TD 06 – Inégalité de Chernoff et Graphes aléatoires (corrigé)

Exercice 1. *Blackjack*

Vous êtes le croupier dans une partie de blackjack et vous soupçonnez un joueur de tricher en comptant les cartes car, sur les quelques premières mains que vous venez de le voir jouer, il gagne 55% du temps (alors, que, sans tricher, la probabilité de gagner une main est 1/2). Cependant, vous voulez attendre d'avoir un peu plus de certitude avant de démasquer le joueur.

1. On suppose que le joueur continue de gagner 55% du temps. Combien de mains devez-vous le laisser jouer avant d'être sûr à 90% qu'il triche ?

☞ On pose $\varepsilon = 0.05$ et $\delta = 0.1$. Supposons que le joueur joue n mains et on note $X_i = 1$ si le joueur a gagné la i-ème main, 0 sinon. On pose $X = \sum i = 1^n X_i$. On va dénoncer le tricheur donc on veut se tromper avec proba au plus δ. On se trompe si le joueur ne trichait pas, auquel cas X_i vaut 1 avec proba 1/2, et donc $\mathbf{E}[X] = n/2$. On veut donc que, avec ces hypothèses, $\mathbf{P}\{X/n - 1/2 \geq \varepsilon\} \leq \delta$. Or $\mathbf{P}\{X/n - 1/2 \geq \varepsilon\} = \mathbf{P}\{X - \mathbf{E}[X] \geq n\varepsilon\}$ donc on peut appliquer Hoeffding et on obtient :

$$\mathbf{P}\{X/n - 1/2 \geq \varepsilon\} = \mathbf{P}\{X - \mathbf{E}[X] \geq n\varepsilon\} \leq e^{-\frac{2\varepsilon^2 n^2}{n}} = e^{-2\varepsilon^2 n} .$$

On veut que ceci soit $\leq \delta$ donc il suffit de prendre $n \geq \frac{1}{2\varepsilon^2}\log(1/\delta) \approx 460$ mains.

Exercice 2. *Random Algorithm*

Suppose you are given a randomized polynomial-time algorithm $\mathcal{A}$ for deciding whether $x \in \{0,1\}^*$ is in the language L or not. Suppose it has the following property. If $x \in L$, then $\mathbf{P}\{\mathcal{A}(x) = 0\} \leq 1/4$ and if $x \notin L$, then $\mathbf{P}\{\mathcal{A}(x) = 1\} \leq 1/3$. Note that the probability here is taken over the randomness used by the algorithm $\mathcal{A}$ and *not* over the input x.

1. Construct a randomized polynomial-time algorithm $\mathcal{B}$ that is allowed to make independent calls to $\mathcal{A}$ such that for all inputs $x \in \{0,1\}^*$, we have $\mathbf{P}\{\mathcal{B}(x) = \mathbf{1}_{x\in L}\} \geq 1 - 2^{-|x|}$. Here $\mathbf{1}_{x\in L} = 1$ if $x \in L$ and 0 otherwise, and $|x|$ denotes the length of the bitstring x.

☞ On pose $n = |x|$. On appelle N fois l'algorithme $\mathcal{A}$, et on note $X_1, \ldots, X_N$ les réponses successives. $\mathcal{B}$ renvoie la réponse majoritaire parmi ces N appels (celle qui apparaît au moins $N/2$ fois ; en cas d'égalité, on choisit arbitrairement). Il faut choisir N.

On veut que la probabilité que l'algorithme $\mathcal{B}$ se trompe soit au plus 2^{-n}. Commençons par le cas où $x \notin L$. On note $p = \mathbf{P}\{\mathcal{A}(x) = 1\}$ sachant que $x \notin L$ (avec donc $p \leq 1/3$). On pose $X = \sum_{i=1}^N X_i$, alors $\mathbf{E}[X] = pN \leq N/3$. On va appliquer la variante de Chernoff (celle avec $\mu_H \geq \mathbf{E}[X]$) en prenant $\delta = 1/2$ et $\mu_H = N/3$ de telle sorte que $(1+\delta)\mu_H = N/2$.

$$\mathbf{P}\{\mathcal{B}(x) = 1\} = \mathbf{P}\{X \geq N/2\} = \mathbf{P}\{X \geq (1+\delta)\mu_H\} \leq e^{-\delta^2\mu_H/3} = e^{-N/36} .$$

Maintenant pour $x \in L$, $\mathbf{E}[X] \geq 3N/4$ donc on applique la variante de Chernoff avec $\mu_L = 3N/4$ et $\delta = 1/3$ de telle sorte que $(1-\delta)\mu_L = N/2$, et donc :

$$\mathbf{P}\{\mathcal{B} = 0\} = \mathbf{P}\{X \leq N+2\} = \mathbf{P}\{X \leq (1-\delta)\mu_L\} \leq e^{-\delta^2\mu_L/2} = e^{-N/24} .$$

Donc la probabilité que $\mathcal{B}$ se trompe est bornée par $\max(e^{-N/24}, e^{-N/36}) = e^{-N/36}$. En prenant $N = 36n/\log e$, on obtient que la probabilité que l'algorithme se trompe est au plus 2^{-n}.

Exercice 3. *Interrupteurs*

Partie I :

1. Montrer qu'il existe une constante $\gamma > 0$ rendant l'énoncé suivant vrai : si une v.a. positive X vérifie $\mathsf{E}[X] = 1$ et $\mathsf{E}[X^2] \leq 3$, alors $\mathsf{P}(X \geq 1/4) \geq \gamma$.
Indication : définir la variable aléatoire $Y = \mathbf{1}_{X\geq 1/4}$ et se ramener à l'inégalité de Cauchy-Schwarz $\mathbb{E}(XY) \leq \sqrt{\mathbb{E}(X^2)\mathbb{E}(Y^2)}$

☞ On écrit

$$1 = \mathsf{E}[X] = \mathsf{E}[X\mathbf{1}_{X<1/4}] + \mathsf{E}[X\mathbf{1}_{X\geq 1/4}] \leq \frac{1}{4} + \mathsf{E}[X\mathbf{1}_{X\geq 1/4}].$$

Par l'inégalité de Cauchy–Schwarz, $\mathsf{E}[X\mathbf{1}_{X\geq 1/4}] \leq \sqrt{\mathsf{E}[X^2]\mathsf{P}(X \geq 1/4)} \leq \sqrt{3}\sqrt{\mathsf{P}(X \geq 1/4)}$. On obtient la minoration voulue pour $\gamma = 3/16$.

2. Soient $(X_1, \dots, X_n)$ des v.a. i.i.d. vérifiant $\mathsf{P}(X_i = 1) = \mathsf{P}(X_i = -1) = \frac{1}{2}$. On pose $Y = \frac{1}{\sqrt{n}}(X_1 + \cdots + X_n)$. Calculer $\mathsf{E}[Y^2]$ et $\mathsf{E}[Y^4]$ et en déduire que

$$\mathsf{E}[|X_1 + \cdots + X_n|] \geq \frac{\gamma}{2}\sqrt{n}.$$

☞ On a $\mathsf{E}[Y^2] = \frac{1}{n} \cdot \mathbf{Var}[Y] = \frac{1}{n} \cdot \sum_i \mathbf{Var}[X_i] = 1$ (par inédependance). On a ensuite

$$\mathsf{E}[Y^4] = \frac{1}{n^2} \cdot \sum_{i,j,k,l=1}^{n} \mathsf{E}[X_i X_j X_k X_l].$$

L'indépendance des X_i et le fait que $\mathsf{E}[X_i] = 0$ implique $\mathsf{E}[X_i X_j X_k X_l] = 0$ dès qu'un indice apparaît une unique fois parmi $\{i, j, k, l\}$. Les seuls termes non nuls sont ceux où $i = j = k = l$ ou $i = j \neq k = l$ ou $i = k \neq j = l$ ou $i = l \neq j = k$. On a donc

$$\mathsf{E}[Y^4] = 1/n^2(n + 3n(n-1)) = 3 - 2/n \leq 3.$$

On applique la question précédente à $X = Y^2$, d'où $\mathsf{P}(Y^2 \geq 1/4) = \mathsf{P}(|X_1 + \cdots + X_n| \geq \frac{\sqrt{n}}{2}) \geq \gamma$. Enfin,

$$\mathsf{E}[|X_1 + \cdots + X_n|] \geq \frac{\sqrt{n}}{2}\mathsf{P}\left(|X_1 + \cdots + X_n| \geq \frac{\sqrt{n}}{2}\right) \geq \frac{\gamma\sqrt{n}}{2}.$$

Partie II :
On considère une grille $n \times n$ d'ampoules ainsi que 3 séries d'interrupteurs : des interrupteurs $a = (a_{ij})_{1 \leq i,j \leq n}$ associés à chaque ampoule, des interrupteurs $b = (b_i)_{1 \leq i \leq n}$ associés à chaque ligne et des interrupteurs $c = (c_j)_{1 \leq j \leq n}$ associés à chaque colonne. Chaque interrupteur prend la valeur -1 ou 1. L'ampoule en position (i, j) est allumée si et seulment si $a_{ij} b_i c_j = 1$. On considère la quantité

$$F(a, b, c) = \sum_{i,j=1}^{n} a_{ij} b_i c_j$$

qui est le nombre d'ampoules allumées moins le nombre d'ampoules éteintes. Enfin, deux joueurs jouent au jeu suivant : le joueur 1 choisit la position des interrupteurs (a_{ij}), puis le joueur 2 choisit la position des interrupteurs (b_i) et (c_j). Le joueur 1 veut minimiser $F(a, b, c)$ et joueur 2 veut le maximiser. On considère donc

$$V(n) = \min_{a \in \{-1,1\}^{n \times n}} \max_{b,c \in \{-1,1\}^n} F(a, b, c).$$

3. Montrer que $V(n) = O(n^{3/2})$ en considérant le cas où le joueur 1 joue au hasard.
☞ Soit $(a_{ij})_{1 \leq i,j \leq n}$ des v.a. i.i.d. de loi uniforme sur $\{-1, 1\}$. Quel que soit le choix de b et c, on a

$$\mathsf{P}(F(a, b, c) \geq t) \leq \exp(-t^2/2n^2)$$

par l'inégalité de Chernoff (en effet, $F(a, b, c)$ est la somme de n^2 v.a. de loi uniforme sur $\{-1, 1\}$). Par la borne de l'union,

$$\mathsf{P}(\max_{b,c} F(a, b, c) \geq t) \leq 4^n \exp(-t^2/2n^2).$$

Lorsque $t > \sqrt{2n^3 \log 4}$, cette probabilité est < 1 et donc $\mathsf{P}(\max_{b,c} F(a, b, c) < t) > 0$: il existe donc un choix de a tel que $\max_{b,c} F(a, b, c) < t$, d'où $V(n) = O(n^{3/2})$.

4. Le joueur 2 applique la stratégie suivante : il choisit b au hasard, puis ensuite choisit c de façon à allumer le maximum de lampes. Estimer le nombre moyen de lampes allumées par cette stratégie à l'aide de la question I.2 et en déduire que $V(n) = \Omega(n^{3/2})$.
☞ Fixons $a = (a_{ij})$ et choisissons (b_i) i.i.d. de loi uniforme sur $\{-1, 1\}$. On a alors

$$\max_c F(a, b, c) = \sum_{i=1}^{n} \left| \sum_{j=1}^{n} a_{ij} b_j \right|.$$

En utilisant la linéarité de l'espérance, le fait que $(b_j)_j$ et $(a_{ij} b_j)_j$ ont même loi et la question I.2, il vient

$$\mathsf{E} \max_c F(a, b, c) = n\mathsf{E}\left| \sum_{j=1}^{n} b_j \right| \geq \frac{n^{3/2}\gamma}{2}.$$

En particulier, pour tout choix de a, il existe b tel que $\max_c F(a, b, c) \geq \frac{n^{3/2}\gamma}{2}$.

Exercice 4. *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 $H_{2n,p}$?

   ☞ 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 $\mathsf{E}[N] = \sum \mathsf{P}(A_i) = 2n(1-p)^n$.

3. Dans cette question on pose $p = c\log(n)/n$ pour un nombre réel $c > 0$.
   1. Montrer que si $c > 1$, alors
   $$\lim_{n\to\infty} \mathsf{P}(H_{2n,p} \text{ a un sommet isolé}) = 0.$$
   2. Montrer que si $c < 1$, alors
   $$\lim_{n\to\infty} \mathsf{P}(H_{2n,p} \text{ a un sommet isolé}) = 1.$$

   ☞
   1. Si $c > 1$, on a $\mathsf{E}[N] = 2n\exp(n\log(1 - \frac{c\log n}{n})) \to 0$ et donc $\mathsf{P}(N \geq 1) \leq \mathsf{E}[N] \to 0$.
   2. Si $c < 1$, on calcule
   $$\mathsf{E}[N^2] = \sum_{i,j=1}^{2n} \mathsf{P}(A_i \cap A_j) = 2n(1-p)^n + 2n(n-1)(1-p)^{2n} + 2n^2(1-p)^{2n-1}$$
   d'où il vient que $\mathsf{E}[N^2]/\mathsf{E}[N]^2$ tend vers 1. On utilise l'inégalité de Tchebychev pour conclure que
   $$\mathsf{P}(N = 0) = \mathsf{P}(\mathsf{E}[N] - N \geq \mathsf{E}[N]) \leq \mathsf{P}(|N - \mathsf{E}[N]| \geq \mathsf{E}[N]) \leq \frac{\mathbf{Var}[N]}{\mathsf{E}[N]^2} = \frac{\mathsf{E}[N^2]}{\mathsf{E}[N]^2} - 1 \to 0.$$

4. Dans cette question on pose $p = 1/2$. Montrer qu'il existe une constante $C > 0$ telle que
$$\lim_{n\to\infty} \mathsf{P}\left(\text{tous les sommets de } H_{2n,p} \text{ ont un degré inférieur à } \frac{n}{2} + C\sqrt{n\log n}\right) = 1.$$

   ☞ Le degré d_i du sommet i suit la loi $B(n, 1/2)$. Par l'inégalité de Chernoff I, on a donc
   $$\mathsf{P}(d_i \geq \frac{n}{2} + a) \leq \exp(-2a^2/n).$$
   Ainsi, par la borne de l'union,
   $$\mathsf{P}(\max_i d_i \geq \frac{n}{2} + a) \leq 2n\exp(-2a^2/n).$$
   Cette quantité tend vers 0 si $a = C\sqrt{n\log n}$ avec $2C^2 > 1$.

Exercice 5. *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.

☞ 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 zero 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 $\Pr(X \neq 0) \to 0$ 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 zero quand n tend vers l'infini. Comme X est a valeur entières, positives 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 $\mathbf{P}\{X = 0\} \to 0$ 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 $\frac{\mathbf{Var}[X]}{\mathbf{E}[X]^2} \to 0$.

☞ 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

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

☞ On a

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

On observe ensuite que $\mathbf{E}\left[(X_i - \mathbf{E}[X_i])^2\right] = \mathbf{E}\left[X_i^2\right] - \mathbf{E}[X_i]^2 \leq \mathbf{E}\left[X_i^2\right]$. Mais comme X_i est à valeur dans $0, 1$, on a $\mathbf{E}\left[X_i^2\right] = \mathbf{E}[X_i]$. 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 défini X_i la variables 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 $\mathbf{Var}[X] \leq \mathbf{E}[X] + \sum_{i \neq j} \mathbf{E}\left[(X_i - \mathbf{E}[X_i])(X_j - \mathbf{E}\left[X_j\right])\right]$. Fixons $i \neq j$ et considérons $\mathbf{E}\left[(X_i - \mathbf{E}[X_i])(X_j - \mathbf{E}\left[X_j\right])\right] = \mathbf{E}\left[X_i X_j\right] - \mathbf{E}[X_i]\,\mathbf{E}\left[X_j\right]$. On a $\mathbf{E}\left[X_i X_j\right] = \mathbf{P}\left\{C_i \text{ et } C_j \text{ sont des cliques}\right\} = p^k$, où k est le nombre d'arêtes nécessaires pour que C_i et C_j soient des cliques. Ce nombre d'arêtes va dépendre nu 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 $\mathbf{E}\left[(X_i - \mathbf{E}[X_i])(X_j - \mathbf{E}\left[X_j\right])\right] = 0$.
- Si $|C_i \cap C_j| = 2$, alors $k = 11$ et $\mathbf{E}\left[(X_i - \mathbf{E}[X_i])(X_j - \mathbf{E}\left[X_j\right])\right] = p^{11}(1-p)$. Il y a $\binom{n}{4} \cdot \binom{4}{2} \cdot \binom{n-4}{2}$ tels couples (C_i, C_j).
- Si $|C_i \cap C_j| = 3$, alors $k = 9$ et $\mathbf{E}\left[(X_i - \mathbf{E}[X_i])(X_j - \mathbf{E}\left[X_j\right])\right] = p^9(1-p^3)$. Il y a $\binom{n}{4} \cdot \binom{4}{3} \cdot \binom{n-4}{1}$ 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 $\mathbf{Var}[X] \leq \mathbf{E}[X] + \binom{n}{4} \cdot \binom{4}{2} \cdot \binom{n-4}{2} p^{11}(1-p) + \binom{n}{4} \cdot \binom{4}{3} \cdot \binom{n-4}{1} p^9(1-p^3)$. 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 $\Pr(X \neq 0) \to 0$ quand n tend vers l'infini.