TD 03 – Variables Aléatoires, Markov, Chebyshev et Chernoff

Exercice 1. *Inégalité de Jensen*

Soit f une fonction convexe et X une variable aléatoire à valeurs réelles.
L'*inégalité de Jensen*[1] affirme : $\mathbf{E}[f(X)] \geq f(\mathbf{E}[X])$. En supposant que f soit $\mathcal{C}^1$, montrer cette inégalité.

Exercice 2. *Coquilles dans un TD (ça n'arrive jamais !)*

1. Une feuille de TD contient 4 coquilles. À chaque relecture, une coquille non corrigée est corrigée avec probabilité $\frac{1}{3}$. Les relectures et les corrections sont indépendantes les unes des autres.
   Combien de relectures faut-il faire au minimum pour que la probabilité qu'il ne reste aucune coquille soit supérieure à 0,9 ?
2. À quel problème vu en cours vous fait penser la situation précédente ?
3. Soit X une variable aléatoire d'espérance $\mu = 10$ et d'écart-type $\sigma = 5$.
   Montrer que pour tout $n \geq 50$, on a $\mathbf{P}\{\mu - n < X < \mu + n\} \geq 0{,}99$.

Exercice 3. *Un tri pas sot*

Le tri par seaux est un algorithme de tri très simple dont la complexité moyenne est linéaire.
Étant donnés k et m deux entiers avec $k \geq m$, on souhaite trier $n = 2^m$ entiers, tirés uniformément et indépendamment sur $\{0, \ldots, 2^k - 1\}$. L'algorithme est le suivant :

a) on effectue un pré-tri en répartissant selon certaines règles les n entiers dans n seaux ;
b) on appelle un algorithme de tri simple (en temps quadratique, par exemple le tri par insertion) dans chaque seau ;
c) on concatène dans l'ordre les listes triées obtenues dans chaque seau.

Pour que l'algorithme soit exact, il faut bien sûr que le pré-tri soit fait de sorte que pour tous $i < j$, tous les éléments du seau i sont inférieurs à tous les éléments du seau 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 4. *Réduisons les erreurs*

Soit $L \subseteq \{0,1\}^*$ un langage, et $\mathcal{A}$ un algorithme probabiliste qui décide en temps polynomial si une entrée $x \in \{0,1\}$ est dans le langage L ou non. On suppose que $\mathcal{A}$ a la propriété suivante :

$$\text{si } x \in L, \text{ alors } \mathbf{P}\{\mathcal{A}(x) = 0\} \leq \frac{1}{4} \qquad \text{si } x \in L, \text{ alors } \mathbf{P}\{\mathcal{A}(x) = 1\} \leq \frac{1}{3}.$$

Attention, cette probabilité vient de l'aléatoire lié à l'algorithme $\mathcal{A}$, *pas* du choix de l'entrée x.
Pour tout $x \in \{0,1\}^*$, on note $|x|$ sa longueur et on définit $\mathbf{1}_{x \in L}$ qui vaut 1 si $x \in L$ et 0 sinon. Construisez un algorithme probabiliste polynomial $\mathcal{B}$ (qui peut faire un ou des appels indépendants à $\mathcal{A}$) tel que pour toute entrée $x \in \{0,1\}^*$, on a :

$$\mathbf{P}\{\mathcal{B}(x) = \mathbf{1}_{x \in L}\} \geq 1 - 2^{-|x|}.$$

1. Johan Jensen, mathématicien et ingénieur danois (1859 – 1925).

Exercice 5. *Top Chrono*

Soit $\mathcal{A}$ un algorithme déterministe qui prend en entrée une chaîne de n bits et dont l'espérance du temps d'exécution est $\mathcal{O}(n^2)$ si l'entrée est choisie aléatoirement de manière uniforme.

1. Soit $f(n)$ une fonction tendant vers $+\infty$ avec n. Montrer que la probabilité que le temps d'exécution soit supérieur à $n^2 f(n)$ tend vers zéro quand n tend vers l'infini.
2. Que pouvons nous en déduire sur le temps d'exécution dans le pire cas ?

Exercice 6. *Chebychev d'ordre supérieur*

L'inégalité de Chebyshev [2] utilise la variance d'une variable aléatoire pour borner son écart par rapport à l'espérance. On peut également utiliser des moments d'ordre supérieur.

1. Supposons que l'on ait une variable aléatoire X et un entier pair k pour lequel $\mathbf{E}\left[(X - \mathbf{E}[X])^k\right]$ est finie. Prouver que :
$$\mathbf{P}\left\{|X - \mathbf{E}[X]| > t\sqrt[k]{\mathbf{E}\left[(X - \mathbf{E}[X])^k\right]}\right\} \leq \frac{1}{t^k}.$$
2. Qu'est-ce qui nous empêche d'obtenir une inégalité similaire pour le cas où k est impair ? Trouver un contre exemple pour $k = 1$.

Exercice 7. *Fonctions génératrices*

Étant donnée une variable aléatoire discrète X à valeurs entières, on appelle *fonction génératrice de X* la fonction $G_X(z) := \mathbf{E}\left[z^X\right]$.

1. Donner la fonction génératrice sous forme de série entière.
   Que peut-on dire de $G_X(1)$, $G'_X(1)$ et $G''_X(1)$? Exprimer la variance à l'aide de G_X.
2. Soient X et Y deux variables aléatoires discrètes à valeur dans $\mathbb{N}$.
   Si X et Y sont indépendantes, que peut-on dire de G_{X+Y} ?

On considère maintenant X une variable aléatoire suivant une loi de Poisson pour $\lambda > 0$, c'est-à-dire telle que $\mathbf{P}\{X = k\} = \mathcal{C}(\lambda)\frac{\lambda^k}{k!}$.

3. Donner une autre expression pour $G_X(z)$.
4. Montrer que $\mathcal{C}(\lambda) = e^{-\lambda}$.
5. Calculer la fonction génératrice de X. En déduire $\mathbf{E}[X]$ et $\mathbf{Var}[X]$.
6. Reprendre la question précédente en supposant que X est une loi binomiale $\mathcal{B}(n, p)$.

Exercice 8. *Intégration*

Axel souhaite participer à un club de sa nouvelle école (un seul, pour des raisons de temps !). Pendant la semaine d'intégration, les n clubs proposent chacun une activité de découverte, dans un ordre aléatoire. Après chaque activité, Axel peut décider soit de s'inscrire à ce club (et de ne pas aller aux activités de découverte suivantes), soit de ne pas s'y inscrire et de continuer à découvrir des clubs (tout choix est définitif).

Bien sûr, Axel aimerait choisir le meilleur club. Iel décide d'utiliser la stratégie suivante : d'abord, participer à m activités, sans inscription ; puis, après la m-ème activité, s'inscrire au premier club qui lui plait strictement plus que tous ceux déjà découverts (on considère qu'il n'y a pas d'ex-aequo).

1. Montrer que la probabilité qu'Axel choisisse le meilleur club est
$$\mathsf{P}_{n,m} = \frac{m}{n}\sum_{j=m+1}^{n}\frac{1}{j-1}.$$
2. En déduire que $\lim_n \max_m \mathsf{P}_{n,m} \geq 1/e$.

2. Pafnuty Chebyshev, mathématicien russe (1821 – 1894).