TD 06 – Graphes aléatoires et partiel de l'an dernier

Exercice 1. *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 arrête 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 $\frac{p}{p_0} \to 0$ quand $n \to +\infty$;

— $p = \omega(p_0)$ signifie $\frac{p_0}{p} \to 0$ quand $n \to +\infty$.

1. Pour p quelconque, calculer $\mathbf{E}[X]$, où X est le nombre de cliques du graphe G.
2. Soit $p = o(p_0)$, montrer que $\Pr(X \neq 0) \to 0$ quand n tend vers l'infini.

On suppose maintenant $p = \omega(p_0)$, et on veut montrer que $\mathbf{P}\{X = 0\} \to 0$ 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 $\frac{\mathbf{Var}[X]}{\mathbf{E}[X]^2} \to 0$.
4. Soit X_i des variables aléatoires à valeur dans $0, 1$ (et non indépendantes). Montrer que

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

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

Exercice 2. *Graphe 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 $V = \{1, \ldots, 2n\}$ et

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

1. Quelle est la loi du nombre d'arêtes de $H_{2n,p}$?
2. Quelle est l'espérance du nombre de sommets isolés de $H_{2n,p}$?
3. Dans cette question on pose $p = c \log(n)/n$ pour un nombre réel $c > 0$.
   1. Montrer que si $c > 1$, alors
   $$\lim_{n \to \infty} \mathbb{P}(H_{2n,p} \text{ a un sommet isolé}) = 0.$$
   2. Montrer que si $c < 1$, alors
   $$\lim_{n \to \infty} \mathbb{P}(H_{2n,p} \text{ a un sommet isolé}) = 1.$$
4. Dans cette question on pose $p = 1/2$. Montrer qu'il existe une constante $C > 0$ telle que

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

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 $\varepsilon > 0$,

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

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 $\mathbf{E}\left[X_k\right]$
2. Soit $\varepsilon > 0$. Montrer que $\lim_{n\to+\infty} \mathbf{P}\left\{\omega(G) \geq (2+\varepsilon)\log n\right\} = 0$ (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 $\lim_{n\to+\infty} \mathbf{P}\left\{\omega(G) \leq (2-\varepsilon)\log n\right\} = 0$. *Indice : on pourra utiliser les questions 3 et 4 de l'exercice K4.*

Exercice 4. *Estimer l'intersection avec un rectangle*

Let $P \subset \mathbb{Z}^2$ of size n. Our objective is to be able to answer quickly queries of the form what is the fraction of points in P that are in the rectangle $r = [a_1, b_1] \times [a_2, b_2]$? We write $r[P] = \frac{|P\cap r|}{n}$ for this fraction. We consider a simple data structure to approximate $r[P]$ efficiently for any query r. The data structure is just a random subset $S \subset P$ of size m. On query r, the estimate for $r[P]$ we output is $\frac{|S\cap r|}{m}$. The structure S defines an ε-approximation if for all queries r, we have $|r[P] - \frac{|S\cap r|}{m}| \leq \varepsilon$.

1. What m should we take to obtain an ε-approximation with probability $1-\delta$?

Exercice 5. *Algorithme probabiliste pour calculer la médiane*

On étudie un algorithme probabiliste[1] pour déterminer la médiane d'un ensemble $E = \{x_1, \ldots, x_n\}$ de n nombres réels en temps $O(n)$. On rappelle que m est une médiane de E si au moins $\lceil n/2 \rceil$ des élements de E sont inférieurs ou égaux à m, et au moins $\lceil n/2 \rceil$ des élements de E sont supérieurs ou égaux à m. Pour simplifier on suppose n impair (ce qui fait que la médiane est unique) et on suppose aussi que les éléments de E sont tous distincts.

Voici comment fonctionne l'algorithme

(a) Soit $(Y_i)_{1\leq i\leq n}$ une suite de v.a. i.i.d. de loi de Bernoulli de paramètre $n^{-1/4}$. On considère le sous-ensemble aléatoire de E défini par $F = \{x_i : Y_i = 1\}$. Si card $F \leq \frac{2}{3}n^{3/4}$ ou card $F \geq 2n^{3/4}$ on répond «`ERREUR 1`».

(b) On trie F et on appelle d le $\lfloor \frac{1}{2}n^{3/4} - \sqrt{n} \rfloor$ème plus petit élément de F, et u le $\lfloor \frac{1}{2}n^{3/4} - \sqrt{n} \rfloor$ème plus grand élément de F.

(c) On détermine le rang de d et de u dans E (l'élément minimal a rang 1, l'élément maximal a rang n), que l'on note respectivement r_d et r_u. Si $r_d > n/2$ ou $r_u < n/2$ on répond «`ERREUR 2`».

(d) On note $G = \{x_i \in E : d < x_i < u\}$. Si card $G \geq 4n^{3/4}$ on répond «`ERREUR 3`».

(e) On trie G et on renvoie le $(\lceil n/2 \rceil - r_d)$ème élement de G.

1. Justifier pourquoi l'algorithme retourne la médiane en temps $O(n)$ lorsqu'il ne répond pas de message d'erreur.
2. Montrer que pour $i \in \{1,2,3\}$, on a

$$\lim_{n\to\infty} \Pr\left(\text{l'algorithme retourne «}\texttt{ERREUR } i\text{»}\right) = 0.$$

   Pour simplifier l'analyse et éviter d'écrire des symobles $\lfloor\cdot\rfloor$ ou $\lceil\cdot\rceil$, on pourra supposer implicitement que des nombres tels que $\sqrt{n}$, $\frac{1}{2}n^{3/4}$, ... sont des entiers

1. Remarque : il existe un algorithme déterministe de même performance