TD 05 (corrigé)

Exercice 1.

Fonctions de répartition

Définitions : Soit une variable aléatoire réelle X de densité de probabilité f_X,

— la **fonction de répartition** associée est :

$$F_X(x) \;=\; \int_{-\infty}^{x} f_X(t)\mathrm{d}t\,,$$

— l'**espérance** de X est définie par

$$\mathbf{E}\left[X\right] \;=\; \int_{-\infty}^{+\infty} t f_X(t)\mathrm{d}t$$

(si l'intégrale est absolument convergente),

— la **variance** de X est définie par

$$\mathbf{Var}\left[X\right] \;=\; \mathbf{E}\left[(X - \mathbf{E}\left[X\right])^2\right] \;=\; \mathbf{E}\left[X^2\right] - \mathbf{E}\left[X\right]^2$$

(si $\mathbf{E}\left[X^2\right]$ existe).

1. Donner la densité, la fonction de répartition, l'espérance et la variance d'une variable aléatoire U suivant la loi uniforme sur $[a,b]$ pour $a < b$.

☞ Densité : $f_U = \frac{1}{b-a}\mathbb{1}_{[a,b]}$,
Fonction de répartition : $F_U(x) = \frac{x-a}{b-a}$ sur $[a,b]$ et 0 ailleurs,
Espérance : $\mathbf{E}\left[U\right] = \frac{a+b}{2}$,
Variance : $\mathbf{Var}\left[U\right] = \frac{(b-a)^2}{12}$.

Soit U une v.a. de loi uniforme sur $[0,2]$. Soit $X := \sqrt{U}$.

2. Calculer la fonction de répartition F_X de X.

☞

$$\begin{aligned} F_X(z) &= \mathbf{P}\left\{X \leq z\right\} = \mathbf{P}\left\{\sqrt{U} \leq z\right\} = \mathbf{P}\left\{U \leq z^2\right\} = F_U\left(z^2\right) \\ &= \begin{cases} \frac{z^2}{2} & \text{si } z \in [0,\sqrt{2}] \\ 1 & \text{si } z \geq \sqrt{2} \\ 0 & \text{si } z < 0 \end{cases} \end{aligned}$$

3. Calculer la densité f_X de X.

☞ $f_X(x) = F_X'(x) = x$ sur $[0,\sqrt{2}]$.

4. Même questions pour $Y := 1/U$.

☞ $F_Y(y) = \mathbf{P}\left\{Y \leq y\right\} = \mathbf{P}\left\{\frac{1}{U} \leq y\right\} = \mathbf{P}\left\{U \geq \frac{1}{y}\right\} = 1 - \mathbf{P}\left\{U \leq \frac{1}{y}\right\}$

$$F_Y = \begin{cases} 0 & \text{si } y \leq \frac{1}{2} \\ 1 - \frac{1}{2y} & \text{si } y \geq \frac{1}{2} \end{cases}$$

$f_y(y) = \frac{1}{2y^2}$ sur $[\frac{1}{2}, +\infty[$.

5. Quelle est l'espérance de Y ?

☞ $E(Y) = \int_{\frac{1}{2}}^{\infty} y f_Y(y) dy = \int_{\frac{1}{2}}^{\infty} \frac{1}{2y} dy = +\infty$

Exercice 2.

Records

Soient $U_1, \ldots, U_n$ des variables aléatoires indépendantes, uniformément distribuées sur $[0,1]$.
Pour $i \in \{1, \ldots, n\}$, on dit que U_i est un **record** si pour tout $j \leq i$, on a $U_i \leq U_j$.
Calculer l'espérance du nombre de records dans la suite $U_1, \ldots, U_n$.

☞ Soit X_i la v.a. valant 1 si U_i est un record, 0 sinon. Alors :

$$\mathbf{P}\{X_i = 1\} = \mathbf{P}\left\{\bigcap_{j<i} U_i \leq U_j\right\} = \int_{u_i=0}^{1} (1-u_i)^{i-1} du_i = \left[\frac{-(1-u_i)^i}{i}\right]_0^1 = \frac{1}{i}\,.$$

Puis, soit $X = \sum_{i=1}^n X_i$ le nombre de records. Alors :

$$\mathbf{E}[X] = \sum_{i=1}^n \mathbf{E}[X_i] = \sum_{i=1}^n \mathbf{P}\{X_i = 1\} = \sum_{i=1}^n \frac{1}{i} = \mathcal{O}(\log n)\,.$$

Exercice 3. *Interrupteurs*

1. Montrer qu'il existe une constante $\gamma > 0$ rendant l'énoncé suivant vrai :

   « Si une v.a. positive X vérifie $\mathbf{E}[X] = 1$ et $\mathbf{E}\left[X^2\right] \leq 3$, alors $\mathbf{P}\{X \geq 1/4\} \geq \gamma$. »

   Indication : définir la variable aléatoire $Y = \mathbf{1}_{X \geq 1/4}$ *et se ramener à l'inégalité de Cauchy-Schwarz.* $\mathbb{E}(XY) \leq \sqrt{\mathbb{E}(X^2)\mathbb{E}(Y^2)}$

   ☞ On écrit

   $$1 = \mathsf{E}[X] = \mathsf{E}[X\mathbf{1}_{X<1/4}] + \mathsf{E}[X\mathbf{1}_{X\geq 1/4}] \leq \frac{1}{4} + \mathsf{E}[X\mathbf{1}_{X\geq 1/4}].$$

   Par l'inégalité de Cauchy–Schwarz, $\mathsf{E}[X\mathbf{1}_{X\geq 1/4}] \leq \sqrt{\mathsf{E}[X^2]\mathsf{P}(X \geq 1/4)} \leq \sqrt{3}\sqrt{\mathsf{P}(X \geq 1/4)}$. On obtient la minoration voulue pour $\gamma = 3/16$.

2. Soient $(X_1, \ldots, X_n)$ des v.a. i.i.d. vérifiant $\mathbf{P}\{X_i = 1\} = \mathbf{P}\{X_i = -1\} = \frac{1}{2}$.
   On pose $Y = \frac{1}{\sqrt{n}}(X_1 + \cdots + X_n)$. Calculer $\mathbf{E}\left[Y^2\right]$ et $\mathbf{E}\left[Y^4\right]$ et en déduire que :

   $$\mathbf{E}\left[|X_1 + \cdots + X_n|\right] \geq \frac{\gamma}{2}\sqrt{n}.$$

   ☞ On a $\mathsf{E}[Y^2] = \frac{1}{n} \cdot \mathbf{Var}[Y] = \frac{1}{n} \cdot \sum_i \mathbf{Var}[X_i] = 1$ (par inédependance). On a ensuite

   $$\mathsf{E}[Y^4] = \frac{1}{n^2} \cdot \sum_{i,j,k,l=1}^{n} \mathsf{E}[X_i X_j X_k X_l].$$

   L'indépendance des X_i et le fait que $\mathsf{E}[X_i] = 0$ implique $\mathsf{E}[X_i X_j X_k X_l] = 0$ dès qu'un indice apparaît une unique fois parmi $\{i,j,k,l\}$. Les seuls termes non nuls sont ceux où $i=j=k=l$ ou $i=j\neq k=l$ ou $i=k\neq j=l$ ou $i=l\neq j=k$. On a donc

   $$\mathsf{E}[Y^4] = 1/n^2(n + 3n(n-1)) = 3 - 2/n \leq 3.$$

   On applique la question précédente à $X = Y^2$, d'où $\mathsf{P}(Y^2 \geq 1/4) = \mathsf{P}(|X_1 + \cdots + X_n| \geq \frac{\sqrt{n}}{2}) \geq \gamma$. Enfin,

   $$\mathsf{E}[|X_1 + \cdots + X_n|] \geq \frac{\sqrt{n}}{2}\mathsf{P}\left(|X_1 + \cdots + X_n| \geq \frac{\sqrt{n}}{2}\right) \geq \frac{\gamma\sqrt{n}}{2}.$$

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 $a_{ij} \times b_i \times c_j = 1$. On considère la quantité

$$\mathbf{F}(a,b,c) = \sum_{i,j=1}^{n} a_{ij} b_i c_j$$

qui est le nombre d'ampoules allumées moins le nombre d'ampoules éteintes.
Deux joueuses jouent au jeu suivant :

1. la joueuse 1 choisit la position des interrupteurs (a_{ij}),
2. puis la joueuse 2 choisit la position des interrupteurs (b_i) et (c_j).

La joueuse 1 veut minimiser $\mathbf{F}(a,b,c)$ et la joueuse 2 veut le maximiser. On considère donc :

$$\mathbf{V}(n) = \min_{a\in\{-1,1\}^{n\times n}} \max_{b,c\in\{-1,1\}^n} \mathbf{F}(a,b,c).$$

3. Montrer que $\mathbf{V}(n) = \mathcal{O}(n^{3/2})$ en considérant le cas où la joueuse 1 joue au hasard.

☞ Soit $(a_{ij})_{1\le i,j\le n}$ des v.a. i.i.d. de loi uniforme sur $\{-1,1\}$. Quel que soit le choix de b et c, on a

$$\mathsf{P}(F(a,b,c) \ge t) \le \exp(-t^2/2n^2)$$

par l'inégalité de Chernoff (en effet, $F(a,b,c)$ est la somme de n^2 v.a. de loi uniforme sur $\{-1,1\}$). Par la borne de l'union,

$$\mathsf{P}(\max_{b,c} F(a,b,c) \ge t) \le 4^n \exp(-t^2/2n^2).$$

Lorsque $t > \sqrt{2n^3 \log 4}$, cette probabilité est < 1 et donc $\mathsf{P}(\max_{b,c} F(a,b,c) < t) > 0$: il existe donc un choix de a tel que $\max_{b,c} F(a,b,c) < t$, d'où $V(n) = O(n^{3/2})$.

4. La joueuse 2 applique la stratégie suivante : elle 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 *(indication : utiliser la question 2)* et en déduire que $\mathbf{V}(n) = \Omega(n^{3/2})$.

☞ Fixons $a = (a_{i,j})$ et choisissons (b_i) i.i.d. de loi uniforme sur $\{-1,1\}$. On a alors

$$\max_c F(a,b,c) = \sum_{i=1}^n \left| \sum_{j=1}^n a_{ij} b_j \right|.$$

En utilisant la linéarité de l'espérance, le fait que $(b_j)_j$ et $(a_{ij}b_j)_j$ ont même loi et la question I.2, il vient

$$\mathsf{E} \max_c F(a,b,c) = n\mathsf{E} \left| \sum_{j=1}^n b_j \right| \ge \frac{n^{3/2}\gamma}{2}.$$

En particulier, pour tout choix de a, il existe b tel que $\max_c F(a,b,c) \ge \frac{n^{3/2}\gamma}{2}$.

Exercice 4. *Arrondi*

Soit U un ensemble à n éléments. On appelle recouvrement de U un ensemble $\mathcal{S} = (S_1, \cdots, S_m)$ de parties de U qui vérifie $\bigcup S_i = U$. Étant donné $\mathcal{S}$ un recouvrement de U, on note $\mathrm{OPT}(\mathcal{S})$ le cardinal minimal d'un sous-ensemble de $\mathcal{S}$ qui est encore un recouvrement de U.

1. Expliquer rapidement pourquoi $\mathrm{OPT}(\mathcal{S})$ est la solution du problème d'optimisation suivant :

$$\text{Minimiser } \sum_{i=1}^m x_i \text{ sous les contraintes } x_i \in \{0,1\} \text{ et } \sum_{i=1}^m x_i \mathbf{1}_{S_i} \ge \mathbf{1}_U \tag{1}$$

☞ Les sous-ensembles de $\mathcal{S}$ sont en bijections avec $\{0,1\}^m$. On a donc, pour tout sous-ensemble $\mathcal{T} \subseteq \mathcal{S}$, $\#\mathcal{T} = \sum x_i$ avec x_i qui vaut 1 si et seulement si $S_i \in \mathcal{T}$.

De plus, $\mathcal{T}$ est un recouvrement de U si et seulement si pour tout $y \in U$, on a $\mathbf{1}_{\mathcal{T}}(y) = 1 = \mathbf{1}_U(y)$. Or, $\mathbf{1}_{\mathcal{T}}(y) = 1$ si et seulement si $\sum_{i=1}^m x_i \mathbf{1}_{S_i}(y) \ge 1$ (car y doit être dans au moins un S_i qui est gardé dans $\mathcal{T}$, et il peut être dans plusieurs S_i différents).

On considère le problème suivant qui est une relaxation de (1) :

$$\text{Minimiser } \sum_{i=1}^m z_i \text{ sous les contraintes } z_i \in [0,1] \text{ et } \sum_{i=1}^m z_i \mathbf{1}_{S_i} \ge \mathbf{1}_U \tag{2}$$

Alors que le problème (1) est NP-difficile, le problème (2) peut être résolu en temps polynomial par les méthodes de programmation linéaire.

2. Soit k un entier, et $(z_i)_{1\le i\le m}$ qui minimisent (2). Soient $(X_{i,j})_{1\le i\le m, 1\le j\le k}$ des variables aléatoires indépendantes vérifiant $\mathbf{P}\left\{X_{i,j} = 1\right\} = z_i$, $\mathbf{P}\left\{X_{i,j} = 0\right\} = 1 - z_i$. On définit un sous-ensemble $\mathcal{T} \subset \mathcal{S}$ par la condition :

$$S_i \in \mathcal{T} \iff \exists j \in \{1, \dots, k\} \,:\, X_{i,j} = 1\,.$$

Montrer que :

$$\mathbf{E}\left[\#\mathcal{T}\right] \le k\, \mathrm{OPT}(\mathcal{S})\,.$$

☞ Par linéarité de l'espérance, on a :

$$\begin{aligned}
\mathbf{E}\left[\#\mathcal{T}\right] &= \sum_{i=1}^{m} \mathbf{P}\left\{S_i \in \mathcal{T}\right\} \\
&= \sum_{i=1}^{m} \mathbf{P}\left\{X_{i,1} + \ldots + X_{i,k} \geq 1\right\} \\
&\leq \sum_{i=1}^{m} \mathbf{E}\left[X_{i,1} + \ldots + X_{i,k}\right] \\
&\leq k \times \sum_{i=1}^{m} z_i \,.
\end{aligned}$$

Puis, les (z_i) minimisent (2), qui est une relaxation de (1), donc $\sum z_i \leq \mathrm{OPT}(\mathcal{S})$.

3. Déterminer une valeur de $c > 0$ telle que, si on pose $k = \lfloor c \log n \rfloor$, on ait :

$$\mathbf{P}\left\{\mathcal{T} \text{ est un recouvrement de } U\right\} \geq 1 - \frac{1}{n}\,.$$

☞ Comme d'habitude, avec la borne de l'union, c'est plus facile de partir de :

$$\mathbf{P}\left\{\mathcal{T} \text{ n'est pas un recouvrement de } U\right\} \leq \sum_{y \in U} \mathbf{P}\left\{y \text{ n'apparait pas dans } \mathcal{T}\right\}.$$

On note A_y l'événement $\{y \notin \bigcup_{T \in \mathcal{T}} T\}$. Alors on a :

$$A_y = \left\{ \sum_{i \text{ t.q. } y \in S_i} \sum_{j=1}^{k} X_{i,j} = 0 \right\}$$

Tous les $X_{i,j}$ sont indépendants et suivent une loi de Bernouilli, la somme des paramètres est donc $\mu := \sum_{i \text{ t.q. } y \in S_i} k \times z_i$, et on a (en supposant U non vide, mais sinon tout cela n'est pas très intéressant) $\sum z_i \geq 1$, donc $\mu \geq k$.
Avec Chernoff II appliquée avec $\epsilon = 1$, on obtient

$$\mathbf{P}\left\{A_y\right\} \leq \exp(-\frac{\mu}{3}) \leq \exp(-\frac{k}{3}) \leq \exp(-\frac{c\log(n)}{3}) = n^{-c/3}\,.$$

En particulier, pour $c = 6$, on a bien $\mathbf{P}\left\{\mathcal{T} \text{ n'est pas un recouvrement de } U\right\} \leq \frac{1}{n}$.
Cas $\delta = 1$. On peut ne pas utiliser Chernoff et avoir une version + directe. On a

$$\begin{aligned}
\mathrm{P}(A_y) &= \mathrm{P}(\bigcap_{i:y\in S_i} \bigcap_{j=1}^{k} \{X_{i,j} = 0\}) && (3)\\
&= \prod_{i:y\in S_i} \prod_{j=1}^{k} \mathrm{P}(X_{i,j} = 0) && (4)\\
&= \prod_{i:y\in S_i} \prod_{j=1}^{k} (1 - z_i) && (5)\\
&\leq \prod_{i:y\in S_i} \prod_{j=1}^{k} e^{-z_i} && (6)\\
&= e^{-\sum_{i:y\in S_i} \sum_{j=1}^{k} z_i} && (7)\\
&= e^{-\mu} \leq e^{-k} \leq e^{-(c \ln n + 1)} && (8)\\
&= n^{-c} e^{-1} && (9)
\end{aligned}$$

et borne de l'union :

$$\begin{aligned}
\mathrm{P}(\bigcup_y A_y) &\leq n \mathrm{P}(A_y) && (10)\\
&= n^{1-c} e^{-1} && (11)
\end{aligned}$$

On peut prendre $c > 2$ qui fonctionne pour n pas trop petit.