

## TD 03 – Variables Aléatoires, Markov, Chebyshev et Chernoff (corrigé)

**Exercice 1.***Inégalité de Jensen*

Soit  $f$  une fonction convexe et  $X$  une variable aléatoire à valeurs réelles.

L'inégalité de Jensen<sup>1</sup> affirme :  $\mathbf{E}[f(X)] \geq f(\mathbf{E}[X])$ . En supposant que  $f$  soit  $\mathcal{C}^1$ , montrer cette inégalité.

☞ Soit  $\mu = \mathbf{E}[X]$ . Comme  $f$  est dérivable et convexe, on sait qu'elle est supérieure à sa tangente en  $\mu$ , i.e. pour tout  $x \in \text{Dom}(f)$ , on a :

$$f(x) \geq f(\mu) + f'(\mu)(x - \mu).$$

En prenant l'espérance des deux côtés, on obtient :

$$\begin{aligned} \mathbf{E}[f(X)] &\geq \mathbf{E}[f(\mu) + f'(\mu)(X - \mu)] && \text{(convexité de } f) \\ &= \mathbf{E}[f(\mu)] + f'(\mu)(\mathbf{E}[X] - \mu) && \text{(linéarité de l'espérance)} \\ &= f(\mu) + 0 && \text{car } \mu = \mathbf{E}[X] \\ &= f(\mathbf{E}[X]). \end{aligned}$$

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

- 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?

☞ Pour une coquille donne  $i$ , on a :

$$\begin{aligned} &\mathbf{P}\{\text{"La coquille numéro } i \text{ est corrigée en au plus } n \text{ relectures"}\} \\ &= 1 - \mathbf{P}\{\text{"Elle n'est pas corrigée après } n \text{ relectures"}\} \\ &= 1 - \left(\frac{2}{3}\right)^n. \end{aligned}$$

Comme les corrections sont indépendantes, on obtient :

$$\mathbf{P}\{\text{"Les 4 coquilles sont corrigées en au plus } n \text{ relectures"}\} = \left(1 - \left(\frac{2}{3}\right)^n\right)^4.$$

Enfin, puisque  $n$  est entier, on a :  $\left(1 - \left(\frac{2}{3}\right)^n\right)^4 \geq 0,9 \iff n \geq 10$  (presque 9, mais rigoureusement c'est 10).

- À quel problème vu en cours vous fait penser la situation précédente?

☞ C'est le problème du collectionneur de vignettes / coupons. La variable  $X_i$  comptant le nombre de corrections nécessaires pour corriger la coquille  $i$  suit une loi géométrique. La variable qui nous intéresse à la question précédente est  $\max(X_i)$ . La seule différence avec le problème du collecteur de coupons est qu'ici les variables  $X_i$  sont indépendantes (on peut corriger plusieurs coquilles dans la même relecture), alors qu'elles ne le sont pas dans le cas du collectionneur (on ne peut pas avoir plusieurs vignettes en un seul tirage).

- Soit  $X$  une variable aléatoire d'espérance  $\mu = 10$  et d'écart-type  $\sigma = 5$ . Montrer que pour tout  $n \geq 50$ , on a  $\mathbf{P}\{\mu - n < X < \mu + n\} \geq 0,99$ .

☞ On remarque que l'événement en question est égal à  $\{|X - \mu| < n\}$ . Or l'inégalité de Chebyshev nous donne ( $n \geq 0$  et  $\mu = 10$ ) :

$$\mathbf{P}\{|X - \mu| \geq n\} \leq \frac{\sigma^2}{n^2}$$

donc  $\mathbf{P}\{|X - 10| < n\} \geq 1 - \frac{25}{2500} = 0,99$ .

**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  $\{0, \dots, 2^k - 1\}$ . L'algorithme est le suivant :

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

- on effectue un pré-tri en répartissant selon certaines règles les  $n$  entiers dans  $n$  seaux ;
- on appelle un algorithme de tri simple (en temps quadratique, par exemple le tri par insertion) dans chaque seau ;
- 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  $j$ .

- 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).

☞ On suppose que les seaux sont numérotés de 0 à  $n-1$ . L'élément  $x$  peut être vu comme un nombre en binaire écrit sur  $k$  bits. On regarde les  $m$  bits de poids forts (ce qui revient à diviser par  $2^{k-m}$ ), cela nous donne un nombre  $s_x$  entre 0 et  $n-1$ . On met alors  $x$  dans le seau numéro  $s_x$ . Ainsi, on est sûr que si  $s_x < s_y$ , alors  $x < y$ , donc la condition d'exactitude du pré-tri est respectée.

- Soit  $X_i$  la v.a. comptant le nombre d'éléments dans le seau  $i$  après le pré-tri. Quelle loi suit  $X_i$  ?

☞  $X_i$  suit une loi binomiale de paramètres  $(n, \frac{1}{n})$ , car les  $n$  éléments d'entrée sont choisis indépendamment et uniformément et le pré-tri est effectué de façon à ce que chaque élément ait une probabilité  $\frac{1}{n}$  de se retrouver dans chacun des seaux.

- Prouver que la complexité en moyenne est  $\mathcal{O}(n)$ .

☞ Le pré-tri coûte un temps constant pour chaque élément, donc un temps linéaire en  $n$  au total. Ensuite, le tri du seau numéro  $i$  coûte  $c(X_i)^2$  pour une certaine constante  $c$ . Or,  $X_i$  suit une loi binomiale  $(n, \frac{1}{n})$ . En notant  $p = \frac{1}{n}$ , on a donc :

$$\mathbb{E}[X_i^2] = \mathbb{E}[X_i]^2 + \text{Var}[X_i] = (np)^2 + np(1-p) = np(np+1-p) = 1 \cdot (1+1-\frac{1}{n}) = 2 - \frac{1}{n} < 2.$$

Donc l'espérance du temps passé dans la deuxième étape est au plus

$$\mathbb{E}\left[\sum_{i=0}^{n-1} c(X_i)^2\right] = c \times \sum_{i=0}^{n-1} \mathbb{E}[X_i^2] \leq 2cn.$$

L'espérance du temps d'exécution totale est donc linéaire en  $n$ .

#### 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 :

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

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 :

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

☞ L'algorithme  $\mathcal{B}$  va appeler  $N$  fois l'algorithme  $\mathcal{A}$ , et renvoyer la réponse majoritaire (choisie arbitrairement en cas d'égalité). Il faut donc choisir  $N$  qui vérifie les conditions. On note  $X_1, \dots, X_N$  les réponses successives de  $\mathcal{A}$ . La réponse renvoyée par  $\mathcal{B}$  apparaît donc au moins  $N/2$  fois. Posons  $n = |x|$ . 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  $p = \mathbf{P}\{\mathcal{A}(x) = 1\}$  (comme  $x \notin L$ , ici on a donc  $p \leq \frac{1}{3}$ ). Posons  $X = \sum_{i=1}^N X_i$ . Alors  $\mathbb{E}[X] = pN \leq \frac{N}{3}$ . On applique la variante de Chernoff (avec  $\mu_H \geq \mathbb{E}[X]$ ) en prenant  $\delta = 1/2$  et  $\mu_H = \frac{N}{3}$ , de telle sorte que  $(1+\delta) \times \mu_H = \frac{N}{2}$ .

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

- Maintenant pour  $x \in L$ . Avec les mêmes notations, on a  $\mathbb{E}[X] \geq \frac{3N}{4}$  donc on applique la variante de Chernoff avec  $\mu_L = \frac{3N}{4}$  et  $\delta = \frac{1}{3}$ , de telle sorte que  $(1-\delta) \times \mu_L = \frac{N}{2}$ . On obtient :

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

Donc la probabilité que  $\mathcal{B}$  se trompe est bornée par  $\max(e^{-\frac{N}{24}}, e^{-\frac{N}{36}}) = e^{-\frac{N}{36}}$ .

En prenant  $N = \frac{36n}{\log e}$ , on obtient que la probabilité que l'algorithme  $\mathcal{B}$  se trompe est au plus  $2^{-n}$ .

**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.

 Le but ici est d'utiliser l'inégalité de Markov. Soit  $X$  le temps d'exécution de l'algorithme.

$$\mathbf{P}\{X \geq n^2 \cdot f(n)\} \leq \frac{\mathbf{E}[X]}{n^2 \cdot f(n)} \leq \frac{c \cdot n^2}{n^2 \cdot f(n)} \leq \frac{c}{f(n)} \xrightarrow{n \rightarrow \infty} 0.$$

2. Que pouvons nous en déduire sur le temps d'exécution dans le pire cas ?

 Pour avoir une borne supérieure sur le temps dans le pire cas, on utilise le fait que les entrées sont distribuées uniformément. Comme chaque entrée est choisie avec probabilité  $1/2^n$ , on a que si  $\mathbf{P}\{X \geq t\}$  est non nulle, elle doit être au moins égale à  $1/2^n$  (car au moins une entrée donnera un temps de calcul supérieur à  $t$ ). On a vu à la question précédente que

$$\mathbf{P}\{X \geq n^2 \cdot f(n)\} \leq \frac{c}{f(n)}.$$


Pour que cette quantité soit inférieure à  $1/2^n$ , il faut que  $f(n) \geq c2^n$ . On en déduit que le temps d'exécution dans le pire cas est borné par  $cn^2 2^n = O(n^2 2^n)$ .

**Exercice 6.***Chebyshev d'ordre supérieur*

L'inégalité de Chebyshev<sup>2</sup> 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  $\mathbf{E}[(X - \mathbf{E}[X])^k]$  est finie. Prouver que :

$$\mathbf{P}\left\{|X - \mathbf{E}[X]| > t \sqrt[k]{\mathbf{E}[(X - \mathbf{E}[X])^k]}\right\} \leq \frac{1}{t^k}.$$

 Soit  $Y = (X - \mathbf{E}[X])^k$ . Supposons dans un premier temps que  $X$  n'est pas constante. Alors  $Y$  n'est pas constante égale à 0, et comme  $Y$  est positive, on en déduit que  $\mathbf{E}[Y] \neq 0$ . Par l'inégalité de Markov, on a :

$$\mathbf{P}\{Y \geq t^k \mathbf{E}[Y]\} \leq \frac{\mathbf{E}[Y]}{t^k \mathbf{E}[Y]} = \frac{1}{t^k}.$$

Or, en passant à la racine  $k$ -ième, on obtient :

$$\mathbf{P}\{Y \geq t^k \mathbf{E}[Y]\} = \mathbf{P}\{\sqrt[k]{Y} \geq t \sqrt[k]{\mathbf{E}[Y]}\} = \mathbf{P}\{|X - \mathbf{E}[X]| \geq t \sqrt[k]{\mathbf{E}[(X - \mathbf{E}[X])^k]}\}$$

où la deuxième égalité est vraie car  $k$  est pair. En recombinaison des deux relations que l'on a trouvées, on obtient ce qui est demandé :

$$\mathbf{P}\left\{|X - \mathbf{E}[X]| > t \sqrt[k]{\mathbf{E}[(X - \mathbf{E}[X])^k]}\right\} \leq \frac{1}{t^k}.$$

Supposons maintenant que  $X$  est constante, alors  $X = \mathbf{E}[X]$ , et  $\mathbf{P}\{|X - \mathbf{E}[X]| > t \sqrt[k]{\mathbf{E}[(X - \mathbf{E}[X])^k]}\} = 0$ , et le résultat est toujours vrai.

2. Qu'est-ce qui nous empêche d'obtenir une inégalité similaire pour le cas où  $k$  est impair ? Trouver un contre exemple pour  $k = 1$ .

 Si  $k$  est impair, alors  $(X - \mathbf{E}[X])^k$  peut prendre des valeurs négatives, ce qui nous empêche d'appliquer l'inégalité de Markov.

Pour le contre exemple, on peut prendre  $X = \pm 1$  avec probabilité  $1/2$  pour chaque. On a alors  $\mathbf{E}[X] = 0$  et  $\mathbf{E}[X - \mathbf{E}[X]] = 0$ . Ainsi, pour tout  $t > 0$ , on a  $t \sqrt[k]{\mathbf{E}[(X - \mathbf{E}[X])^k]} = 0$ . En particulier, pour  $t = 3$ , on a  $\mathbf{P}\{|X - \mathbf{E}[X]| > 3 \sqrt[k]{\mathbf{E}[(X - \mathbf{E}[X])^k]}\} = \mathbf{P}\{|X| > 0\} = 1 \not\leq \frac{1}{3}$ .

**Exercice 7.***Fonctions génératrices*

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

---

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

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  $G_X$ .

☞ Par définition de l'espérance, on a

$$G_X(z) := \mathbb{E}[z^X] = \sum_{k \geq 0} z^k \mathbf{P}\{X = k\}.$$

On a en plus les propriétés :

- $G_X(1) = 1$ ,
- le rayon  $R$  de convergence de cette série est donc supérieur ou égal à 1,
- $G'_X(1) = \mathbb{E}(X)$  (dans le cas où  $R > 1$ ),
- $G''_X(1) = \mathbb{E}(X(X-1))$  (dans le cas où  $R > 1$ ),
- $V(X) = G''_X(1) + G'_X(1) - G'_X(1)^2$  (dans le cas où  $R > 1$ ).

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  $G_{X+Y}$ ?

☞ Si  $X, Y$  sont indépendantes et à valeur dans  $\mathbb{N}$ , on a  $G_{X+Y} = G_X G_Y$ .

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  $G_X(z)$ .

☞ On calcule :

$$\begin{aligned} G_X(z) &= \sum_{k \geq 0} z^k \mathbf{P}\{X = k\} \\ &= \mathcal{C}(\lambda) \sum_{k \geq 0} \frac{(\lambda z)^k}{k!} \\ &= \mathcal{C}(\lambda) \exp \lambda z. \end{aligned}$$

4. Montrer que  $\mathcal{C}(\lambda) = e^{-\lambda}$ .

☞  $X$  est une variable aléatoire, on a donc  $\sum_{k \geq 0} \mathbf{P}\{X = k\} = 1$ . Or  $G_X(1) = \sum_{k \geq 0} \mathbf{P}\{X = k\} = \mathcal{C}(\lambda) \exp \lambda$ , d'où le résultat  $\mathcal{C}(\lambda) = \exp -\lambda$ .

5. Calculer la fonction génératrice de  $X$ . En déduire  $\mathbb{E}[X]$  et  $\mathbf{Var}[X]$ .

☞ La fonction génératrice de  $X$  est donc  $G_X(z) = \exp \lambda(z-1)$ . On a en toute généralité :

$$G'_X(z) = \sum_{k \geq 1} k z^{k-1} \mathbf{P}\{X = k\} \quad \text{et} \quad G'_X(1) = \mathbb{E}[X],$$

$$G''_X(z) = \sum_{k \geq 2} k(k-1) z^{k-2} \mathbf{P}\{X = k\} \quad \text{et} \quad G''_X(1) = \mathbb{E}[X^2] - \mathbb{E}[X].$$

Ainsi,

- $\mathbb{E}[X] = G'_X(1) = \lambda$ ,
- $\mathbf{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = G''_X(1) + G'_X(1) - G'_X(1)^2 = \lambda^2 + \lambda - \lambda^2 = \lambda$ .

6. Reprendre la question précédente en supposant que  $X$  est une loi binomiale  $\mathcal{B}(n, p)$ .

☞ On suppose maintenant que  $X$  suit une loi binomiale de paramètre  $(n, p)$ . On calcule sa fonction génératrice :

$$\begin{aligned} G_X(z) &= \sum_{k \geq 0} z^k \mathbf{P}\{X = k\} \\ &= \sum_{k \geq 0} z^k \binom{n}{k} p^k (1-p)^{n-k} \\ &= (pz + (1-p))^n \end{aligned}$$

et ses deux premières dérivées successives :

$$G'_X(z) = np(pz + 1 - p)^{n-1} \quad \text{et} \quad G''_X(z) = n(n-1)p^2(pz + 1 - p)^{n-2}$$

On en déduit alors facilement son espérance et sa variance :

$$\mathbb{E}[X] = np \quad \text{et} \quad \mathbf{Var}[X] = np(1-p).$$

**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. Ici 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

$$P_{n,m} = \frac{m}{n} \sum_{j=m+1}^n \frac{1}{j-1}.$$

 Soit  $X_j$  l'évènement "le  $j$ -ième club est le meilleur", alors pour tout  $j$  on a  $\mathbf{P}\{X_j\} = \frac{1}{n}$ .

Soit  $E_j$  l'évènement "le  $j$ -ième club est le meilleur et est choisi". Si  $j \leq m$ ,  $\mathbf{P}\{E_j\} = 0$ . Sinon, si  $j > m$ , alors  $\mathbf{P}\{E_j\} = \frac{1}{n} \cdot \frac{m}{j-1}$ . En effet, le meilleur club est en  $j$ -ième position avec probabilité  $\frac{1}{n}$ , et il est choisi si parmi les  $j-1$  clubs vus précédemment, le meilleur est parmi les  $m$  premiers clubs, ce qui arrive avec probabilité  $\frac{m}{j-1}$ .

2. En déduire que  $\lim_n \max_m P_{n,m} \geq 1/e$ .

 Les bornes se trouvent par des calculs d'intégrales :

$$\frac{m}{n} (\ln(n) - \ln(m)) \leq P(n, m) \leq \frac{m}{n} (\ln(n-1) - \ln(m-1))$$

Le maximum de la fonction  $\frac{\ln(x)}{x}$  est atteint pour  $x = e$ . Donc on prend  $m = \frac{n}{e}$ , et on obtient  $\lim_n \max_m P(n, m) \geq 1/e$ .