Estimation de la taille d'un graphe par marches aléatoires

Matthieu Dinot, Dorian Boully

Préliminaires

Montrons deux résultats permettant de légitimer les hypothèses faites dans **T2** et **T3**.

Spectre de la matrice d'adjacence d'un graphe d-régulier connexe

Une interversion de limites

1 Partie théorique

T1. Soient $s \in \{0, \ldots, \tau - 1\}$ 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 $\left(\mathbb{P}(X_{s+1}^t = i)\right)_{i \in V} = P \cdot (X_s^t = i)_{i \in V}$. Une récurrence immédiate montre alors que

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

où δ désigne le symbole de Krönecker.

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

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

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

$$P^\tau = {}^t O \Delta^\tau O \xrightarrow[\tau \to +\infty]{} P^\infty := {}^t O \, \mathrm{Diag}(1, 0, \ldots, 0) \, O$$

car la convergence a lieu coefficient par coefficient. La convergence de $(P^\tau)_{\tau \geqslant 1}$ pour la norme subordonnée à $\|\cdot\|_2$ entraine 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 :

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

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 ${}^tO = O^{-1}$, qui n'est autre que la matrice d'une base orthonormée de diagonalisation de P dans la base canonique. On conclut comme attendu que

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

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

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

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

T3. Notons

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

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 :

$$\begin{aligned}
\mathbb{P}((C_\ell - C_{\ell-1} > n)\cap(C_{\ell-1}=m)) &= \sum_{(y_1,\dots,y_n)\in\mathcal{A}_m} \mathbb{P}\left((C_\ell - C_{\ell-1} > n)\cap \bigcap_{1\leqslant t\leqslant m}(Y_t = y_t)\right)\\
&= \sum_{\substack{(y_1,\dots,y_m)\in\mathcal{A}_m\\ (y_{m+1},\dots,y_{m+n})\in\mathcal{B}_n(\{y_1,\dots,y_m\})}} \mathbb{P}\left(\bigcap_{1\leqslant t\leqslant m+n}(Y_t=y_t)\right)\\
&= \sum_{\substack{(y_1,\dots,y_m)\in\mathcal{A}_m\\ (y_{m+1},\dots,y_{m+n})\in\mathcal{B}_n(\{y_1,\dots,y_m\})}} \prod_{1\leqslant t\leqslant m+n}\mathbb{P}(Y_t=y_t)\\
&= \sum_{\substack{(y_1,\dots,y_m)\in\mathcal{A}_m\\ (y_{m+1},\dots,y_{m+n})\in\mathcal{B}_n(\{y_1,\dots,y_m\})}} \frac{1}{N^{m+n}}\\
&= \frac{1}{N^{m+n}} \sum_{(y_1,\dots,y_m)\in\mathcal{A}_m} \operatorname{Card}\mathcal{B}_n(\{y_1,\dots,y_m\})\\
&= \frac{1}{N^{m+n}} \sum_{(y_1,\dots,y_m)\in\mathcal{A}_m} n!\binom{N-(m-(\ell-1))}{n}\\
&= \frac{(N-m+\ell-1)(N-m+\ell-2)\cdots(N-m+\ell-n)}{N^n}\frac{\operatorname{Card}\mathcal{A}_m}{N^m}\\
&= \frac{(N-m+\ell-1)(N-m+\ell-2)\cdots(N-m+\ell-n)}{N^n}\mathbb{P}(C_{\ell-1}=m).
\end{aligned}$$

On trouve comme attendu :

$$\mathbb{P}((C_\ell - C_{\ell-1} > n) \mid (C_{\ell-1} = m)) = \frac{(N-m+\ell-1)(N-m+\ell-2)\cdots(N-m+\ell-n)}{N^n}. \tag{3}$$

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.

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 $(a_N)_{N\geqslant 1}, (b_N)_{N\geqslant 1}$ 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 $N \geqslant 1$

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

D'après la formule de Stirling, on a :

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

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 $b_N = bN^{1/2} + o(N^{1/2})$

$$\begin{aligned} v_n(\log u_n - \log v_n) &= b_N + \frac{b(2a+b)}{2} - \frac{b_N(b_N + a_N)}{N} + o(1) \\ &= b_N + \frac{b(2a+b)}{2} - b(a+b) + o(1) \\ &= b_N - \frac{b^2}{2} + o(1). \end{aligned} \tag{4}$$

De plus, pour la même raison,

$$b_N \log u_n = b_N \log N - ab + o(1). \tag{5}$$

Enfin, en utilisant les développements asymptotiques précédents (4 et 5)

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

Cela montre que

$$\mathbb{P}((C_\ell - C_{\ell-1} > b_N) \mid (C_{\ell-1} = a_N)) \xrightarrow[N\to+\infty]{} e^{-ab-b^2/2}. \tag{6}$$

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

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

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 (6) avec $a = (2x)^{1/2}$ et $b = 2^{1/2}\left[(x+y)^{1/2} - x^{1/2}\right]$. On a

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

ce qui donne

$$\mathbb{P}\left(\left(\frac{C_\ell^2 - C_{\ell-1}^2}{2N} > y\right) \middle| \left(\frac{C_{\ell-1}^2}{2N} = \frac{\lfloor (2Nx)^{1/2} \rfloor^2}{2N}\right)\right) \xrightarrow[N\to+\infty]{} e^{-y}. \tag{7}$$