**Examen partiel**
15 avril 2024

On pourra utiliser les variantes suivantes des inégalités de CHERNOFF.

(I) Si $X$ est la somme de $n$ variables aléatoires i.i.d. de loi de BERNOULLI de paramètre 1/2, alors pour tout $a > 0$

$$\mathbf{P}(X \geqslant \frac{n}{2} + a) \leqslant \exp(-2a^2/n)$$
$$\mathbf{P}(X \leqslant \frac{n}{2} - a) \leqslant \exp(-2a^2/n)$$

(II) Si $X$ est une somme de variables aléatoires indépendantes ayant chacune une loi de BERNOULLI, alors en posant $\mu = \mathbf{E}[X]$, pour tout $0 < \varepsilon < 1$,

$$\mathbf{P}(|X - \mu| \geqslant \varepsilon\mu) \leqslant 2\exp(-\mu\varepsilon^2/3)$$

## Exercice 1 Graphes orientés aléatoires

Étant donnés $n \in \mathbf{N}$ et $p \in [0, 1]$, on définit un graphe orienté $\Gamma_{n,p}$, sans boucle ni arête multiple, de la manière suivante. On se donne une famille $(X_{i,j})_{1\leqslant i,j\leqslant n, i\neq j}$ de variables aléatoires i.i.d. de loi de BERNOULLI de paramètre $p$ et on pose $\Gamma_{n,p} = (V, E)$ avec $V = \{1, \ldots, n\}$ et

$$E = \{(i, j) \in V^2 \, : \, i \neq j \text{ et } X_{i,j} = 1\}.$$

Étant donné un sommet $x \in V$, on note $d_{in}(x)$ son degré entrant (c'est à dire le nombre d'arêtes de la forme $(y, x)$ et $d_{out}(x)$ son degré sortant (c'est à dire le nombre d'arêtes de la forme $(x, y)$.

1. Quelle est la loi du nombre d'arêtes de $\Gamma_{n,p}$ ?
2. Quelle est l'espérance du nombre de sommets isolés ? (Un sommet est dit isolé si son degré entrant et son degré sortant sont nuls.)
3. Dans cette question on pose $p = c\log(n)/n$ pour un nombre réel $c > 0$.
   (a) Montrer que si $c > 1/2$, alors
   $$\lim_{n\to\infty} \mathbf{P}(\Gamma_{n,p} \text{ a un sommet isolé}) = 0.$$
   (b) Montrer que si $c < 1/2$, alors
   $$\lim_{n\to\infty} \mathbf{P}(\Gamma_{n,p} \text{ a un sommet isolé}) = 1.$$
4. Montrer qu'il existe une constante $C$ telle que, quelle que soit la suite $(p_n)$ dans $[0, 1]$,
$$\lim_{n\to\infty} \mathbf{P}(\text{pour tout sommet } x \text{ de } \Gamma_{n,p_n} \text{ on a } |d_{in}(x) - d_{out}(x)| \leqslant C\sqrt{n \log n}) = 1$$

## Exercice 2 Points aléatoires dans l'hypercube

On définit la *distance de Hamming* $d$ sur l'hypercube $H_n = \{-1, 1\}^n$ comme suit : si $x = (x_i)$ et $y = (y_i)$ sont dans $H_n$, alors $d(x, y)$ est le nombre de coordonnées $i \in \{1, \ldots, n\}$ telles que $x_i \neq y_i$.

1. Montrer que si $X$ et $Y$ sont des variables aléatoires indépendantes de loi uniforme sur $H_n$, alors pour tout $\delta > 0$,
$$\mathbf{P}\left(d(X, Y) \leqslant \frac{n}{2} - \delta\right) \leqslant \exp(-2\delta^2/n).$$
2. En appliquant cette inégalité pour $\delta = n/4$, montrer le résultat suivant : pour tout entier $n$, on peut trouver $\exp(n/16)$ vecteurs de norme 1 dans l'espace euclidien $\mathbf{R}^n$ qui sont deux à deux à distance au moins 1.

**Exercice 3 Sous-ensembles indépendants dans un graphe**

Soit $G = (V, E)$ un graphe non orienté, sans boucle ni arête multiple. On note $\alpha(G)$ le cardinal maximal d'un sous-ensemble de sommets deux à deux non adjacents. On suppose $|E| \geqslant |V|/2$. On pose $d = 2|E|/|V|$ et on considère le processus suivant.

(i) On supprime chaque sommet (ainsi que les arêtes adjacentes) indépendamment avec probabilité $1 - 1/d$.

(ii) Pour chaque arête restante, on la supprime ainsi qu'une de ses extrémités choisie arbitrairement.

1. Soit $X$ le nombre de sommets (resp. $Y$ le nombre d'arêtes) dans le graphe obtenu après l'étape (i). Calculer $\mathbf{E}[X]$ et $\mathbf{E}[Y]$.
2. Démontrer que $\alpha(G) \geqslant \frac{|V|^2}{4|E|}$.

**Exercice 4 Distance minimale dans un nuage de points**

On étudie un algorithme probabiliste pour déterminer, étant donné un ensemble $E \subset [0,1]^2$ de $n$ points, la distance minimale entre deux points distincts de $E$. La distance utilisée est la distance euclidienne, définie pour $p = (x, y)$ et $p' = (x', y')$ par

$$d(p, p') = \sqrt{(x - x')^2 + (y - y')^2}.$$

L'algorithme utilise un recouvrement de $[0,1]^2$ par des $\varepsilon$-carrés (c'est-à-dire des carrés de côté $\varepsilon$). Il est constitué de $N^2$ $\varepsilon$-carrés $\{C_{s,t} \,:\, 0 \leqslant s < N,\ 0 \leqslant t < N\}$, où $N = \lceil 1/\varepsilon \rceil$ et

$$C_{s,t} = \{(x, y) \in [0,1]^2 \,:\, s\varepsilon \leqslant x < (s+1)\varepsilon,\ t\varepsilon \leqslant y < (t+1)\varepsilon\}.$$

On dit que deux $\varepsilon$-carrés $C_{s,t}$ et $C_{s',t'}$ sont proches si $|s - s'| \leqslant 2$ et $|t - t'| \leqslant 2$. On utilise un dictionnaire pour déterminer quel(s) point(s) d'un ensemble $A$ appartient à un $\varepsilon$-carré donné. La création du dictionnaire prend un temps $O(|A|)$ et sa consultation prend un temps $O(1)$.

On commence par numéroter les points $E = \{p_1, \ldots, p_n\}$ selon un ordre aléatoire, choisi uniformément parmi les $n!$ permutations possibles. Pour $2 \leqslant i \leqslant n$, l'algorithme calculera la distance minimale entre deux points parmi $\{p_1, \ldots, p_i\}$, notée $\delta_i$. Voici l'algorithme.

(I) Initialiser en posant $\delta_2 = d(p_1, p_2)$ et en créant le dictionnaire associé à $\varepsilon = \delta_2/2$ et à l'ensemble $A = \{p_1, p_2\}$.

(II) Pour $i$ allant de 3 à $n$

  (a) Pour $\varepsilon = \delta_{i-1}/2$, déterminer quel $\varepsilon$-carré $C_{s,t}$ contient $p_i$, puis calculer $d(p_i, p_j)$ pour tous les points $p_j$ $(j < i)$ qui sont dans un $\varepsilon$-carré proche de $C_{s,t}$.

  (b) Si on détecte à l'étape (a) un $j < i$ tel que $d(p_i, p_j) < \delta_{i-1}$, poser $\delta_i = d(p_i, p_j)$, effacer le dictionnaire précédent et créer un nouveau dictionnaire associé à $\varepsilon = \delta_i/2$ et $A = \{p_1, \ldots, p_i\}$.
  Sinon, poser $\delta_i = \delta_{i-1}$ et ajouter $p_i$ au dictionnaire.

1. Expliquer brièvement pourquoi l'algorithme calcule la distance minimale.
2. Pour $3 \leqslant i \leqslant n$, on définit une variable booléenne aléatoire $X_i$ comme valant 1 si $\delta_i < \delta_{i-1}$ et 0 sinon. Expliquer brièvement pourquoi la complexité $C$ de l'algorithme est $O(n + \sum_{i=3}^{n} iX_i)$.
3. Montrer que la complexité moyenne vérifie $\mathbf{E}[C] = O(n)$.