

$$a_N \underset{+\infty}{\sim} aN^{1/2} \quad \text{et} \quad b_N \underset{+\infty}{\sim} bN^{1/2}.$$

D'après la question précédente, on a :

$$\begin{aligned}\mathbb{P}((C_\ell - C_{\ell-1} > b_N) \mid (C_{\ell-1} = a_N)) &= \frac{(N - a_N + \ell - 1)(N - a_N + \ell - 2) \cdots (N - a_N + \ell - b_N)}{N^{b_N}} \\ &= \frac{(N - (a_N - (\ell - 1)))!}{N^{b_N}(N - b_N - (a_N - (\ell - 1)))!}\end{aligned}$$

Posons, pour  $N \geq 1$

$$\begin{aligned}u_N &= N - (a_N - (\ell - 1)) \\ v_N &= N - b_N - (a_N - (\ell - 1)).\end{aligned}$$

D'après la formule de Stirling, on a :

$$\begin{aligned}\mathbb{P}((C_\ell - C_{\ell-1} > b_N) \mid (C_{\ell-1} = a_N)) &\sim \frac{\sqrt{2\pi u_N}}{\sqrt{2\pi v_N}} \frac{\exp[u_N \log u_N - u_N]}{\exp[b_N \log N + v_N \log v_N - v_N]} \\ &\sim \exp[u_N \log u_N + v_N - u_N - b_N \log N - v_N \log v_N].\end{aligned}$$

On cherche un développement asymptotique en  $o(1)$  de l'argument de l'exponentielle. On procède par étapes :

$$\begin{aligned}\log u_N &= \log N - \frac{a_N - (\ell - 1)}{N} - \frac{(a_N - (\ell - 1))^2}{2N^2} + o(1/N) \\ &= \log N - \frac{a_N - (\ell - 1)}{N} - \frac{(aN^{1/2} + o(N^{1/2}))^2}{2N^2} + o(1/N) \\ &= \log N - \frac{a_N - (\ell - 1)}{N} - \frac{a^2}{2N} + o(1/N).\end{aligned}$$

De même,

$$\begin{aligned}\log v_N &= \log N - \frac{b_N + a_N - (\ell - 1)}{N} - \frac{(b_N + a_N - (\ell - 1))^2}{2N^2} + o(1/N) \\ &= \log N - \frac{b_N + a_N - (\ell - 1)}{N} - \frac{(a + b)^2}{2N} + o(1/N).\end{aligned}$$

Puis

$$\log u_N - \log v_N = \frac{b_N}{N} + \frac{b(2a + b)}{2N} + o(1/N).$$

D'où, en utilisant le fait que  $a_N = aN^{1/2} + o(N^{1/2})$  et  $b_N = bN^{1/2} + o(N^{1/2})$

$$\begin{aligned}v_N(\log u_N - \log v_N) &= b_N + \frac{b(2a + b)}{2} - \frac{b_N(b_N + a_N)}{N} + o(1) \\ &= b_N + \frac{b(2a + b)}{2} - b(a + b) + o(1) \\ &= b_N - \frac{b^2}{2} + o(1).\end{aligned}\tag{6}$$

De plus, pour la même raison,

$$b_N \log u_N = b_N \log N - ab + o(1).\tag{7}$$

Enfin, en utilisant les développements asymptotiques précédents (6 et 7)

$$\begin{aligned}
u_N \log u_N + v_N - u_N - b_N \log N - v_N \log v_N &= v_N (\log u_N - \log v_N) + b_N \log u_N - b_N - b_N \log N \\
&= b_N - \frac{b^2}{2} + b_N \log N - ab - b_N - b_N \log N + o(1) \\
&= -ab - \frac{b^2}{2} + o(1).
\end{aligned}$$

Cela montre que

$$\mathbb{P}((C_\ell - C_{\ell-1} > b_N) \mid (C_{\ell-1} = a_N)) \xrightarrow{N \rightarrow +\infty} e^{-ab-b^2/2}. \quad (8)$$

Étudions maintenant la suite

$$\mathbb{P} \left( \left( \frac{C_\ell^2 - C_{\ell-1}^2}{2N} > y \right) \middle| \left( \frac{C_{\ell-1}^2}{2N} = \frac{\lfloor (2Nx)^{1/2} \rfloor^2}{2N} \right) \right)$$

où  $x, y > 0$ . On écrit pour cela

$$\begin{aligned}
&\mathbb{P} \left( \left( \frac{C_\ell^2 - C_{\ell-1}^2}{2N} > y \right) \middle| \left( \frac{C_{\ell-1}^2}{2N} = \frac{\lfloor (2Nx)^{1/2} \rfloor^2}{2N} \right) \right) \\
&= \mathbb{P} \left( C_\ell > \left( \lfloor (2Nx)^{1/2} \rfloor^2 + (2Ny) \right)^{1/2} \middle| C_{\ell-1} = \lfloor (2Nx)^{1/2} \rfloor \right) \\
&= \mathbb{P} \left( C_\ell - C_{\ell-1} > \left( \lfloor (2Nx)^{1/2} \rfloor^2 + (2Ny) \right)^{1/2} - \lfloor (2Nx)^{1/2} \rfloor \middle| C_{\ell-1} = \lfloor (2Nx)^{1/2} \rfloor \right).
\end{aligned}$$

Or,

$$\lfloor (2Nx)^{1/2} \rfloor \underset{+\infty}{\sim} (2Nx)^{1/2}$$

et

$$\begin{aligned}
\left( \lfloor (2Nx)^{1/2} \rfloor^2 + (2Ny) \right)^{1/2} - \lfloor (2Nx)^{1/2} \rfloor &= (2N(x+y) + o(N))^{1/2} - (2Nx)^{1/2} + o(N^{1/2}) \\
&= (2N)^{1/2} \left[ (x+y)^{1/2} - x^{1/2} \right] + o(N^{1/2}).
\end{aligned}$$

On peut donc appliquer (8) avec  $a = (2x)^{1/2}$  et  $b = 2^{1/2} [(x+y)^{1/2} - x^{1/2}]$ . On a

$$ab + b^2/2 = 2 [x(x+y)]^{1/2} - 2x + (x+y) - 2 [x(x+y)]^{1/2} + x = y,$$

ce qui donne

$$\mathbb{P} \left( \left( \frac{C_\ell^2 - C_{\ell-1}^2}{2N} > y \right) \middle| \left( \frac{C_{\ell-1}^2}{2N} = \frac{\lfloor (2Nx)^{1/2} \rfloor^2}{2N} \right) \right) \xrightarrow{N \rightarrow +\infty} e^{-y}. \quad (9)$$