TD 04 – Moments and Midterm Preparation

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. *Comparer Markov, Chebyshev, Chernoff bounds*

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 3. *Sondage*

On veut faire un sondage d'opinion pour estimer la proportion p de la population qui est en accord avec le Président. Supposons que l'on interroge n personnes choisies uniformément et indépendamment au hasard, et que chacune d'elle réponde par "Oui, je suis d'accord" ou "Non, je ne suis pas d'accord". Étant donnés $\theta > 0$ et $0 < \delta < 1$, on souhaite trouver une estimation $\overline{X}$ de p telle que

$$\mathbf{P}\left\{|\overline{X} - p| \leq \theta\right\} > 1 - \delta \,.$$

Par exemple pour $1 - \delta = 0.95$, on pourra ainsi dire que le sondage a une précision de θ à 95%.

1. Que choisir comme estimation $\overline{X}$ de p?

2. Combien de personnes doit-on interroger pour que l'estimation $\overline{X}$ vérifie nos conditions ? Autrement dit, donner une borne inférieure sur n en termes de θ et δ. On remarquera que cette borne ne dépend pas de la taille de la population totale.

3. Calculer la valeur de n obtenue grâce à votre borne pour les paramètres $\theta = 0.2$ et $1 - \delta = 95\%$.

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 $H_{2n,p}$?

2. Quelle est l'espérance du nombre de sommets isolés de $H_{2n,p}$?

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

$$\lim_{n\to\infty} \mathsf{P}(H_{2n,p} \text{ a un sommet isolé}) = 1.$$

4. Dans cette question on pose $p = 1/2$. Montrer qu'il existe une constante $C > 0$ telle que

$$\lim_{n\to\infty} \mathsf{P}\left(\text{tous les sommets de } H_{2n,p} \text{ ont un degré inférieur à } \frac{n}{2} + C\sqrt{n \log n}\right) = 1.$$

Exercice 5. *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 $\{0, \ldots, 2^k - 1\}$ 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 $\mathcal{O}(n)$.