TD 04 – Moments et Inégalités de concentration (corrigé)

Exercice 1. *Running Time*

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 zero 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.

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

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 nul, elle doit être au moins égale à $1/2^n t$ (car au moins une entrée donnera un temps de calcul supérieur à t). On a vu à la question précédente que

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

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 2. *Coquilles dans un TD*

1. Une feuille de TD contient 4 coquilles. À chaque relecture, une coquille non corrigée est corrigée avec probabilité $1/3$. Les relectures sont indépendantes les unes des autres. Combien de relectures faut-il faire au minimum pour que la probabilité qu'il ne subsiste aucune coquille soit supérieure à 0,9 ?

☞ On a : $\mathbf{P}\{$"La coquille numéro i est corrigée en au plus n relectures"$\}$
$= 1 - \mathbf{P}\{$"Elle n'est pas corrigée après n relectures"$\} = 1 - (\frac{2}{3})^n$.

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

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

2. À quel autre problème vu en cours vous fait penser la situation précédente ?

☞ C'est le problème du collectionneur de vignettes. 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'on ne peut pas avoir plusieurs vignettes avec une vignette).

3. Soit X une variable aléatoire d'espérance $\mu = 10$ et d'écart-type $\sigma = 5$. Montrer que pour tout $n \geq 50$, la probabilité que de l'événement $\{10 - n < X < 10 + n\}$ est au moins égale à 0,99.

☞ On remarque que l'événement en question est égal à $\{|X - 10| < n\}$. 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. *Chebychev d'ordre supérieur*

L'inégalité de Chebyshev 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 :

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

☞ 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,

$$\mathbf{P}\left\{Y\ge t^k\mathbf{E}[Y]\right\}\le\frac{\mathbf{E}[Y]}{t^k\mathbf{E}[Y]}=\frac{1}{t^k}\,.$$

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

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

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

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

Supposons maintenant que X est constante, alors $X=\mathbf{E}[X]$, et $\mathbf{P}\left\{|X-\mathbf{E}[X]|>t\sqrt[k]{\mathbf{E}[(X-\mathbf{E}[X])^k]}\right\}=0$, donc 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 $t\sqrt[1]{\mathbf{E}[(X-\mathbf{E}[X])^1]}=0$. En particulier, pour $t=3$, on a $\mathbf{P}\left\{|X-\mathbf{E}[X]|>3\sqrt[1]{\mathbf{E}[(X-\mathbf{E}[X])^1]}\right\}=\mathbf{P}\{|X|>0\}=1\not\le\frac{1}{3}$.

Exercice 4. *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 $X\ge n/4$.

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).

☞ La variable X est une somme de n variables de Bernouilli de paramètre $1/6$. On a donc $\mathbf{E}[X]=\frac{n}{6}$ et $\mathbf{Var}[X]=\frac{5n}{36}$. On obtient :

Markov : $\mathbf{P}\{X\ge n/4\}\le\frac{\mathbf{E}[X]}{n/4}=2/3$.

Chebyshev : $\mathbf{P}\left\{X\ge\frac{n}{4}\right\}=\mathbf{P}\left\{X-\frac{n}{6}\ge\frac{n}{12}\right\}\le\frac{\mathbf{Var}[X]}{(n/12)^2}=\frac{20}{n}$.

Chernoff (en utilisant la variante #2 où les X_i sont des variable aléatoires indépendantes à valeurs dans $[0,1]$: $\mathbf{P}\{X\ge(1+\varepsilon)\mathbf{E}[X]\}\le\exp\left(-\frac{\varepsilon^2}{2+\varepsilon}\mathbf{E}[X]\right)$) :

$$\mathbf{P}\left\{X\ge\frac{n}{4}\right\}=\mathbf{P}\left\{X\ge\left(1+\frac{1}{2}\right)\mathbf{E}[X]\right\}\le\exp\left(-\frac{n}{60}\right)\,.$$

Exercice 5. *Fonction Génératrice*

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

1. Dîtes des choses intelligentes sur la fonction génératrice d'une variable aléatoire discrète (ex : donnez la fonction génératrice sous forme de série entière, une formule de l'espérance de X, de sa variance...).

☞ Par définition de l'espérance, $G_X(z):=\mathbb{E}(z^X)=\sum_{k\ge 0}z^k\mathbb{P}(X=k)$. 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 $R>1$)
- Si X, Y sont indépendantes, à valeur dans $\mathbb{N}$, $G_{X+Y}=G_XG_Y$.

Soit X une variable aléatoire suivant une loi de Poisson : $\mathbb{P}(X=k)=\mathcal{C}(\lambda)\frac{\lambda^k}{k!}$, avec $\lambda>0$.

2. Donner une autre expression pour $G_X(z)$.
☞

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

3. 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 $G_X(1) = \sum_{k\geq 0}\mathbf{P}\{X=k\} = C(\lambda)\exp\lambda$, d'où le résultat $C(\lambda) = \exp -\lambda$.

4. Calculer la fonction génératrice de X. En déduire $\mathbf{E}[X]$ et $\mathbf{Var}[X]$.
☞ La fonction génératrice de X est donc $G_X(z) = \exp\lambda(z-1)$. On a en toute généralité :

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

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

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

5. 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 :

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

On en déduit alors facilement son espérance et sa variance :

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

Exercice 6. *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 ?
☞ Soit $X_i = 1$ si le i-ième lancer donne pile, 0 sinon. Soit $\overline{X_n} = \frac{1}{n}\sum_{i=1}^n X_i$. On veut approximer p par $\overline{X_n}$, en choisissant soigneusement n. Notre but est donc d'obtenir :

$$\mathbf{P}\{|\overline{X_n} - p| \leq 0.1\} \geq 0.9$$

ou, autrement dit, $\mathbf{P}\{|\overline{X_n} - p| \geq 0.1\} \leq 0.1$.
Or, $\mathbf{E}[X_i] = p$ et $\mathbf{Var}[X_i] = p(1-p)$ car il s'agit de variables de Bernouilli, donc $\mathbf{E}[\overline{X_n}] = np/n = p$ et $\mathbf{Var}[\overline{X_n}] = \frac{1}{n^2}\sum_{i=1}^n \mathbf{Var}[X_i] = \frac{p(1-p)}{n}$.
Donc, en utilisant l'inégalité de Chebyshev,

$$\begin{aligned} \mathbf{P}\{|\overline{X_n} - p| \geq 0.1\} &\leq \frac{\mathbf{Var}[X]}{0.1^2} \\ &\leq \frac{p(1-p)}{0.01n} \\ &\leq \frac{100}{n}\frac{1}{4} \text{ car } p(1-p) \leq 1/4 \\ &\leq \frac{25}{n} \end{aligned}$$

Il suffit donc d'avoir $25/n \leq 0.1$, autrement dit $n \geq 250$.

Exercice 7. *Probabilités conditionnelles*

Soit Y une variable aléatoire prenant des valeurs entières, positives ou nulles, et dont l'espérance est strictement positive. Le but de cet exercice est de prouver la relation suivante :

$$\frac{\mathbf{E}[Y]^2}{\mathbf{E}[Y^2]} \leq \mathbf{P}\{Y \neq 0\} \leq \mathbf{E}[Y] \ .$$

1. On voudrait une variable aléatoire X qui corresponde informellement à $(Y|Y \neq 0)$. Comment la définir proprement (on pourra changer d'espace de probabilité) ?

   ☞ Soit Ω l'espace de probabilité sur lequel Y est défini. On considère $\Omega' = \Omega \setminus \{\omega | Y(\omega) = 0\}$, avec $\mathbf{P'}(\omega) = \mathbf{P}\{\omega\} / \mathbf{P}\{Y \neq 0\}$. On défini X sur Ω' par $X(\omega) = Y(\omega)$. Alors on a bien la relation voulue :

   $$\mathbf{P'}(X = k) = \mathbf{P}\{Y = k\} / \mathbf{P}\{Y \neq 0\} = \mathbf{P}\{Y = k | Y \neq 0\} \ .$$

2. Comparer $\mathbf{E}[X]^2$ et $\mathbf{E}\left[X^2\right]$.

   ☞ En utilisant par exemple Jensen, comme $x \to x^2$ est convexe, on a : $(\mathbf{E}[X])^2 \leq \mathbf{E}\left[X^2\right]$.

3. Conclure.

   ☞ En écrivant les choses selon la définition de l'espérance (avec la somme sur toutes les valeurs possibles), on obtient : $\mathbf{E}[X] = \mathbf{E}[Y] / \mathbf{P}\{Y \neq 0\}$ et $\mathbf{E}\left[X^2\right] = \mathbf{E}\left[Y^2\right] / \mathbf{P}\{Y \neq 0\}$. En réinjectant dans l'inégalité de la question précédente :

   $$\begin{aligned}(\mathbf{E}[X])^2 \leq \mathbf{E}\left[X^2\right] &\Leftrightarrow \frac{\mathbf{E}[Y]^2}{\mathbf{P}\{Y \neq 0\}^2} \leq \frac{\mathbf{E}\left[Y^2\right]}{\mathbf{P}\{Y \neq 0\}} \\ &\Leftrightarrow \frac{\mathbf{E}[Y]^2}{\mathbf{E}[Y^2]} \leq \mathbf{P}\{Y \neq 0\}\end{aligned}$$

   Ce qui donne la borne inférieure. La borne supérieure s'obtient facilement avec Markov car $\mathbf{P}\{Y \neq 0\} = \mathbf{P}\{Y \geq 1\} \leq \mathbf{E}[Y] / 1$.

Exercice 8. *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 $X_i \in [0,1]$.

☞ La fonction $x \to e^{\lambda x}$ est convexe donc pour $x \in [0,1]$, $e^{\lambda x} \leq (1-x)e^0 + xe^\lambda = (1-x) + xe^\lambda$.
En particulier,

$$\mathbf{E}\left[e^{\lambda X}\right] \leq \mathbf{E}\left[(1-X) + Xe^\lambda\right] = (1-p) + pe^\lambda = \mathbf{E}\left[e^{\lambda Y}\right] \ .$$

Maintenant, on a tous les outils pour reprendre la preuve de la borne de Chernoff : soit X_i des variables aléatoires indépendantes à valeurs dans $[0,1]$ avec $\mathbf{E}[X_i] = p_i$. Alors

$$\mathbf{E}\left[e^{\lambda X_i}\right] \leq (1-p_i) + p_i e^\lambda = 1 + p_i(e^\lambda - 1) \leq e^{p_i(e^\lambda - 1)} \ .$$

Soit $X = \sum X_i$ et $\mu = \mathbf{E}[X] = \sum_{i=1}^n p_i$. Alors de même que dans la preuve originale,

$$\mathbf{E}\left[e^{\lambda X}\right] = \prod_i \mathbf{E}\left[e^{\lambda X_i}\right] \leq \prod_i e^{p_i(e^\lambda - 1)} = e^{\mu(e^\lambda - 1)}$$

On applique Markov à $e^{\lambda X}$:

$$\begin{aligned}\mathbf{P}\{X \geq (1+\delta)\mu\} &= \mathbf{P}\left\{e^{\lambda X} \geq e^{\lambda(1+\delta)\mu}\right\} \\ &\leq \frac{\mathbf{E}\left[e^{\lambda X}\right]}{e^{\lambda(1+\delta)\mu}} \\ &\leq \frac{e^{\mu(e^\lambda - 1)}}{e^{\lambda(1+\delta)\mu}}\end{aligned}$$

En posant $\lambda = \ln(1+\delta) > 0$, on obtient

$$\mathbf{P}\{X \geq (1+\delta)\mu\} \leq \left(\frac{e^\delta}{(1+\delta)^{1+\delta}}\right)^\mu \ .$$