TD 05 – Inégalité de Chernoff: applications

Exercice 1. *Tester la pièce*

1. On considère une pièce qui tombe sur pile avec probabilité p. Combien de fois doit-on lancer la pièce pour déterminer p à ± 0.1 avec probabilité au moins 0.9 ?

Exercice 2. *BucketSort*

Le tri par seaux est un algorithme de tri très simple dont la complexité moyenne est linéaire. Les hypothèses sont les suivantes : nous avons $n = 2^m$ nombres à trier, tirés uniformément et indépendamment sur $\{0, \dots, 2^k - 1\}$ avec $k \geq m$ (k est supposé connu). L'algorithme procède en deux étapes : d'abord, il effectue un pré-tri en jetant selon certaines règles les n éléments dans n seaux ; ensuite, il appelle un algorithme de tri simple (en temps quadratique, tel que tri par insertion ou tri par sélection) dans chaque seau. Enfin, il concatène dans l'ordre les listes triées obtenues dans chaque seau. Pour que l'algorithme soit exact, il faut que le pré-tri soit fait de telle sorte que tous les éléments du seau i soient inférieurs à tous les éléments du seau j, pour $i < j$.

1. Donner une façon très simple de faire le pré-tri tout en respectant la condition énoncée ci-dessus. On veut que le choix du seau pour un élément x soit effectué en temps constant (on suppose ici que les opérations arithmétiques peuvent être effectuées en temps constant).
2. Soit X_i la v.a. comptant le nombre d'éléments dans le seau i après le pré-tri. Quelle loi suit X_i ?
3. Prouver que la complexité en moyenne est $\mathcal{O}(n)$.

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 $\mathbb{E}(XY) \leq \sqrt{\mathbb{E}(X^2)\mathbb{E}(Y^2)}$
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
$$\mathsf{E}[|X_1 + \cdots + X_n|] \geq \frac{\gamma}{2}\sqrt{n}.$$

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 seulement 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.
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 $V(n) = \Omega(n^{3/2})$.

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. *Chernoff Bound Interval*

Let X be an arbitrary random variable with $0 \leq X \leq 1$ and $\mathbf{E}[X] = p$. Consider the random variable $Y \in \{0, 1\}$ with $\mathbf{P}\{Y = 1\} = p$. Show that for any $\lambda > 0$, $\mathbf{E}\left[e^{\lambda X}\right] \leq \mathbf{E}\left[e^{\lambda Y}\right]$. *Hint : you may want to use the convexity of the exponential function.*

Using this fact, show that the Chernoff bound we saw in class still holds if we replace the condition $X_i \in \{0, 1\}$ by $X_i \in [0, 1]$.