Corrigé du devoir

Exercice 1 : (5 pts)

Un joueur tire 2 cartes dans un jeu de 32 cartes.

1. Quelle est la probabilité qu'il obtienne une paire d'as ?

   Commençons par modéliser le problème. Posons Ω l'ensemble des parties $\{a, b\}$ de deux cartes tirées parmi les 32 cartes. Il contient $\binom{32}{2} = 496$ éléments. On suppose les tirages indépendants et les toutes les cartes équiprobables. On attribue donc à chaque élément de Ω la probabilité $\frac{1}{496}$.

   Répondons maintenant à la question : il y a $\binom{4}{2} = 6$ paires de deux as possibles. La probabilité de tirer deux as est donc $\frac{6}{496} \approx 0,012$.

   Après le premier tirage, le joueur peut conserver ses deux cartes ou en jeter une et en tirer une nouvelle dans le reste du paquet.

2. Quelle est désormais la probabilité de détenir finalement une paire d'as ?
   (on supposera le joueur assez intelligent pour ne pas jeter un as s'il en a déjà deux.)

   Il faudrait a priori modifier l'univers Ω pour décrire ce nouveau problème. On peut éviter cela et décrire la situation à l'aide d'un arbre. Pour obtenir une paire d'as, il faut soit tirer deux as, soit en tirer un seul et jeter l'autre carte pour un as.
   Il y a 4×28 couples de cartes contentant un seul as. Une fois la mauvaise carte jetée, il reste 3 chances sur 30 de tirer un as parmi les 30 cartes restantes. La probabilité totale d'obtenir deux as est :

   $$\frac{6}{496} + \frac{4 \times 28}{496}\frac{3}{30} \approx 0,035.$$

3. Toujours avec la même règle, quelle est la probabilité d'obtenir une paire de deux cartes de même valeur ?

   Reprenons l'idée de la question précédente : soit on tire directement une paire, soit ce n'est pas le cas et on jette l'une des deux en espérant retirer une carte de la même valeur que celle qu'on a gardée.
   Il y a 8 valeurs différentes et 6 paires possibles pour une valeur donnée. Cela fait $6 \times 8 = 48$ paires possibles. Et il y a donc $496 - 48 = 448$ combinaisons sans paire. Comme précédemment, il y a dans ce cas 3 chances sur 30 de tirer une bonne carte parmi les 30 restantes. La probabilité totale est :

   $$\frac{48}{496} + \frac{448}{496}\frac{3}{30} \approx 0,187.$$

Exercice 2 : test d'indépendance (15 pts)

On a programmé un générateur de nombres aléatoires permettant de simuler des tirages indépendants selon la loi uniforme sur $[0,1]$. Le caractère uniforme des nombres tirés a été validé et nous souhaiterions vérifier statistiquement que les tirages successifs obtenus sont deux à deux indépendants.

Nous noterons X et Y les résultats de deux tirages successifs de notre générateur. Nous considérons que ces variables suivent la même loi $\mathcal{U}([0,1])$ (d'espérance $\frac{1}{2}$ et de variance $\frac{1}{12}$). Afin de vérifier expérimentalement qu'elles sont indépendantes, nous supposerons dans tout le problème qu'elles le sont et établirons alors des résultats théoriques que nous pourrons comparer aux résultats expérimentaux. Par exemple, on doit théoriquement avoir $\mathbf{E}(XY) = \mathbf{E}(X)\mathbf{E}(Y)$ et nous pouvons regarder si cela est validé par l'expérience.

Nous noterons $Z = XY$ le produit de nos deux variables.

I. **Espérance de Z** (5 pts)
Nous l'avons dit, en supposant que X et Y sont indépendantes, on a

$$\mathbf{E}(Z) = \mathbf{E}(X)\mathbf{E}(Y) = \tfrac{1}{2} \times \tfrac{1}{2} = \tfrac{1}{4}.$$

1. Démontrer que la variance de Z vérifie :

$$\mathbf{V}(Z) = \mathbf{V}(X)\mathbf{V}(Y) + \mathbf{V}(X)\mathbf{E}(Y)^2 + \mathbf{E}(X)^2\mathbf{V}(Y).$$

Calculer ainsi sa valeur.

$$\mathbf{V}(Z) = \mathbf{E}(Z^2) - \mathbf{E}(Z)^2 = \mathbf{E}(X^2Y^2) - \mathbf{E}(X)^2\mathbf{E}(Y)^2.$$

Comme X et Y sont indépendantes, X^2 et Y^2 le sont aussi, donc $\mathbf{E}(X^2Y^2) = \mathbf{E}(X^2)\mathbf{E}(Y^2)$. Ainsi

$$\mathbf{V}(Z) = \mathbf{E}(X^2)\mathbf{E}(Y^2) - \mathbf{E}(X)^2\mathbf{E}(Y)^2 = (\mathbf{V}(X)+\mathbf{E}(X)^2)(\mathbf{V}(Y)+\mathbf{E}(Y)^2) - \mathbf{E}(X)^2\mathbf{E}(Y)^2$$
$$= \mathbf{V}(X)\mathbf{V}(Y) + \mathbf{V}(X)\mathbf{E}(Y)^2 + \mathbf{E}(X)^2\mathbf{V}(Y).$$

Pour nos variables, $\mathbf{E}(X) = \mathbf{E}(Y) = \frac{1}{2}$ et $\mathbf{V}(X) = \mathbf{V}(Y) = \frac{1}{12}$ et on obtient $\mathbf{V}(Z) = \frac{7}{144}$.

Nous avons tiré 1000 couples de valeurs (X, Y) et avons calculé pour chacun le produit XY. La moyenne de ces 1000 nombres obtenus était égale à $0,263$.

2. On note $Z_1, Z_2, \ldots, Z_{1000}$ des tirages indépendants de même loi que Z et on note $M = \frac{1}{1000}\sum_{i=1}^{1000} Z_i$ leur moyenne.
Énoncer le théorème central limite pour M.

Les variables Z_i sont indépendantes, de même loi, d'espérance $\frac{1}{4}$ et de variance $\frac{7}{144}$. En estimant que 1000 est assez grand, on peut considérer que $\frac{M-\frac{1}{4}}{\sqrt{\frac{7}{144}}/\sqrt{1000}}$ suit approximativement la loi normale $\mathcal{N}(0,1)$.

3. Déterminer le nombre a pour lequel

$$\mathbf{P}(|M - \tfrac{1}{4}| < a) \approx 0,99.$$

$$\mathbf{P}(|M - \tfrac{1}{4}| < a) = \mathbf{P}\left(\left|\frac{M - \frac{1}{4}}{\sqrt{\frac{7}{144}}/\sqrt{1000}}\right| < \frac{a}{\sqrt{\frac{7}{144}}/\sqrt{1000}}\right) \approx 2F_{\mathcal{N}}\left(\frac{a}{\sqrt{\frac{7}{144}}/\sqrt{1000}}\right) - 1$$

d'après le TCL. Pour que cette probabilité soit de l'ordre de $0,99$, il faut que $F_{\mathcal{N}}\left(\frac{a}{\sqrt{\frac{7}{144}}/\sqrt{1000}}\right) \approx 0,995$ et on obtient avec les tables que $\frac{a}{\sqrt{\frac{7}{144}}/\sqrt{1000}} \approx 2,57$. On obtient $a \approx 0,018$.

4. La moyenne expérimentale de $0,263$ semble-t-elle très éloignée de la moyenne théorique de $\frac{1}{4}$? Peut-on avoir des doutes sur l'indépendance des variables X et Y ?

Notre résultat précédent montre que dans 99% des cas, la moyenne M de 100 tirages de la variable Z est comprise entre $0,25 - 0,018 = 0,232$ et $0,25 + 0,018 = 0,268$. Autrement dit, il serait étonnant (mais néanmoins possible) d'obtenir une moyenne M en dehors de cet intervalle. La valeur de $0,263$ obtenue expérimentalement est comprise dans cet intervalle. L'écart à la moyenne théorique de $0,25$ n'a donc rien d'extraordinaire et nous n'avons pas de raison de remettre en cause l'égalité $\mathbf{E}(Z) = \mathbf{E}(XY) = \mathbf{E}(X)\mathbf{E}(Y)$. Nous n'avons pour le moment pas d'élément pour douter de l'indépendance de X et Y.

II. **Loi de Z** (4 pts)

Notre premier test ne porte que sur l'espérance de Z, nous aimerions un test plus précis portant sur la répartition des valeurs de Z. Commençons par déterminer la loi de Z.

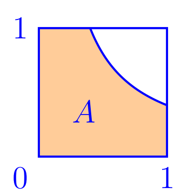
1. Toujours en supposant les variables X et Y indépendantes, montrer que la fonction de répartition de Z est définie par

$$\forall t \in [0,1], \quad F_Z(t) = t - t\ln(t).$$

Soit $t \in [0,1]$.

$$F_Z(t) = \mathbf{P}(Z \leqslant t) = \mathbf{P}(XY \leqslant t).$$

C'est un problème de couple de variables. On a supposé les variables X et Y indépendantes ; la densité de leur couple est le produit de leurs densités. On commence par représenter l'ensemble $\{XY \leqslant t\}$ en n'oubliant pas que X et Y sont à valeurs dans $[0,1]$.



Alors

$$\mathbf{P}(XY \leqslant t) = \iint_A f_X(x) f_Y(y)\,dxdy.$$

Comme X et Y suivent la loi uniforme sur $[0,1]$, leurs densités sont constantes égales à 1 et l'intégrale sur A est simplement égale à l'aire de A. Pour la calculer, on peut décomposer A en l'union d'un rectangle et de la partie située sous la courbe d'équation $y = \frac{t}{x}$. Ainsi :

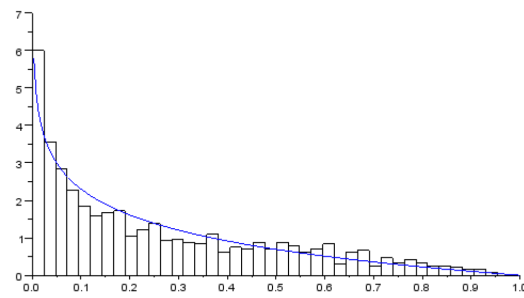
$$F_Z(t) = t + \int_t^1 \tfrac{t}{x} dx = t - t\ln(t).$$

2. En déduire la densité de Z.

La densité f_Z de Z est simplement la dérivée de F_Z :

$$\forall t \in [0,1], \quad f_Z(t) = F_Z'(t) = 1 - \ln(t) - \tfrac{t}{t} = -\ln(t).$$

Nous avons représenté ci-contre cette densité ainsi que les 1000 valeurs de Z obtenues expérimentalement réparties dans un histogramme. On constate qu'elles se répartissent à peu près selon la densité théorique mais pas parfaitement. Nous allons maintenant essayer de savoir si les écarts observés sont significatifs.



3. Calculer la probabilité théorique $p = \mathbf{P}(Z > \frac{1}{2})$.

$$p = 1 - \mathbf{P}(Z \leqslant \tfrac{1}{2}) = 1 - F_Z(\tfrac{1}{2}) = 1 - (\tfrac{1}{2} - \tfrac{1}{2}\ln(\tfrac{1}{2})) \approx 0,153.$$

4. On effectue 1000 tirages de la variable Z et on note N le nombre de résultats supérieurs à $\frac{1}{2}$.
Quelle est la loi de N ? On précisera bien ses paramètres.

Toujours en supposant les tirages indépendants, on reconnaît une loi binomiale : N est le nombre de succès parmi 1000 essais. Chaque succès a pour probabilité p, donc $N \sim \mathcal{B}(1000, p)$.

5. Estimer à l'aide du TCL le nombre b tel que $\mathbf{P}(N < b) \approx 0,99$.

Appliquons le TCL et plus précisément le théorème de Moivre-Laplace : en estimant que 1000 est assez grand, $\frac{N-1000p}{\sqrt{1000p(1-p)}}$ suit approximativement la loi normale $\mathcal{N}(0,1)$.Alors

$$\mathbf{P}(N < b) = \mathbf{P}(\frac{N - 1000p}{\sqrt{1000p(1-p)}} < \frac{b - 1000p}{\sqrt{1000p(1-p)}}) \approx F_{\mathcal{N}}(\frac{b - 1000p}{\sqrt{1000p(1-p)}}).$$

D'après les tables, cette probabilité est de l'ordre de $0,99$ si $\frac{b-1000p}{\sqrt{1000p(1-p)}}$ est de l'ordre de $2,33$. On en déduit $b \approx 180$.

6. Sur nos 1000 tirages, il y en avait 201 qui étaient supérieurs à $\frac{1}{2}$. Que peut-on en déduire ?

   D'après la question précédente, sous l'hypothèse d'indépendance de X et Y, la valeur de N est inférieure à 180 dans 99% des cas. Avoir obtenu $N = 201$ est donc surprenant, c'est statistiquement très éloigné de la valeur moyenne théorique de 153. Même si une telle valeur est possible, elle est très peu probable et on peut se poser des questions sur notre hypothèse d'indépendance.

7. Conclusion : peut-on douter de l'indépendance des variables X et Y ? Si oui, essayer de préciser à partir de l'histogramme, le lien de dépendance qu'il pourrait y avoir entre X et Y.

   La moyenne expérimentale de Z était assez proche de la moyenne théorique si on suppose l'indépendance de X et Y. En revanche, la valeur de N est statistiquement trop élevée par rapport à ce qu'on devrait observer si X et Y étaient indépendantes. On peut donc douter de cette indépendance.
   Ce second test est plus précis que le premier car il porte sur la répartition des valeurs de Z. Si on observe l'histogramme des données observées, on constate qu'il est un peu plus élevé que la densité théorique pour les grandes valeurs de Z et un peu plus bas pour les valeurs plus basses. Notre test avec N a montré que cet écart était significatif. Il semble donc que $Z = XY$ prend statistiquement plus de grandes valeurs que dans le cas théorique de l'indépendance. On peut donc penser que X et Y ont tendance à avoir des valeurs élevées en même temps. Le fait que X soit élevé augmenterait la probabilité que Y le soit aussi.

La méthode proposée dans ce problème n'est pas la méthode standard pour tester l'indépendance de deux variables. En général, on teste sur plusieurs parties A et B si on a bien $\mathbf{P}(X \in A \text{ et } Y \in B) = \mathbf{P}(X \in A)\mathbf{P}(Y \in B)$. Cette façon de faire s'automatise dans un test appelé test du χ^2.