

TD 07 – Graphes aléatoires (corrigé)

Exercice 1.*Grphe Aléatoire Bipartite*

Soit $0 < p < 1$ et $n \in \mathbb{N}^*$. On définit un graphe aléatoire non orienté $H_{2n,p}$ de la manière suivante. On se donne une famille $\{X_{i,j} : 1 \leq i \leq n, n+1 \leq j \leq 2n\}$ de v.a. i.i.d. de loi de Bernoulli de paramètre p . On pose alors $H_{2n,p} = (V, E)$, avec $V = \{1, \dots, 2n\}$ et

$$E = \{(i, j) : X_{i,j} = 1\} \subset \{1, \dots, n\} \times \{n+1, \dots, 2n\}.$$

1. Quelle est la loi du nombre d'arêtes de $H_{2n,p}$?

☞ Le nombre d'arêtes de $H_{2n,p}$ suit la loi $B(n^2, p)$.

2. Quelle est l'espérance du nombre de sommets isolés de $H_{2n,p}$?

☞ Soit N le nombre de sommets isolés. Si A_i est l'événement « le sommet i est isolé », on a par linéarité de l'espérance, on a $E[N] = \sum P(A_i) = 2n(1-p)^n$.

3. Dans cette question on pose $p = c \log(n)/n$ pour un nombre réel $c > 0$.

1. Montrer que si $c > 1$, alors

$$\lim_{n \rightarrow \infty} P(H_{2n,p} \text{ a un sommet isolé}) = 0.$$

2. Montrer que si $c < 1$, alors

$$\lim_{n \rightarrow \infty} P(H_{2n,p} \text{ a un sommet isolé}) = 1.$$

☞

1. Si $c > 1$, on a $E[N] = 2n \exp(n \log(1 - \frac{c \log n}{n})) \rightarrow 0$ et donc $P(N \geq 1) \leq E[N] \rightarrow 0$.
2. Si $c < 1$, on calcule

$$E[N^2] = \sum_{i,j=1}^{2n} P(A_i \cap A_j) = 2n(1-p)^n + 2n(n-1)(1-p)^{2n} + 2n^2(1-p)^{2n-1}$$

d'où il vient que $E[N^2]/E[N]^2$ tend vers 1. On utilise l'inégalité de Tchebychev pour conclure que

$$P(N=0) = P(E[N] - N \geq E[N]) \leq P(|N - E[N]| \geq E[N]) \leq \frac{\text{Var}[N]}{E[N]^2} = \frac{E[N^2]}{E[N]^2} - 1 \rightarrow 0.$$

4. Dans cette question on pose $p = 1/2$. Montrer qu'il existe une constante $C > 0$ telle que

$$\lim_{n \rightarrow \infty} P\left(\text{tous les sommets de } H_{2n,p} \text{ ont un degré inférieur à } \frac{n}{2} + C\sqrt{n \log n}\right) = 1.$$

☞ Le degré d_i du sommet i suit la loi $B(n, 1/2)$. Par l'inégalité de Chernoff I, on a donc

$$P(d_i \geq \frac{n}{2} + a) \leq \exp(-2a^2/n).$$

Ainsi, par la borne de l'union,

$$P(\max_i d_i \geq \frac{n}{2} + a) \leq 2n \exp(-2a^2/n).$$

Cette quantité tend vers 0 si $a = C\sqrt{n \log n}$ avec $2C^2 > 1$.

Exercice 2.*K4*

Soit G un graphe aléatoire de loi $G_{n,p}$. L'objectif de cet exercice est de montrer qu'il y a un seuil $p_0 := n^{-2/3}$ tel que pour $p = o(p_0)$, le graphe G n'a pas de clique de taille 4 avec bonne probabilité, et que pour $p = \omega(p_0)$, le graphe G a au moins une clique de taille 4 avec bonne probabilité.

Rappels / définitions :

- Un graphe aléatoire G suit la loi $G_{n,p}$ s'il a n sommets et que chaque arête est présente dans G avec probabilité p ;
- une clique de taille 4 est un ensemble de 4 sommets tous reliés deux à deux par des arêtes;
- $p = o(p_0)$ signifie $\frac{p}{p_0} \rightarrow 0$ quand $n \rightarrow +\infty$;
- $p = \omega(p_0)$ signifie $\frac{p_0}{p} \rightarrow 0$ quand $n \rightarrow +\infty$.

1. Pour p quelconque, calculer $\mathbf{E}[X]$, où X est le nombre de cliques du graphe G .

☞ On a $\binom{n}{4}$ ensembles possibles de 4 sommets. Pour chacun de ces ensembles, on définit X_i qui vaut 1 si c'est une clique et zéro sinon. On a $X = \sum_i X_i$. Et $\mathbf{E}[X_i] = \mathbf{P}\{X_i = 1\} = p^6$. D'où

$$\mathbf{E}[X] = \binom{n}{4} p^6.$$

2. Soit $p = o(p_0)$, montrer que $\Pr(X \neq 0) \rightarrow 0$ quand n tend vers l'infini.

☞ On a $\mathbf{E}[X] \leq n^4 p^6 = o(n^{4-6 \times 2/3}) = o(1)$. Donc $\mathbf{E}[X]$ tend vers zéro quand n tend vers l'infini. Comme X est une valeur entière, positive ou nulle, on conclut que $\mathbf{P}\{X \neq 0\} = \Pr(X \geq 1) \leq \mathbf{E}[X]$ tend aussi vers 0.

On suppose maintenant $p = \omega(p_0)$, et on veut montrer que $\mathbf{P}\{X = 0\} \rightarrow 0$ quand n tend vers l'infini.

3. Montrer que $\mathbf{P}\{X = 0\} \leq \frac{\mathbf{Var}[X]}{\mathbf{E}[X]^2}$. Il suffira donc de montrer que $\frac{\mathbf{Var}[X]}{\mathbf{E}[X]^2} \rightarrow 0$.

☞ On utilise l'inégalité de Chebychev

$$\Pr(X = 0) \leq \mathbf{P}\{|X - \mathbf{E}[X]| \geq \mathbf{E}[X]\} \leq \frac{\mathbf{Var}[X]}{\mathbf{E}[X]^2}.$$

4. Soit X_i des variables aléatoires à valeur dans $0, 1$ (et non indépendantes). Montrer que

$$\mathbf{Var}\left[\sum_i X_i\right] \leq \mathbf{E}\left[\sum_i X_i\right] + \sum_{i \neq j} \mathbf{E}[(X_i - \mathbf{E}[X_i])(X_j - \mathbf{E}[X_j])].$$

☞ On a

$$\begin{aligned} \mathbf{Var}\left[\sum_i X_i\right] &= \mathbf{E}\left[\left(\sum_i (X_i - \mathbf{E}[X_i])\right)^2\right] \\ &= \mathbf{E}\left[\sum_{i,j} (X_i - \mathbf{E}[X_i]) \cdot (X_j - \mathbf{E}[X_j])\right] \\ &= \sum_i \mathbf{E}[(X_i - \mathbf{E}[X_i])^2] + \sum_{i \neq j} \mathbf{E}[(X_i - \mathbf{E}[X_i])(X_j - \mathbf{E}[X_j])]. \end{aligned}$$

On observe ensuite que $\mathbf{E}[(X_i - \mathbf{E}[X_i])^2] = \mathbf{E}[X_i^2] - \mathbf{E}[X_i]^2 \leq \mathbf{E}[X_i^2]$. Mais comme X_i est à valeur dans $0, 1$, on a $\mathbf{E}[X_i^2] = \mathbf{E}[X_i]$. D'où l'inégalité.

5. En déduire que $\mathbf{Var}[X] = o(\mathbf{E}[X]^2)$ et conclure.

☞ On note C_i les ensembles de 4 sommets du graphe G , et on définit X_i la variable aléatoire qui vaut 1 si C_i forme une clique et 0 sinon. On a $X = \sum_i X_i$ et d'après la question précédente $\mathbf{Var}[X] \leq \mathbf{E}[X] + \sum_{i \neq j} \mathbf{E}[(X_i - \mathbf{E}[X_i])(X_j - \mathbf{E}[X_j])]$. Fixons $i \neq j$ et considérons $\mathbf{E}[(X_i - \mathbf{E}[X_i])(X_j - \mathbf{E}[X_j])] = \mathbf{E}[X_i X_j] - \mathbf{E}[X_i] \mathbf{E}[X_j]$. On a $\mathbf{E}[X_i X_j] = \mathbf{P}\{C_i \text{ et } C_j \text{ sont des cliques}\} = p^k$, où k est le nombre d'arêtes nécessaires pour que C_i et C_j soient des cliques. Ce nombre d'arêtes va dépendre du nombre de sommets communs entre C_i et C_j . On distingue donc les cas suivants

- Si $|C_i \cap C_j| = 0$ ou $|C_i \cap C_j| = 1$, alors $k = 12$ et $\mathbf{E}[(X_i - \mathbf{E}[X_i])(X_j - \mathbf{E}[X_j])] = 0$.
- Si $|C_i \cap C_j| = 2$, alors $k = 11$ et $\mathbf{E}[(X_i - \mathbf{E}[X_i])(X_j - \mathbf{E}[X_j])] = p^{11}(1-p)$. Il y a $\binom{n}{4} \cdot \binom{n-4}{2} \cdot \binom{n-4}{2}$ tels couples (C_i, C_j) .
- Si $|C_i \cap C_j| = 3$, alors $k = 9$ et $\mathbf{E}[(X_i - \mathbf{E}[X_i])(X_j - \mathbf{E}[X_j])] = p^9(1-p^3)$. Il y a $\binom{n}{4} \cdot \binom{n-4}{3} \cdot \binom{n-4}{1}$ tels couples (C_i, C_j) .
- Le cas $|C_i \cap C_j| = 4$ est impossible car $C_i \neq C_j$.

On a donc $\mathbf{Var}[X] \leq \mathbf{E}[X] + \binom{n}{4} \cdot \binom{n-4}{2} \cdot \binom{n-4}{2} p^{11}(1-p) + \binom{n}{4} \cdot \binom{n-4}{3} \cdot \binom{n-4}{1} p^9(1-p^3)$. Chacun des trois termes de cette somme est un $o(\mathbf{E}[X]^2)$ (car $p = \omega(p_0)$), d'où la réponse à la question. On conclut grâce aux questions précédentes que $\Pr(X \neq 0) \rightarrow 0$ quand n tend vers l'infini.

Exercice 3.

Clique Number

Le clique number $\omega(G)$ d'un graphe G est l'entier k maximal tel que G contient une clique de taille k (c'est-à-dire un ensemble de k sommets tous reliés deux à deux par des arêtes). Soit G un graphe aléatoire de loi $G_{n,1/2}$ (i.e. G a n sommets et chaque arête de G est présente avec probabilité $1/2$). L'objectif de cet exercice est de montrer que pour tout $\varepsilon > 0$,

$$\lim_{n \rightarrow +\infty} \mathbf{P}\{(2 - \varepsilon) \log n \leq \omega(G) \leq (2 + \varepsilon) \log n\} = 1.$$

Asymptotiquement, $\omega(G)$ est donc de l'ordre de $2 \log n$.

Remarque : La méthode utilisée pour montrer ce résultat ressemble beaucoup à celle utilisée dans l'exercice K4.

1. Pour un entier k quelconque, on définit X_k la variable aléatoire comptant le nombre de cliques de taille k dans G . Calculer $\mathbf{E}[X_k]$

☞ Une solution pour l'exercice peut être trouvée ici : https://www.win.tue.nl/~nikhil/courses/2015/2W008/probabilistic_method-2.pdf, section 3.2.

2. Soit $\varepsilon > 0$. Montrer que $\lim_{n \rightarrow +\infty} \mathbf{P}\{\omega(G) \geq (2 + \varepsilon) \log n\} = 0$ (pour simplifier, on pourra supposer que $k = (2 + \varepsilon) \log n$ est un entier).
3. On considère maintenant $k = (2 - \varepsilon) \log n$ (encore une fois, on le suppose entier). Montrer que $\lim_{n \rightarrow +\infty} \mathbf{P}\{\omega(G) \leq (2 - \varepsilon) \log n\} = 0$. *Indice : on pourra utiliser les questions 3 et 4 de l'exercice K4.*

Exercice 4.

Approximation de Poisson

On se place dans le modèle *Balls and Bins* où l'on jette m balles au hasard dans n paniers. Le problème est que les v.a. X_i représentant le nombre de balles dans le i -ème panier ne sont pas indépendantes (intuitivement, car $X_1 + \dots + X_n = m$). On voudrait approximer le modèle *Balls and Bins* par le modèle *Approximation de Poisson*, dans lequel Y_1, \dots, Y_n sont des variables aléatoires indépendantes qui suivent chacune une loi de Poisson de moyenne $\mu = m/n$ (la variable Y_i est donc pensée pour être la version simplifiée de X_i).

1. Montrer que $Y = \sum_{i=1}^n Y_i$ suit une loi de Poisson dont on précisera le paramètre.

☞ Facile par récurrence sur n : la somme de v.a. de Poisson indépendantes Y_i de paramètre μ_i est une v.a. de Poisson de param $\sum_i \mu_i$. Voir par exemple Lemma 5.2 dans MU, p. 96. ($\mathbf{P}\{Y_i = j\} = e^{-\mu} \mu^j / j!$)

2. Montrer que la distribution de (Y_1, \dots, Y_n) conditionnée au fait que $Y = m$ est la même que la distribution de (X_1, \dots, X_n) .

Note : on peut en fait obtenir un résultat légèrement plus général. Si (X_1, \dots, X_n) représente la charge de n paniers après avoir lancé au hasard k balles, et que les Y_i sont n v.a. indépendantes suivant chacune une loi de Poisson de paramètre m/n , alors la distribution de (Y_1, \dots, Y_n) conditionnée au fait que $Y = k$ est la même que la distribution de (X_1, \dots, X_n) , indépendamment de la valeur de m .

☞ voir Thm 5.6 dans MU 5.4, p. 100

Fixing $k_1 \dots k_n$ summing to k , the probability that all $X_i = k_i$ is

$$\frac{k!}{k_1! \dots k_n!} \frac{1}{n^k}$$

Now, as Y_i are independent and follow a Poisson law, the probability that all $Y_i = k_i$ knowing that their sum $Y = k$ is

$$\frac{\prod_i e^{-m/n} \mu^{k_i} / k_i!}{e^{-m} m^k / k!} = \frac{k!}{k_1! \dots k_n!} \frac{1}{n^k}$$

3. Soit f une fonction sur n variables, à valeurs réelles positives ou nulles. Prouver que

$$\mathbf{E}[f(X_1, \dots, X_n)] \leq e\sqrt{m} \mathbf{E}[f(Y_1, \dots, Y_n)] .$$

On pourra prouver comme étape intermédiaire que $m! < e\sqrt{m} \left(\frac{m}{e}\right)^m$.

☞ voir Thm 5.7 dans MU 5.4, p. 101

Inequality : Since $\ln x$ is concave we have

$$\int_{i-1}^i \ln x dx \geq \frac{1}{2} (\ln(i-1) + \ln i)$$

Then

$$\int_1^n \ln x = n \ln n - 1 \geq \ln(n!) - \frac{1}{2} \ln n$$

Then exponentiate.

Then,

$$\mathbf{E}[f(Y_1, \dots, Y_n)] \geq \mathbf{E}[f(X_1, \dots, X_n)] \mathbf{P}\{Y = m\}$$

And use the inequality

4. En déduire le corollaire suivant : soit \mathcal{E} un événement qui dépend de la charge des paniers. Supposons que \mathcal{E} arrive avec probabilité p dans l'Approximation de Poisson, c'est-à-dire si la charge des paniers est (Y_1, \dots, Y_n) . Alors \mathcal{E} arrive avec probabilité au plus $pe\sqrt{m}$ dans le modèle *Balls and Bins*, c'est-à-dire si la charge des paniers est (X_1, \dots, X_n) .

☞ Voir corollaire 5.9, MU p. 102

5. *Application* : On jette n balles dans n paniers selon le modèle *Balls and Bins*. Montrer qu'avec probabilité au moins $1 - 1/n$ (pour n assez grand), la charge maximale est $\geq \ln n / \ln \ln n$.

☞ Lemma 5.12 MU p. 103

Poisson with $m = n$: probability that no bin exceed charge M is $(1 - \frac{1}{eM})^n \leq e^{-n/(eM)}$