# Devoir

*Documents et calculatrice sont autorisés.*
*Le barème est donné à titre indicatif.*

## Exercice 1 : double pile (7 pts)

On lance une pièce jusqu'à ce qu'on obtienne deux « Pile » de suite et on note $X$ le nombre de lancers qui ont été nécessaires. Par exemple, si la suite des lancers a donné $P\ F\ F\ P\ P$, alors $X = 5$.

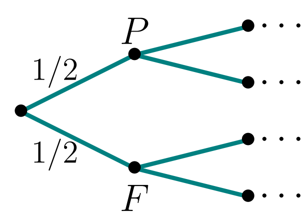
Le but de cet exercice est de déterminer la loi de $X$ ainsi que son espérance. Nous considérerons la pièce équilibrée et nous noterons :

$$\forall n \geqslant 1, \quad p_n = \mathbf{P}(X = n).$$

1. Déterminer les valeurs de $p_1$, $p_2$, $p_3$ et $p_4$.

   Il est impossible d'obtenir deux Piles de suite en un lancer, donc $p_1 = 0$.
   Une seule séquence permet d'obtenir deux Piles de suite en deux lancers : $PP$. Donc $p_2 = \mathbf{P}(PP) = \frac{1}{4}$.
   Une seule séquence permet d'obtenir deux Piles de suite en trois lancers : $FPP$. Donc $p_3 = \mathbf{P}(FPP) = \frac{1}{8}$.
   Deux séquences permettent d'obtenir deux Piles de suite en quatre lancers : $PFPP$ et $FFPP$. Donc $p_4 = \mathbf{P}(PFPP) + \mathbf{P}(FFPP) = \frac{1}{16}$.

   Pour déterminer la loi de $X$, nous allons déterminer une relation de récurrence entre les probabilités $p_n$. Pour cela, nous allons raisonner sur le début des tirages en distinguant deux cas, selon le résultat du premier lancer.

2. Justifier que pour $n \geqslant 2$,

$$\mathbf{P}(X = n \mid \text{le premier lancer est tombé sur « Face »}) = \mathbf{P}(X = n - 1).$$

   Obtenir $PP$ à l'issue de $n$ tirages sachant que le premier tirage est $F$ signifie qu'on a obtenu $PP$ à l'issue des $n-1$ tirages ayant suivi $F$. Autrement dit, le premier tirage n'ayant rien donné d'intéressant, c'est comme si le jeu débutait à partir du second tirage. On obtient bien $\mathbf{P}(X = n \mid \text{le premier lancer est tombé sur « Face »}) = \mathbf{P}(X = n - 1)$

3. Justifier de même que pour $n \geqslant 3$,

$$\mathbf{P}(X = n \mid \text{le premier lancer est tombé sur « Pile »}) = \frac{1}{2}\mathbf{P}(X = n - 2).$$

   Si le premier tirage est $P$, la situation est un peu différente que dans le cas précédent. Pour que la séquence $PP$ apparaisse à l'issue des $n$ tirages (avec $n \geqslant 3$), il ne faut pas qu'elle apparaisse avant. En particulier, il ne faut pas que le second tirage soit également $P$. Donc la séquence commence nécessairement par $PF$, puis est suivie de $n-2$ tirages ayant finalement abouti à $PP$. Ainsi, il faut tirer $F$ puis faire comme si le jeu débutait et se terminait en $n-2$ tirages.

$\mathbf{P}(X = n \mid \text{le premier lancer est tombé sur « Pile »}) = \mathbf{P}(\text{le second tirage est } F \text{ et les } n-2 \text{ suivants se}$

4. En déduire que pour $n \geqslant 3$,
$$p_n = \frac{1}{2}p_{n-1} + \frac{1}{4}p_{n-2}.$$

On décompose la probabilité $p_n$ selon les deux premières branches de l'arbre : $\mathbf{P}(X = n) = \mathbf{P}(F)\mathbf{P}(X = n \mid F) + \mathbf{P}(P)\mathbf{P}(X = n \mid P)$. Ainsi
$$p_n = \frac{1}{2}p_{n-1} + \frac{1}{2} \cdot \frac{1}{2}p_{n-2}.$$

5. On note $(u_n)_n$ la suite de Fibonacci définie par $u_1 = 0$, $u_2 = 1$ et pour $n \geqslant 3$, $u_n = u_{n-1} + u_{n-2}$.
Démontrer que pour tout $n \geqslant 1$, $\quad p_n = \frac{u_n}{2^n}$.

Démontrons le résultat par récurrence. Pour $n = 1$ et 2, on a bien $\frac{u_1}{2} = 0 = p_1$ et $\frac{u_2}{4} = \frac{1}{4} = u_2$. Supposons le résultat vrai aux rangs $n-2$ et $n-1$. Alors, d'après la question précédente :
$$p_n = \frac{1}{2}p_{n-1} + \frac{1}{4}p_{n-2} = \frac{1}{2} \cdot \frac{u_{n-1}}{2^{n-1}} + \frac{1}{4} \cdot \frac{u_{n-2}}{2^{n-2}} = \frac{u_{n-1} + u_{n-2}}{2^n} = \frac{u_n}{2^n}.$$

6. Calculer l'espérance de $X$.
*On admet les résultats suivants :*
$$\forall n,\ u_n = \frac{5-\sqrt{5}}{10}\varphi^n + \frac{5+\sqrt{5}}{10}\left(-\frac{1}{\varphi}\right)^n, \qquad \sum_{n=1}^{+\infty} n x^n = \frac{x}{(x-1)^2},$$
*avec* $\varphi = \frac{1+\sqrt{5}}{2} \approx 1{,}62$.

Remplaçons les termes $u_n$ par l'expression qui est donnée et ramenons le calcul à la série qui est rappelée.
$$\begin{aligned}
\mathbf{E}(X) &= \sum_{n=1}^{+\infty} n p_n = \sum_{n=1}^{+\infty} n \frac{u_n}{2^n} \\
&= \sum_{n=1}^{+\infty} \frac{n}{2^n}\left(\frac{5-\sqrt{5}}{10}\varphi^n + \frac{5+\sqrt{5}}{10}\left(-\frac{1}{\varphi}\right)^n\right) \\
&= \frac{5-\sqrt{5}}{10}\sum_{n=1}^{+\infty} n\left(\frac{\varphi}{2}\right)^n + \frac{5+\sqrt{5}}{10}\sum_{n=1}^{+\infty} n\left(-\frac{1}{2\varphi}\right)^n \\
&= \frac{5-\sqrt{5}}{10}\frac{\varphi/2}{(\varphi/2\ -\ 1)^2} + \frac{5+\sqrt{5}}{10}\frac{-1/2\varphi}{(-1/2\varphi\ -\ 1)^2} \\
&= 6
\end{aligned}$$

Remarque : une autre preuve est possible. Après avoir justifié que la série $\sum_{n=1}^{+\infty} np_n$ est bien convergente, on peut utiliser la relation de récurrence de la question 4 pour en déterminer la somme. Rappelons que $\sum_{k=1}^{+\infty} p_k = 1$ puisque les $p_k$ constituent la densité de $X$.

$$\begin{aligned}
\mathbf{E}(X) &= \sum_{n=1}^{+\infty} np_n \\
&= p_1 + 2p_2 + \frac{1}{2}\sum_{n=3}^{+\infty} np_{n-1} + \frac{1}{4}\sum_{n=3}^{+\infty} np_{n-2} \\
&= \frac{1}{2} + \frac{1}{2}\sum_{k=2}^{+\infty}(k+1)p_k + \frac{1}{4}\sum_{k=1}^{+\infty}(k+2)p_k \\
&= \frac{1}{2} + \left(\frac{1}{2}\sum_{k=2}^{+\infty} kp_k + \frac{1}{2}\sum_{k=2}^{+\infty} p_k\right) + \left(\frac{1}{4}\sum_{k=1}^{+\infty} kp_k + \frac{1}{4}\sum_{k=1}^{+\infty} 2p_k\right) \\
&= \frac{1}{2} + \frac{1}{2}\mathbf{E}(X) + \frac{1}{2} + \frac{1}{4}\mathbf{E}(X) + \frac{1}{2} \\
&= \frac{3}{2} + \frac{3}{4}\mathbf{E}(X)
\end{aligned}$$

On en déduit bien que $\mathbf{E}(X) = 6$.

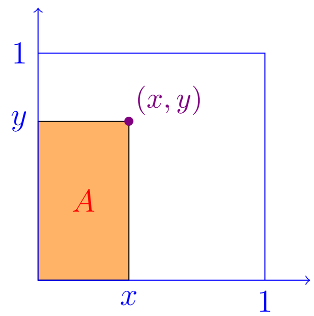
## Exercice 2 : aire aléatoire (6 pts)

On tire au hasard un point $(x, y)$ dans le carré $[0, 1] \times [0, 1]$ et on considère le rectangle dont $(0, 0)$ et $(x, y)$ sont les sommets d'une diagonale. On souhaite étudier l'aire de ce rectangle.

Précisons les hypothèses. Nous supposons que $x$ et $y$ sont tirés uniformément et indépendamment. Nous posons ainsi $X \sim \mathcal{U}([0, 1])$ et $Y \sim \mathcal{U}([0, 1])$ indépendantes. Et nous notons $A = XY$ l'aire du rectangle correspondant.

1. Quelles sont les valeurs prises par $A$ ?

   Commençons par bien nous représenter le problème. La variable $A$ représente l'aire du rectangle ci-dessous, où le point $(x, y)$ a été tiré au hasard dans le carré $[0, 1]^2$.

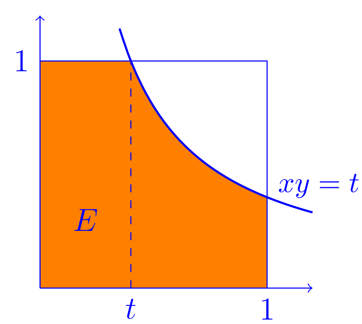



   Les variables $X$ et $Y$ étant à valeurs dans $[0, 1]$, leur produit $A$ est également à valeurs dans $[0, 1]$.

2. Montrer que la fonction de répartition de $A$ est définie par $F_A(t) = t - t\ln(t)$.

   Soit donc $t \in [0, 1]$. Alors, en s'appuyant sur la figure ci-contre et en utilisant l'indépendance de $X$ et $Y$ :

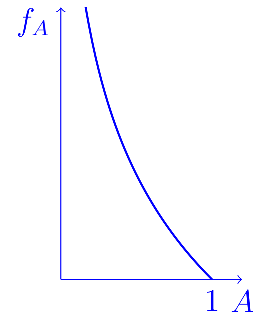
$$
\begin{aligned}
F_A(t) &= \mathbf{P}(A \leqslant t) = \mathbf{P}(XY \leqslant t) \\
&= \mathbf{P}((X, Y) \in E) \\
&= 1 - \int_{x=t}^{1} \int_{y=\frac{t}{x}}^{1} f_X(x) f_Y(y) \mathrm{d}y \mathrm{d}x \\
&= 1 - \int_{x=t}^{1} \int_{y=\frac{t}{x}}^{1} 1 \cdot 1 \ \mathrm{d}y \mathrm{d}x \\
&= 1 - \int_{t}^{1} 1 - \tfrac{t}{x} \ \mathrm{d}x \\
&= 1 - [x - t\ln(x)]_t^1 \\
&= 1 - (1 - t + t\ln(t)) = t - t\ln(t)
\end{aligned}
$$




3. En déduire la densité de $A$ et la représenter.

La densité de $A$ est alors donnée par

$$f_A(t) = F'_A(t) = 1 - \ln(t) - \frac{t}{t} = -\ln(t).$$




4. Déterminer $\mathbf{E}(A)$ ainsi que la médiane de $A$.

   Pour calculer l'espérance de $A$, nous pourrions intégrer $tf_A(t)$ sur $[0,1]$. Mais il y a plus simple. Comme $X$ et $Y$ sont indépendantes, $\mathbf{E}(A) = \mathbf{E}(XY) = \mathbf{E}(X)\mathbf{E}(Y)$. Or $X$ et $Y$ suivent la loi uniforme sur $[0,1]$ dont l'espérance est $\frac{1}{2}$, donc $\mathbf{E}(A) = \frac{1}{4}$.

   La médiane de $A$ est la valeur de $m$ telle que $\mathbf{P}(A \leqslant m) = \frac{1}{2}$, donc telle que $F_a(m) = \frac{1}{2}$. La médiane est donc la valeur $m \in [0,1]$ telle que $m - m\ln(m) = \frac{1}{2}$. À l'aide d'une calculatrice, on obtient que la médiane vaut approximativement $0,19$.

## Exercice 3 : erreur (7 pts)

Soit $X \sim \mathcal{N}(0,1)$ et soit $Y = X^2$. Il n'est pas possible d'exprimer à l'aide d'une formule simple la densité de $Y$. Mais nous allons néanmoins étudier certaines propriétés de $Y$.

1. Calculer $\mathbf{P}(Y \leqslant 2)$.

$$\mathbf{P}(Y \leqslant 2) = \mathbf{P}(X^2 \leqslant 2) = \mathbf{P}(-\sqrt{2} \leqslant X \leqslant \sqrt{2}) = 2F_{\mathcal{N}}(\sqrt{2}) \approx 2 \cdot 0,92 - 1 \approx 0,84.$$

2. Montrer que $\mathbf{E}(Y) = 1$. On pourra utiliser le théorème de transfert et reconnaître la variance de $X$.

$$\mathbf{E}(Y) = \mathbf{E}(X^2) = \int_{-\infty}^{+\infty} x^2 \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{x^2}{2}} \,\mathrm{d}x.$$

   Or $X$ est d'espérance nulle, donc sa variance est donnée par $\mathbf{V}(X) = \mathbf{E}((X-0)^2) = \mathbf{E}(X^2)$. On reconnaît l'espérance de $Y$. Donc $\mathbf{E}(Y) = \mathbf{V}(X) = 1$.

   Raisonnement équivalent avec la formule de König : on sait que $\mathbf{V}(X) = \mathbf{E}(X^2) - \mathbf{E}(X)^2$, donc $\mathbf{E}(Y) = \mathbf{E}(X^2) = \mathbf{V}(X) + \mathbf{E}(X)^2 = 1 + 0^2$.

3. Donner, également à l'aide du théorème de transfert, l'expression sous forme intégrale de la variance de $Y$. On admet pour la suite que $\mathbf{V}(Y) = 2$.

   De même,

$$\mathbf{V}(Y) = \mathbf{E}((Y-1)^2) = \mathbf{E}((X^2-1)^2) = \int_{-\infty}^{+\infty} (x^2-1)^2 \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{x^2}{2}} \,\mathrm{d}x.$$

   En manipulant cette intégrale (développement et intégration par partie), on montre que $\mathbf{V}(Y) = 2$.

4. Soient $Y_1, \ldots, Y_{100}$ 100 variables indépendantes de même loi que $Y$ et soit $S = \sum_1^{100} Y_i$. Appliquer le théorème central limite et déterminer la valeur $\alpha$ pour laquelle

$$\mathbf{P}(S \leqslant \alpha) \approx 0,95.$$

Comme $100 \geqslant 30$ et les variables $Y_i$ sont indépendantes, nous pouvons appliquer le théorème central limite : $\frac{S-100\cdot\mathbf{E}(Y)}{\sqrt{100}\cdot\sigma(Y)} \sim \mathcal{N}(0,1)$ à peu près. Nous considérons donc que $\frac{S-100}{\sqrt{200}} \sim \mathcal{N}(0,1)$. Ainsi

$$\mathbf{P}(S \leqslant \alpha) = \mathbf{P}\left(\frac{S-100}{\sqrt{200}} \leqslant \frac{\alpha-100}{\sqrt{200}}\right) \approx F_{\mathcal{N}}\left(\frac{\alpha-100}{\sqrt{200}}\right).$$

Or $F_{\mathcal{N}}(t) = 0,95$ pour $t \approx 1,65$, donc $\frac{\alpha-100}{\sqrt{200}} \approx 1,65$ et on obtient $\alpha \approx 123,3$.

5. On mesure 100 grandeurs $m_1, m_2, \ldots, m_{100}$. Les imprécisions de mesure sont modélisées par des lois normale de variance 1 : la mesure de $m_i$ est une variable aléatoire $Z_i$ de loi $\mathcal{N}(m_i, 1)$. Estimer la probabilité que la distance euclidienne entre le vecteur $(m_1, m_2, \ldots, m_{100})$ et le vecteur $(Z_1, Z_2, \ldots, Z_{100})$ soit supérieure à 11.

Nous souhaitons estimer $\mathbf{P}(\sqrt{\sum_{i=1}^{100}(Z_i - m_i)^2} \geqslant 11) = \mathbf{P}(\sum_{i=1}^{100}(Z_i - m_i)^2 \geqslant 121)$. Par hypothèse $Z_i - m_i \sim \mathcal{N}(0,1)$, donc les variables $(Z_i - m_i)^2$ ont la même loi que les variables $Y_i$. En les supposant indépendantes (ce qui est raisonnable pour un ensemble de mesures), nous retrouvons les conditions de la question précédente. Nous avions obtenu $\mathbf{P}(S \leqslant 123,3) \approx 0,95$. Comme 121 est proche de $123,3$, on peut affirmer que la probabilité recherchée est de l'ordre de $1 - 0,95 = 0,05$.

*Cette loi est très utilisée en statistique pour tester des hypothèses. L'idée est de mesurer la distance entre les données observées et celles correspondant aux hypothèses faites. Si cette distance est anormalement élevée, c'est-à-dire dans une zone de probabilité très faible, on peut douter des hypothèses qui ont été posées. L'étude de cette loi permet, comme dans la question 5, de déterminer des seuils à partir desquels ont peut remettre en cause des hypothèses statistiques.*

*Quelques valeurs de la* **fonction de répartition** $F_{\mathcal{N}}$ *de la loi* $\mathcal{N}(0,1)$.

| $x$ | 0 | 0,31 | 0,63 | 0,96 | 1,23 | 1,41 | 1,65 | 1,96 | 2,00 | 2,33 |
|---|---|---|---|---|---|---|---|---|---|---|
| $F_{\mathcal{N}}(x)$ | 0,5 | 0,63 | 0,74 | 0,83 | 0,89 | 0,92 | 0,95 | 0,975 | 0,98 | 0,99 |

Les valeurs pour $x < 0$ se déduisent de l'égalité $F_{\mathcal{N}}(-x) = 1 - F_{\mathcal{N}}(x)$.