

TD 04 – Inégalités de Chernoff (1)

Exercice 1.*Comparer les bornes*

On a un dé équilibré à 6 côtés. On lance le dé n fois et on note X le nombre de fois où le dé tombe sur 6. Soit q la probabilité de l'événement $X \geq n/4$.

1. Comparer les bornes supérieures que l'on obtient en utilisant l'inégalité de Markov, l'inégalité de Chebyshev, et une borne de Chernoff (la plus adaptée).

Exercice 2.*Chernoff Interval*

- Soit X une variable aléatoire quelconque avec $0 \leq X \leq 1$ et $\mathbb{E}[X] = p$. Considérons la variable aléatoire $Y \in \{0, 1\}$ telle que $\Pr(Y = 1) = p$. Montrer que pour tout $\lambda > 0$,

$$\mathbb{E}[e^{\lambda X}] \leq \mathbb{E}[e^{\lambda Y}].$$

Indice : on pourra utiliser la convexité de la fonction exponentielle.

- En utilisant ce fait, montrer que la borne de Chernoff vue en cours reste valable si l'on remplace l'hypothèse $X_i \in \{0, 1\}$ par $X_i \in [0, 1]$.

Exercice 3.*Calcul de la médiane*

On étudie un algorithme probabiliste¹ pour déterminer la médiane d'un ensemble $E = \{x_1, \dots, x_n\}$ de n nombres réels en temps $\mathcal{O}(n)$. On rappelle que m est une médiane de E si au moins $\lceil n/2 \rceil$ des éléments de E sont inférieurs ou égaux à m , et au moins $\lfloor n/2 \rfloor$ des éléments 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 $\text{card } F \leq \frac{2}{3}n^{3/4}$ ou $\text{card } F \geq 2n^{3/4}$ 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 $\text{card } G \geq 4n^{3/4}$ on répond «ERREUR 3».
- (e) On trie G et on renvoie le $(\lceil n/2 \rceil - r_d)$ ème élément de G .

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

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

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

Exercice 4.*Interrupteurs*

1. Montrer qu'il existe une constante $\gamma > 0$ rendant l'énoncé suivant vrai :

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

« Si une v.a. positive X vérifie $\mathbf{E}[X] = 1$ et $\mathbf{E}[X^2] \leq 3$, alors $\mathbf{P}\{X \geq 1/4\} \geq \gamma$. »

Indication : définir la variable aléatoire $Y = \mathbf{1}_{X \geq 1/4}$ et se ramener à l'inégalité de Cauchy-Schwarz.
 $\mathbf{E}(XY) \leq \sqrt{\mathbf{E}(X^2)\mathbf{E}(Y^2)}$

2. Soient (X_1, \dots, X_n) des v.a. i.i.d. vérifiant $\mathbf{P}\{X_i = 1\} = \mathbf{P}\{X_i = -1\} = \frac{1}{2}$.

On pose $Y = \frac{1}{\sqrt{n}}(X_1 + \dots + X_n)$. Calculer $\mathbf{E}[Y^2]$ et $\mathbf{E}[Y^4]$ et en déduire que :

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

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 seulement si $a_{ij} \times b_i \times c_j = 1$. On considère la quantité

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

qui est le nombre d'ampoules allumées moins le nombre d'ampoules éteintes.

Deux joueuses jouent au jeu suivant :

1. la joueuse 1 choisit la position des interrupteurs (a_{ij}) ,
2. puis la joueuse 2 choisit la position des interrupteurs (b_i) et (c_j) .

La joueuse 1 veut minimiser $\mathbf{F}(a, b, c)$ et la joueuse 2 veut le maximiser. On considère donc :

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

3. Montrer que $\mathbf{V}(n) = \mathcal{O}(n^{3/2})$ en considérant le cas où la joueuse 1 joue au hasard.
4. La joueuse 2 applique la stratégie suivante : elle 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 (indication : utiliser la question 2) et en déduire que $\mathbf{V}(n) = \Omega(n^{3/2})$.