Sur un graphe aléatoire

Erik Thomas [*]

Résumé

Dans cette note, nous nous intéressons à un modèle de graphe aléatoire. Nous montrons que le nombre de points isolés du graphe converge, sous certaines hypothèses, vers une loi de Poisson. Nous prouvons aussi une loi forte des grands nombres.

1 Introduction au problème

Le point de départ de cette note est un problème qui aurait pu intéresser Édouard Lucas lorsqu'il dénombrait le nombre de façons de placer n couples autour d'une table !

On place n personnes autour d'une table circulaire, chaque personne sympathise avec un de ses voisins (celui de gauche ou de droite avec une probabilité fixée indépendante du choix de la personne). On s'intéresse au nombre de personnes qui n'ont sympathisé avec personne à l'issue du repas.

On peut modéliser la situation avec un graphe aléatoire. Ce genre d'approche n'est pas nouveau et l'article fondateur est sûrement celui d'Erdös et Rényi. Dans leur article [1], les auteurs se placent dans le graphe complet K_n dans lequel les arêtes sont enlevées de manière indépendante avec une probabilité $p \in]0, 1[$. Les auteurs montrent alors que le nombre de sommets isolés (notion définie à la définition 1.1) que, sous certaines hypothèses entre n et p, le nombre de sommets isolés converge en loi vers une loi de Poisson.

Nous nous plaçons dans le cercle discret $[\![0, n-1]\!]$. L'addition est celle de $\mathbf{Z}/n\mathbf{Z}$. Chaque sommet i est relié par une arête aux sommets $i-1$ et $i+1$, que l'on appelera ses voisins. Soit $p \in]0, 1[$. Chaque arête est enlevée, indépendamment les unes des autres, avec une probabilité p. On note $G_{n,p}$ le graphe obtenu.

Définition 1.1. *Sommet isolé*

Soit $j \in [\![0, n-1]\!]$. On dit que le sommet j est un sommet isolé de $G_{n,p}$ s'il n'y a pas d'arête entre $j-1$ et j et entre j et $j+1$.

Pour tout $i \in [\![0, n-1]\!]$, on note $X_{i,n}$ la variable aléatoire qui vaut 1 si le sommet i est isolé et 0 sinon. On note X_n le nombre de sommets isolés de $G_{n,p}$. Il est clair que $X_n = \sum_{i=0}^{n-1} X_{i,n}$.

Le but de cette note est de montrer les deux propositions suivantes.

Proposition 1.1. *Probabilité de présence de points isolés*

On a :

(i) $\lim_{np^2 \to 0} \mathbf{P}(X_n > 0) = 0$;

(ii) $\lim_{np^2 \to +\infty} \mathbf{P}(X_n > 0) = 1$.

Remarque 1. On remarque que la valeur $\frac{1}{\sqrt{n}}$ constitue alors un seuil : le comportement de $\mathbf{P}(X_n > 0)$ dépend de la position de p par rapport à $\frac{1}{\sqrt{n}}$.

Proposition 1.2. *Convergence vers une loi de Poisson*

Soit $c > 0$. On pose $p := \frac{c}{\sqrt{n}} + o\left(\frac{1}{\sqrt{n}}\right)$. La suite $(X_n)_{n\in\mathbf{N}^\star}$ converge vers une loi de Poisson de paramètre c^2.

[*] `erik.thomas@ens-rennes.fr`

Nous allons ensuite nous intéresser à la convergence de la suite $\left(\frac{X_n}{n}\right)_{n\in\mathbf{N}^\star}$.

Proposition 1.3. *Loi des grands nombres*

Soit $p \in]0,1[$. *La suite* $\left(\frac{X_n}{n}\right)_{n\in\mathbf{N}^\star}$ *converge presque-sûrement vers une variable déterministe presque sûrement égale à* p^2.

2 Convergence du nombre de points isolés vers une loi de Poisson

Avant d'établir les propositions 1.1 et 1.2, nous allons établir la proposition suivante.

Proposition 2.1. *On a :*

(i) $\mathbf{E}(X) = np^2$;

(ii) *pour tout* $(i,j) \in [\![0, n-1]\!]^2$ *avec* $i \neq j$,

$$\operatorname{cov}(X_{i,n}, X_{j,n}) = \begin{cases} p^3(1-p) & \textit{si } i \textit{ et } j \textit{ sont voisins} \\ 0 & \textit{sinon} \end{cases};$$

(iii) $\mathbf{V}(X_n) = \mathbf{E}(X_n)(1+2p-3p^2)$.

Remarque 2. La proposition 1.2 devient alors intuitive : lorsque $p = \frac{c}{\sqrt{n}}$, la variable aléatoire X_n est la somme de lois de Bernoulli dont le couplage tend vers 0 lorsque n tend vers l'infini. Il est donc raisonnable de penser que X_n se comporte comme la somme de lois de Bernoulli indépendantes de paramètres $\frac{c^2}{n}$, soit une loi binomiale de paramètres n et $\frac{c^2}{n}$. Or, il est classique qu'une telle suite de variables aléatoires converge en loi vers une loi de Poisson de paramètre c^2.

Démonstration. (i) On a $\mathbf{E}(X_n) = \sum_{i=0}^{n-1} \mathbf{E}(X_{i,n})$. Or, $X_{i,n}$ suit une loi de Bernoulli et $X_{i,n} = 1$ si, et seulement si, les arêtes reliant les sommets $i-1$ et i et les sommets i et $i+1$ ont été enlevées. Par indépendance, on a $\mathbf{E}(X_{i,n}) = p^2$.

(ii) Soient i et j deux entiers compris entre 0 et $n-1$ avec $i \neq j$. On a

$$\operatorname{cov}(X_{i,n}, X_{j,n}) = \mathbf{E}(X_{i,n}X_{j,n}) - \mathbf{E}(X_{i,n})\,\mathbf{E}(X_{j,n}) = \mathbf{E}(X_{i,n}X_{j,n}) - p^4.$$

- Si i et j sont voisins. On a par exemple $j = i+1$. Ainsi, $X_{i,n}X_{j,n} = 1$ si et seulement si les arêtes reliant les sommets $i-1$, i, $j = i+1$ et $i+2$ ont été enlevées, soit $\mathbf{E}(X_{i,n}X_{j,n}) = p^3$.
- Si i et j ne sont pas voisins. On a $X_{i,n}X_{j,n} = 1$ si et seulement si, les arêtes reliant les sommets $i-1$ et i, les sommets i et $i+1$, les sommets $j-1$ et j et les sommets j et $j+1$ ont été enlevées. Ainsi, $\mathbf{E}(X_{i,n}X_{j,n}) = p^4$.

(iii) On a :

$$\begin{aligned}
\mathbf{V}(X_n) &= \mathbf{V}\left(\sum_{i=0}^{n-1} X_{i,n}\right) \\
&= \sum_{i=0}^{n-1} \mathbf{V}(X_{i,n}) + \sum_{i\neq j} \operatorname{cov}(X_{i,n}, X_{j,n}) \\
&= np^2(1-p^2) + 2np^3(1-p) \\
&= \mathbf{E}(X_n)(1+2p-3p^2).
\end{aligned}$$

□

On en déduit alors le corollaire suivant.

Corollaire 2.1. *Soient i et j deux éléments non voisins de $[\![1, n]\!]$. Les variables aléatoires $X_{i,n}$ et $X_{j,n}$ sont indépendantes.*

Remarque 3. On peut montrer plus généralement que si, $i_1, \ldots, i_m$ sont des voisins qui sont deux à deux non voisins, alors les variables aléatoires $X_{i_1}, \ldots, X_{i_m}$ sont indépendantes.

Démonstration. La preuve est immédiate si l'on se souvient que deux variables de Bernoulli sont indépendantes si, et seulement si, elles sont non corrélées. □

Nous prouvons la proposition 1.1.

Démonstration. (i) On utilise l'inégalité de Markov pour obtenir $\mathbf{P}(X_n > 0) = \mathbf{P}(X_n \geq 1) \leq \mathbf{E}(X_n) = np^2$, ainsi $\lim\limits_{np^2 \to 0} \mathbf{P}(X_n > 0) = 0$.

(ii) En utilisant l'inégalité de Bienaymé-Tchebyshev, on a :

$$\begin{aligned}
\mathbf{P}(X_n > 0) &\geq \frac{\mathbf{E}(X_n)^2}{\mathbf{E}(X_n^2)} \\
&\geq 1 - \frac{\mathbf{V}(X_n)}{\mathbf{E}(X_n^2)} \\
&\geq 1 - \frac{\mathbf{V}(X_n)}{\mathbf{E}(X_n)^2} \\
&\geq 1 - \frac{1}{\mathbf{E}(X_n)(1 + o(1))}.
\end{aligned}$$

Il s'ensuit que $\lim\limits_{np^2 \to +\infty} \mathbf{P}(X_n > 0) = 1$.

□

Nous prouvons la proposition 1.2.

Démonstration. • Soit $k \in \mathbf{N}^\star$. On calcule le k-ième moment factoriel de X_n : $\mathbf{E}(X_n(X_n - 1) \cdots (X_n - (k-1)))$. On a :

$$\begin{aligned}
&\mathbf{E}(X_n(X_n - 1) \cdots (X_n - (k-1))) \\
= &\sum_{\substack{i_1, \ldots, i_k \\ \text{deux à deux distincts}}} \mathbf{P}(i_1, \ldots, i_k \text{ isolés}) \\
= &\sum_{\substack{i_1, \ldots, i_k \text{ deux à deux} \\ \text{distincts et non voisins}}} \mathbf{P}(i_1, \ldots, i_k \text{ isolés}) + \sum_{\substack{i_1, \ldots, i_k \text{ deux à deux distincts} \\ \text{au moins deux sommets sont voisins}}} \mathbf{P}(i_1, \ldots, i_k \text{ isolés}).
\end{aligned}$$

La somme $\sum\limits_{\substack{i_1, \ldots, i_k \text{ deux à deux} \\ \text{distincts et non voisins}}}$ contient $n(n-2)(n-4) \cdots (n - 2(k-1))$ termes, ainsi

$$\sum_{\substack{i_1, \ldots, i_k \text{ deux à deux} \\ \text{distincts et non voisins}}} \mathbf{P}(i_1, \ldots, i_k \text{ isolés}) = n(n-2)(n-4) \cdots (n - 2(k-1))p^{2k} \underset{n \to +\infty}{\sim} \mathbf{E}(X_n)^k = c^{2k}.$$

On notera que la ligne précédente assure qu'il existe $\lambda_1 > 0$ tel que :

$$0 \leq \sum_{\substack{i_1, \ldots, i_k \text{ deux à deux} \\ \text{distincts et non voisins}}} \mathbf{P}(i_1, \ldots, i_k \text{ isolés}) \leq \lambda_1 n^k p^{2k} = \lambda_1 c^{2k}. \tag{1}$$

Il reste à montrer que $\lim\limits_{n \to +\infty} \sum\limits_{\substack{i_1, \ldots, i_k \text{ deux à deux distincts} \\ \text{au moins deux sommets sont voisins}}} \mathbf{P}(i_1, \ldots, i_k \text{ isolés}) = 0$. Pour cela, on discute suivant le nombre de voisins. Pour $j \in [\![2, k]\!]$, on introduit

$$S_j := \sum_{\substack{i_1, \ldots, i_k \text{ deux à deux distincts} \\ j \text{ sommets sont voisins}}} \mathbf{P}(i_1, \ldots, i_k \text{ isolés})$$

de sorte que

$$\sum_{\substack{i_1,\dots,i_k \text{ deux à deux distincts} \\ \text{au moins deux sommets sont voisins}}} \mathbf{P}(i_1,\dots,i_k \text{ isolés}) = \sum_{j=2}^{k} S_j.$$

Or, la somme de S_j contient $n \times (n-j-2) \times (n-j-4) \times \cdots \times (n-j-2(k-j))$ termes. Il s'ensuit que

$$\begin{aligned} S_j &= n \times (n-j-2) \times (n-j-4) \times \cdots \times (n-j-2(k-j)) \times p^j \times p^{2(k-j)} \\ &\underset{n\to+\infty}{\sim} n^{k-j+1} p^{2k-j+1} \\ &\underset{n\to+\infty}{\sim} \mathbf{E}(X_n)^{k-j+1} p^{j-1}. \end{aligned}$$

Le calcul ci-dessus assure aussi que :

$$\forall j \in [\![2,k]\!], \exists \lambda_j \geq 0, \quad 0 \leq S_j \leq \lambda_j c^{2(k-j+1)} p^{j-1} \leq \lambda_j c^{2(k-j+1)}. \tag{2}$$

On en déduit que $\lim_{n\to+\infty} S_j = 0$, puis $\lim_{n\to+\infty} \sum_{\substack{i_1,\dots,i_k \text{ deux à deux distincts} \\ \text{au moins deux sommets sont voisins}}} \mathbf{P}(i_1,\dots,i_k \text{ isolés}) = 0$. On a donc montré que

$$\forall k \in \mathbf{N}^\star, \quad \lim_{n\to+\infty} \mathbf{E}(X_n(X_n-1)\cdots(X_n-(k-1))) = c^{2k}.$$

Les lignes (1) et (2) assurent, en posant $C := \max\{\lambda_1,\dots,\lambda_k\}$, aussi que :

$$0 \leq \mathbf{E}(X_n(X_n-1)\cdots(X_n-(k-1)) \leq Cc^{2k} + C\sum_{j=2}^{k} c^{2(k-j+1)} = C\sum_{j=1}^{k} c^{2(k-j+1)}. \tag{3}$$

- On introduit la fonction génératrice factorielle des moments de X_n :

$$\varphi_n(t) := \sum_{k=0}^{+\infty} \frac{\mathbf{E}(X_n^{(k)})}{k!} t^k,$$

où, par commodité, on a posé $X_n^{(k)} := X_n(X_n-1)\cdots(X_n-(k-1))$. L'inégalité (3) assure que φ_n est bien défini sur $\mathbf{R}$.
Soit $Y \hookrightarrow \mathscr{P}(c^2)$. On note φ sa fonction génératrice factorielle définie par

$$\varphi(t) := \sum_{k=0}^{+\infty} \frac{\mathbf{E}(Y^{(k)})}{k!} t^k = \sum_{k=0}^{+\infty} \frac{c^{2k} t^k}{k!} = \exp(c^2 t).$$

Il s'ensuit que φ est définie sur $\mathbf{R}$.
Soit $t \in \mathbf{R}$. Comme pour tout $k \in \mathbf{N}$, $\lim_{n\to+\infty} \mathbf{E}(X_n^{(k)}) = \mathbf{E}(Y^{(k)})$ et grâce à la majoration (3), le théorème de convergence dominée pour les séries assure que

$$\lim_{n\to+\infty} \varphi_n(t) = \varphi(t). \tag{4}$$

- La ligne (4) assure que la suite $(X_n)_{n\in\mathbf{N}^\star}$ converge en loi vers Y.

$\square$

3 Loi des grands nombres

Nous allons prouver la proposition 1.3.

Démonstration. • Nous prouvons d'abord la convergence en loi. Soit $\varepsilon > 0$. D'après l'inégalité de Bienaymé-Tchebyshev et en utilisant la proposition 2.1, on a :

$$\forall n \in \mathbf{N}^\star, \quad \mathbf{P}\left(\left|\frac{X_n}{n} - p^2\right| \geq \varepsilon\right) \leq \frac{\mathbf{V}(X_n)}{n^2\varepsilon^2} = \frac{p^2(1+2p-3p^2)}{n\varepsilon^2}.$$

Par encadrement, $\lim\limits_{n\to+\infty} \mathbf{P}\left(\left|\frac{X_n}{n} - p^2\right| \geq \varepsilon\right) = 0$. Ainsi, la suite $\left(\frac{X_n}{n}\right)_{n\in\mathbf{N}^\star}$ converge en probabilité, donc en loi, vers une variable aléatoire déterministe presque sûrement égale à p^2.

• Nous prouvons d'abord la convergence en loi. Soit $n \in \mathbf{N}^\star$. Pour tout $i \in [\![0, n-1]\!]$, on pose $X'_{i,n} = X_{i,n} - \mathbf{E}(X_{i,n}) = X_{i,n} - p^2$. On pose aussi $X'_n = \sum_{i=0}^{n-1} X'_{i,n}$. Il suffit donc montrer que $\left(\frac{X'_n}{n}\right)_{n\in\mathbf{N}^\star}$ converge presque-sûrement vers 0. On va montrer que $\sum_{n\geq 1} \mathbf{E}\left(\frac{X'^4_n}{n^4}\right)$ converge presque-sûrement, ce qui assure $\left(\frac{X'_n}{n}\right)_{n\in\mathbf{N}^\star}$ converge presque-sûrement vers 0. On a :

$$\mathbf{E}(X'^4_n) = \sum_{1\leq i,j,k,\ell\leq n} \mathbf{E}(X'_{i,n}X'_{j,n}X'_{k,n}X'_{\ell,n}) = S_1 + S_2 + S_3 + S_4,$$

où, pour tout $m \in [\![1, 4]\!]$, on a posé

$$S_m := \sum_{\substack{1\leq i,j,k,\ell\leq n \\ \operatorname{card}\{i,j,k,\ell\}=m}} \mathbf{E}(X'_{i,n}X'_{j,n}X'_{k,n}X'_{\ell,n}).$$

On a

$$S_1 = \sum_{i=1}^{n} \mathbf{E}(X'^4_{i,n}) \underset{n\to+\infty}{=} O(n). \tag{5}$$

Il y a $\binom{4}{2}$ façons de choisir 2 éléments parmi 4 (ceux qui seront égaux), ainsi

$$S_2 = \binom{4}{2} \sum_{\substack{1\leq i,j\leq n \\ i\neq j}} \mathbf{E}(X'^2_{i,n}, X'^2_{j,n}) \underset{n\to+\infty}{=} O(n^2). \tag{6}$$

Il y a $\binom{4}{3}$ façons de choisir 3 éléments parmi 4 (ceux qui seront égaux), ainsi

$$S_3 = \binom{4}{3} \sum_{\substack{1\leq i,j\leq n \\ i\neq j}} \mathbf{E}(X'^3_{i,n}, X'_{j,n}) \underset{n\to+\infty}{=} O(n^2). \tag{7}$$

Pour étudier S_4, on discute suivant le nombre de voisins. Plus formellement, on introduit :

$$S_4^{(1)} := \sum_{\substack{1\leq i,j,k,\ell\leq n \\ \text{tous non voisins}}} \mathbf{E}(X'_{i,n}X'_{j,n}X'_{k,n}X'_{\ell,n}),$$

$$S_4^{(2)} := \sum_{\substack{1\leq i,j,k,\ell\leq n \\ \text{trois voisins, l'autre isolé}}} \mathbf{E}(X'_{i,n}X'_{j,n}X'_{k,n}X'_{\ell,n}),$$

$$S_4^{(3)} := \sum_{\substack{1\leq i,j,k,\ell\leq n \\ \text{deux groupes de voisins}}} \mathbf{E}(X'_{i,n}X'_{j,n}X'_{k,n}X'_{\ell,n}),$$

$$S_4^{(4)} := \sum_{\substack{1\leq i,j,k,\ell\leq n \\ \text{deux voisins, les autres deux isolés}}} \mathbf{E}(X'_{i,n}X'_{j,n}X'_{k,n}X'_{\ell,n})$$

et

$$S_4^{(5)} := \sum_{\substack{1\leq i,j,k,\ell\leq n \\ \text{tous voisins}}} \mathbf{E}(X'_{i,n}X'_{j,n}X'_{k,n}X'_{\ell,n}).$$

On a donc $S_4 = \sum_{m=1}^{5} S_4^{(m)}$. Par indépendance des variables aléatoires car les sommets sont tous non voisins, on a

$$S_4^{(1)} = \sum_{\substack{1\leq i,j,k,\ell\leq n \\ \text{tous non voisins}}} \mathbf{E}(X'_{i,n})\,\mathbf{E}(X'_{j,n})\,\mathbf{E}(X'_{k,n})\,\mathbf{E}(X'_{\ell,n}) = 0. \tag{8}$$

Pour estimer $S_4^{(2)}$, on remarque que la somme contient contient $O(n^2)$ termes (on choix un sommet, puis on place les deux sommets voisins et un autre sommet non voisin), donc

$$S_4^{(2)} \underset{n\to+\infty}{=} O(n^2). \tag{9}$$

Le même raisonnement s'applique pour montrer que

$$S_4^{(3)} \underset{n\to+\infty}{=} O(n^2). \tag{10}$$

On remarque, par exemple, que si les i et j sont voisins et si les sommets k et ℓ sont isolés, alors, par indépendance $\mathbf{E}(X'_{i,n}X'_{j,n}X'_{k,n}X'_{\ell,n}) = \mathbf{E}(X'_{i,n}X'_{j,n})\,\mathbf{E}(X'_{k,n})\,\mathbf{E}(X'_{\ell,n}) = 0$, ainsi

$$S_4^{(4)} = 0. \tag{11}$$

Enfin, on remarque que

$$S_4^{(5)} \underset{n\to+\infty}{=} O(n) \tag{12}$$

car la somme $\sum_{\substack{1\leq i,j,k,\ell\leq n \\ \text{tous voisins}}}$ contient $O(n)$ termes (il suffit de choisir un sommet dans $[\![0, n-1]\!]$, les autres étant voisins).

Les lignes (5), (6), (7), (8), (9), (10), (11) et (12) assurent que $\mathbf{E}(X'_n) \underset{n\to+\infty}{=} O(n^2)$, soit $\mathbf{E}\left(\frac{X'_n 4}{n^4}\right) \underset{n\to+\infty}{=} O\left(\frac{1}{n^2}\right)$, ce qui permet de conclure.

$\square$

Références

[1] P. Erdös, A. Rényi, *On Random Graphs. I*, Publicationes Mathematicae. 6, pp. 290-297, 1959.