Correction de l'Examen du cours: Structures et Algorithmes Aléatoires

Temps conseillé: 1h.

1. Problème 1: on considère le modèle de graphe aléatoire vu en cours $G(n,p)$ où $n \in \mathbb{N}$ et $p \in [0,1]$: Ω est l'ensemble des graphes (simples, i.e. sans boucle ni arête multiple) ayant pour sommets $\{1,\dots,n\}$ et pour tout $G \in \Omega$

$$\mathbb{P}(G(n,p) = G) = p^{e_G}(1-p)^{\binom{n}{2}-e_G},$$

avec e_G le nombre d'arêtes de G. Soit X_n le nombre de sommets isolés dans $G(n,p)$. Soit $c_n = np - \log n$. Toutes les limites sont prises quand n tend vers l'infini.

(a) Montrer que si $c_n \to \infty$ alors $\mathbb{P}(X_n > 0) \to 0$.

(b) Montrer que si $c_n \to -\infty$ alors $\mathbb{P}(X_n > 0) \to 1$.

(c) Montrer que si $c_n \to c$ alors $\mathbb{P}(X_n = k) \to \mathbb{P}(Po(e^{-c}) = k)$ où pour tout $\lambda > 0$, $Po(\lambda)$ est une variable aléatoire de Poisson de paramètre λ.

- On écrit $X_n = \sum_{i=1}^n \mathbb{1}(i \text{ est isolé})$. Attention les variables aléatoires $Y_i = \mathbb{1}(i \text{ est isolé})$ ne sont pas indépendantes! Cependant par linéarité de l'espérance, on a:

$$\mathbb{E}[X_n] = n\mathbb{P}(1 \text{ est isolé}) = n(1-p)^{n-1}.$$

On utilise ensuite la méthode du premier moment $\mathbb{P}(X_n > 0) \le \mathbb{E}[X_n]$ et on voit que si $C_n \to \infty$ alors $\mathbb{E}[X_n] \to 0$.

- Dans ce cas $\mathbb{E}[X_n] \to \infty$. L'idée est alors d'appliquer la méthode du second moment:

$$\mathbb{P}(X_n = 0) \le \frac{Var[X_n]}{\mathbb{E}[X_n]^2},$$

Théorème 39 des notes de cours. Il faut alors calculer la variance:

$$\begin{aligned} Var[X_n] &= \sum_{i=1}^n Var[Y_i] + \sum_{i \ne j} Cov[Y_i, Y_j] \\ &\le n(1-p)^{n-1} + n(n-1)[(1-p)^{2n-3} - (1-p)^{2n-2}] \\ &= n(1-p)^{n-1} + n(n-1)p(1-p)^{2n-3}. \end{aligned}$$

On a donc $Var[X_n] = o(\mathbb{E}[X_n]^2)$ et le résultat en découle.

- Dans ce cas, on a $\mathbb{E}[X_n] \to e^{-c}$. De même $\mathbb{E}[X_n(X_n-1)] = \sum_{i \ne j} \mathbb{P}(Y_iY_j = 1) = n(n-1)(1-p)^{2n-3} \to e^{-2c}$. D'une manière générale:

$$\mathbb{E}[X_n(X_n-1)\dots(X_n-k)] = \sum_{i_1,\dots,i_k} \mathbb{P}(Y_{i_1}\dots Y_{i_k} = 1) \to e^{-kc},$$

où la somme est sur les indices $i_1, i_2, \dots i_k$ deux à deux disjoints. On veut montrer que la fonction génératrice de X_n tend vers celle d'une variable de Poisson de paramètre e^{-c}, c'est à dire que $g_{X_n}(t) = \mathbb{E}[t^{X_n}] \to g_{Po(e^{-c})}(t) = e^{e^{-c}(t-1)}$. On a vu en cours que $g^{(k)}_{X_n}(1) = \mathbb{E}[X_n(X_n-1)\dots(X_n-k)]$ et $g^{(k)}_{Po(e^{-c})}(1) = e^{-kc}$. Le résultat voulu découle donc de l'analycité de la fonction génératrice.

2. Problème 2: dans un graphe $G = (V, E)$ ayant pour ensemble de sommets V et pour arêtes E, on note d_v le degré du sommet v: $d_v = |\{w \in V,\ (v, w) \in E\}|$. Un ensemble stable est un ensemble de sommets n'ayant pas d'arête entre eux. La taille d'un ensemble stable est le cardinal de l'ensemble. Soit $\alpha(G)$ la taille maximale d'un ensemble stable de G.

   (a) Montrer que $\alpha(G) \geq \sum_{v \in V} \frac{1}{d_v+1}$.

   (b) Soit $n \geq m$. On définit alors e comme suit: soit q, r le quotient et reste dans la division de n par m, $n = mq + r$ avec $0 \leq r < m$. Alors $e = r\binom{q+1}{2} + (m-r)\binom{q}{2}$. Soit $G_{n,e}$ le graphe ayant n sommets et e arêtes, constitué par la réunion disjointe de m cliques ayant q ou $q+1$ sommets. En particluier $\alpha(G_{n,e}) = m$. Montrer que pour tout graphe H ayant n sommets et e arêtes, on a $\alpha(H) \geq m$.

   (c) Montrer que pour tout graphe H ayant n sommets et e arêtes, si $\alpha(H) = m$ alors H est isomorphe à $G_{n,e}$ (c'est à dire, il existe une bijection ϕ de l'ensemble des sommets de H vers l'ensemble des sommets de $G_{n,e}$ telle que si (i, j) est une arête de H alors $(\phi(i), \phi(j))$ est une arête de $G_{n,e}$).

   - L'idée est de considérer un algorithme simple qui produise un ensemble stable. L'algorithme est le suivant: à chaque itération, l'algorithme choisit un sommet au hasard dans le graphe, le met dans l'ensemble et retire ce sommet et tous ses voisins avant de passer à l'étape suivante. L'ensemble aléatoire I produit par cet algorithme est clairement un ensemble stable. Pour $v \in V$, on définit $X_v = \mathbb{1}(v \in I)$ et $X = |I| = \sum_{v \in V} X_v$. Une autre manière de décrire l'algorithme est la suivante: au début, tous les sommets sont verts; à chaque itération l'algorithme choisit un sommet au hasard dans le graphe, si le sommet choisi est vert, il est mis dans I et ses voisins sont coloriés en rouge, par contre si le sommet choisi est rouge, on ne fait rien; on passe ensuite à l'itération suivante. On a alors pour tout $v \in V$,
   $$\mathbb{E}[X_v] = \frac{1}{1+d_v},$$
   car pour que v appartienne à I, il faut que v soit le premier choisi par l'algorithme parmi tous ses voisins. On a donc
   $$\mathbb{E}[X] = \sum_{v \in V} \frac{1}{d_v+1},$$
   donc il existe I tel que $\sum_{v \in V} \frac{1}{d_v+1} \leq |I| \leq \alpha(G)$.

   - Pour $G_{n,e}$, on a $\sum_{v \in V}(d_v+1)^{-1} = m$ car chaque clique contribue de 1 dans la somme. Si on fixe $e = \sum_v d_v/2$, alors $\sum_v (d_v+1)^{-1}$ est minimisé quand les d_v sont aussi proches les uns des autres que possible. On a donc pour un tel H,
   $$\alpha(H) \geq \sum_v \frac{1}{d_v+1} \geq m.$$

- Pour avoir $\alpha(H) = m$, il faut que les deux inégalités ci-dessus soient des égalités. La seconde égalité implique que les d_v soient aussi proches que possible les uns des autres. Soit $X = |I|$ comme dans la question 1. On suppose que $\alpha(H) = \mathbb{E}[X]$ mais comme $\alpha(H) \geq X$ pour toute réalisation de l'algorithme, X doit être une constante déterministe. On suppose maintenant que H n'est pas une union de cliques. Alors il existe $x, y, z \in V$ avec $\{x, y\}, \{x, z\} \in E$ et $\{y, z\} \notin E$. Soit I l'ensemble obtenu si l'algorithme commence en choisissant x puis y et z et I' celui obtenu par l'algorithme qui commence en choisissant dans l'ordre y, z, puis x. On suppose que la suite des choix est la même dans les deux cas. Alors I et I' sont identiques sauf que $x \in I$ et $y, z \notin I$ tandis que $x \notin I'$ et $y, z \in I'$. Donc X n'est pas constant. Nous venons de montrer que $\alpha(H) = \mathbb{E}[X]$ implique que H est une union de cliques d'où le résultat (connu sous le nom de Théorème de Turán).