



## TOP 09

### AUTOUR DE LA NOTION D'INDICATRICE.

Les fichiers TOP sont réservés aux étudiants qui préparent le Top 5. Ils sont plus difficiles et demandent déjà une bonne maîtrise du reste du programme (cours, exercices, TD et méthodes). Même si le contenu de ces exercices dépasse le cadre du programme de ECG2, ils peuvent inspirer une série de questions d'un texte de concours.

Cet exercice concerne les chapitres 12 et 14 (variables à densités et estimation).

#### Exercice 1.- *Préambule.*

Soient  $(\Omega, \mathcal{A}, P)$  un espace probabilisé et  $A$  un évènement.

On appelle *variable indicatrice* de l'évènement  $A$  la variable aléatoire notée  $\mathbb{1}_A$  définie par

$$\mathbb{1}_A(\omega) = \begin{cases} 1, & \text{si } \omega \in A \\ 0, & \text{si } \omega \notin A \end{cases}$$

1. Reconnaître la loi de  $\mathbb{1}_A$  et son espérance.
2. Soient  $A$  et  $B$  deux évènements. Montrer que  $\mathbb{1}_A \times \mathbb{1}_B = \mathbb{1}_{A \cap B}$ .

Soit  $I$  un intervalle de  $\mathbb{R}$ . On introduit la *fonction indicatrice* de  $I$  avec la formule

$$\chi_I(x) = \begin{cases} 1, & \text{si } x \in I \\ 0, & \text{si } x \notin I \end{cases}$$

Soient  $X$  une variable aléatoire réelle de densité  $f$  et  $s \in \mathbb{R}$ . On remarque alors que

$$\mathbb{1}_{[X \geq s]} = \chi_{[s; +\infty[}(X).$$

3. Soient  $X$  une variable aléatoire réelle de densité  $f$  et  $s \in \mathbb{R}$ . Soit  $Y = g(X)$  une autre variable aléatoire. Justifier soigneusement, que  $\mathbb{1}_{[X \geq s]}Y$  admet une espérance si et seulement si l'intégrale

$$\int_{-\infty}^{+\infty} \chi_{[s; +\infty[}(x)g(x)f(x)dx = \int_s^{+\infty} g(x)f(x)dx$$

converge.

#### Exercice 2.-

1. *Question de cours* : Rappeler l'inégalité de Markov.

Un fabricant cherche à optimiser le prix de vente  $p$  de son produit. Il suppose que chaque client potentiel est prêt à payer un prix aléatoire  $X$  pour acheter le produit. Si  $X$  est supérieur ou égal à  $p$ , la vente a lieu le fabricant encaisse  $p$  euros ; sinon la vente n'a pas lieu et le fabricant n'encaisse aucun chiffre d'affaires.

On suppose que  $X$  admet une densité de probabilité  $f$  continue sur  $\mathbb{R}_+$  et une espérance finie. On note  $F$  la fonction de répartition et  $Z_p$  le chiffre d'affaires réalisé si le prix de son produit est fixé à la valeur  $p$  et on définit la fonction  $g$  sur  $\mathbb{R}_+$  par  $g(p) = E(Z_p)$ .

2. Exprimer  $Z_p$  à l'aide de  $\mathbb{1}_{(X \geq p)}$  puis montrer que  $g(p) = p(1 - F(p))$ .
3. Montrer que  $g(p)$  tend vers 0 lorsque  $p$  tend vers  $+\infty$ .
4. Que vaut  $g(0)$  ? Dessiner l'allure de la courbe de  $g$ . En déduire que  $g$  admet un maximum sur  $\mathbb{R}_+$ , atteint en une valeur  $p^* > 0$ .
5. Expliciter une équation vérifiée  $p^*$ .
6. Quel chiffre d'affaires maximal peut espérer le fabricant si  $X$  suit une loi exponentielle de paramètre  $\lambda > 0$  ?

### Exercice 3.-

On étudie le jeu suivant :  $X_1$  et  $X_2$  étant deux variables indépendantes de loi uniforme sur  $[0; 1]$ , un joueur observe d'abord  $X_1$ . S'il décide d'en rester là, il gagne la valeur observée. S'il décide de continuer, il observe et gagne  $X_2$ . On note  $G$  la v.a. correspondant au gain du joueur.

1. Sa première stratégie est de toujours continuer et observer  $X_2$ . Quelle est alors l'espérance de son gain ?
2. Dans un deuxième temps, il décide de continuer et d'observer  $X_2$  si et seulement si  $X_1 \leq s$ , où  $s \in [0; 1]$  est un seuil qu'il se fixe à l'avance.
  - a. Justifier que
 
$$G = \mathbb{1}_{X_1 \geq s} X_1 + \mathbb{1}_{X_1 < s} X_2.$$
  - b. Écrire un programme Python qui demande à l'utilisateur d'entrer le seuil  $s$  et simule le gain.
  - c. Quelle est l'espérance du gain dans ce cas ?
  - d. Quelle valeur  $s$  doit-il choisir pour maximiser cette espérance ?
3. Si, dans une variante du jeu précédent, on pouvait observer  $X_1$  et  $X_2$  avant de prendre une décision, quelle serait la stratégie et combien rapporterait-elle en moyenne ?

### Exercice 4.- *Le collectionneur.*

Cet exercice s'inspire très largement d'un problème de l'excellent ouvrage *Introduction aux probabilités*, Quentin Berger *et al.*, Dunod (2021).

Un collectionneur souhaite compléter un album (panini) de  $N$  vignettes différentes, numérotées de 1 à  $N$  et vendues dans des paquets séparés. Chaque paquet ne contient qu'une seule vignette et on suppose que chaque paquet, indépendamment les uns des autres, possède la même probabilité de contenir l'une des  $N$  vignettes existantes.

La question est alors de savoir combien de paquets faut-il acheter pour avoir *de bonnes chances* de compléter l'album.



On introduit alors les variables aléatoires suivantes :

- Pour  $k \in \mathbb{N}^*$ , on note  $X_k$  la variable aléatoire correspondant au numéro de la vignette présente dans le  $k$ -ième paquet acheté ;
- $T_N$  la variable aléatoire égale au nombre de paquets achetés pour obtenir les  $N$  vignettes différentes ;
- Pour  $k \in \mathbb{N}$ , on note  $Y_k$  la variable aléatoire au nombre de vignettes manquantes (pour compléter la collection) après l'achat de  $k$  paquets. En particulier,  $Y_0 = N$  et  $Y_1 = N - 1$ .

1. a. Qui sont les deux joueurs sur les photos ?

*Indication : ils sont tous les deux attaquants et Tony Vairelles n'a jamais joué à Bordeaux (contrairement à ce qu'affirme sa page Wikipédia).*

- b. À quels titres des Girondins sont-ils associés ?

2. Quelle est, pour tout  $k \in \mathbb{N}^*$ , la loi de  $X_k$  ?

3. Déterminer la loi de  $Y_2$ .

4. Justifier que

$$[T_n > k] = [Y_k \geq 1].$$

5. On note  $A_i^{(k)}$  l'évènement "la vignette numéro  $i$  n'a pas encore été trouvée après l'achat de  $k$  paquets".

- a. En écrivant  $A_i^{(k)}$  comme une intersection faisant intervenir les variables  $X_j$  (pour  $1 \leq j \leq k$ ), déterminer  $P(A_i^{(k)})$ .

- b. Justifier que

$$Y_k = \sum_{i=1}^N \mathbb{1}_{A_i^{(k)}}.$$

- c. En déduire que

$$E(Y_k) = N \times \left(1 - \frac{1}{N}\right)^k.$$

- d. (i) Montrer que, pour tout  $x \in \mathbb{R}$ ,  $1 - x \leq e^{-x}$ .

- (ii) En déduire, à l'aide de l'inégalité de Markov, que

$$P(T_N > k) \leq N e^{-k/N}.$$

- (iii) Soit  $\varepsilon > 0$ . Déterminer  $K_0 = K_0(\varepsilon, N)$  (c'est à dire un entier  $K_0$  qui dépende de  $N$  et de  $\varepsilon$ ) tel que, si  $k \geq K_0$ , on a

$$P(T_n > k) \leq \frac{1}{N^\varepsilon}.$$

- (iv) Interpréter ce résultat si  $N$  est grand.

6. a. Déterminer la variance de  $\mathbb{1}_{A_i^{(k)}}$  et  $\text{Cov}(\mathbb{1}_{A_i^{(k)}}, \mathbb{1}_{A_j^{(k)}})$ .

- b. Montrer que<sup>1</sup>

$$V(Y_k) = \sum_{i=1}^N V(\mathbb{1}_{A_i^{(k)}}) + \sum_{\substack{1 \leq i, j \leq N \\ i \neq j}} \text{Cov}(\mathbb{1}_{A_i^{(k)}}, \mathbb{1}_{A_j^{(k)}}).$$

- c. Dédurre des deux questions précédentes que

$$V(Y_k) \leq E(Y_k).$$

7. a. Vérifier que  $[Y_k = 0] \subset [|Y_k - E(Y_k)| \geq E(Y_k)]$ .

- b. En déduire que

$$P(Y_k = 0) \leq \frac{1}{N \left(1 - \frac{1}{N}\right)^k}.$$

8. Soit  $\varepsilon > 0$ .

- a. Montrer que, si  $k \leq K_1 = (1 - \varepsilon)N \ln(N)$ , alors

$$P(T_N \leq k) \leq \frac{1}{N \left(1 - \frac{1}{N}\right)^{(1-\varepsilon)N \ln(N)}}.$$

- b. Montrer que

$$N \left(1 - \frac{1}{N}\right)^{(1-\varepsilon)N \ln(N)} \sim N^\varepsilon, \quad N \rightarrow +\infty.$$

- c. Interpréter.

### Exercice 5.-

On rappelle que si  $X$  est une variable aléatoire **admettant une variance**, alors la loi faible des grands nombres permet d'affirmer que, notant  $(X_1, \dots, X_n)$  un  $n$ -échantillon de  $X$ , on a

$$\forall t > 0, \quad P\left(\left|\frac{X_1 + \dots + X_n}{n} - E(X)\right| > t\right) \xrightarrow{n \rightarrow +\infty} 0.$$

La preuve de ce résultat repose sur l'inégalité de Bienaymé-Tchebychev (on invite d'ailleurs à la reprendre avant d'attaquer ce problème).

L'objectif est de montrer que **ce résultat reste vrai** lorsque  $X$  **admet une espérance mais pas de variance**.

Pour simplifier les calculs, on se place dans la situation d'une variable  $X$  centrée.

On considère donc une variable aléatoire  $X$  de densité<sup>2</sup>  $f$ , d'espérance nulle dont on a un  $n$ -échantillon  $(X_1, X_2, \dots, X_n)$ .

1. Cette question apparaît aussi dans le sujet **HEC 2018**.

2. Tout ceci fonctionne de manière analogue dans le cas d'une variable discrète

**Partie 1 - Des variables tronquées.** Soit  $A > 0$  fixé. Pour tout  $k \in \llbracket 1; n \rrbracket$ , on introduit les variables

$$Y_k = X_k \cdot \mathbb{1}_{(|X_k| \leq A)}, \quad \text{et} \quad Z_k = X_k \cdot \mathbb{1}_{(|X_k| > A)}.$$

1. Quelle relation a-t-on entre  $X_k, Y_k$  et  $Z_k$  ?
2. Montrer que  $Y_k$  admet un moment d'ordre 2 et que  $E(Y_k^2) \leq A^2$ .
3. Montrer que

$$\lim_{A \rightarrow +\infty} E(Y_k) = E(X_k) = 0 \quad \text{et que} \quad \lim_{A \rightarrow +\infty} E(Z_k) = 0.$$

**Partie 2 - Inégalités.** Dans toute la suite, on considère  $\varepsilon > 0$  fixé.

4. Soient  $x, y \in \mathbb{R}$ . Montrer que

$$|x + y| > t \implies \left( \left[ |x| > \frac{t}{2} \right] \text{ ou } \left[ |y| > \frac{t}{2} \right] \right)$$

5. On note alors

$$\bar{X}_n = \frac{X_1 + \dots + X_n}{n}, \quad \bar{Y}_n = \frac{Y_1 + \dots + Y_n}{n} \quad \text{et} \quad \bar{Z}_n = \frac{Z_1 + \dots + Z_n}{n}.$$

Déduire de la question précédente que

$$P(|\bar{X}_n| > t) \leq P\left(|\bar{Y}_n| > \frac{t}{2}\right) + P\left(|\bar{Z}_n| > \frac{t}{2}\right).$$

6. a. En réutilisant la Question (4), montrer que

$$P\left(|\bar{Z}_n| > \frac{t}{2}\right) \leq nP\left(|Z_1| > \frac{nt}{2}\right).$$

- b. En déduire, à l'aide de l'inégalité de Markov, qu'il existe  $A_1 > 0$ , tel que, si  $A \geq A_1$ ,

$$P\left(|\bar{Z}_n| > \frac{t}{2}\right) \leq \frac{\varepsilon}{2}.$$

7. a. Montrer que

$$E(\bar{Y}_n^2) = \frac{1}{n^2} \left( \sum_{k=1}^n E(Y_k^2) + 2 \sum_{1 \leq i < j \leq n} E(Y_i Y_j) \right).$$

- b. Montrer que, pour  $i \neq j$ ,

$$E(Y_i Y_j) = E(Y_i)E(Y_j) = E(Y_1)^2$$

En déduire que

$$2 \sum_{1 \leq i < j \leq n} E(Y_i Y_j) \leq n(n-1)E(Y_1)^2.$$

- c. Obtenir ensuite que

$$E(\bar{Y}_n^2) \leq \frac{A^2}{n} + E(Y_1)^2.$$

- d. Justifier de l'existence d'un  $A_2 > 0$  tel que pour  $A \geq A_2$ ,

$$E(Y_1)^2 \leq \frac{16\varepsilon}{t^2}.$$

- e. Utiliser une inégalité de Markov pour obtenir que

$$P\left(|\bar{Y}_n| > \frac{t}{2}\right) \leq \frac{t^2 A^2}{4n} + \frac{\varepsilon}{4}.$$

8. Montrer que si  $A \geq \max(A_1, A_2)$ ,

$$P(|\bar{X}_n| > t) \leq \frac{t^2 A^2}{4n} + \frac{3\varepsilon}{2}.$$

9. Conclure.

Pour que ces exercices soient profitables, il n'est pas recommandé d'utiliser la correction trop tôt. La recherche de la solution a beaucoup plus d'intérêt que la solution elle-même.

### Correction 1.-

Soient  $(\Omega, \mathcal{A}, P)$  un espace probabilisé et  $A$  un évènement.

On appelle *variable indicatrice* de l'évènement  $A$  la variable aléatoire notée  $\mathbb{1}_A$  définie par

$$\mathbb{1}_A(\omega) = \begin{cases} 1, & \text{si } \omega \in A \\ 0, & \text{si } \omega \notin A \end{cases}$$

1. La variable  $\mathbb{1}_A$  prend la valeur 0 ou 1 ; c'est donc une variable de Bernoulli. Son paramètre est égal à la probabilité de prendre la valeur 1

$$P(\{\omega : \mathbb{1}_A(\omega) = 1\}) = P(\{\omega : \omega \in A\}) = P(A).$$

En particulier,

$$E(\mathbb{1}_A) = P(A), \quad V(\mathbb{1}_A) = P(A)(1 - P(A)).$$

2. Soient  $A$  et  $B$  deux évènements.

$$\begin{aligned} \mathbb{1}_A(\omega) \cdot \mathbb{1}_B(\omega) = 1 &\iff \mathbb{1}_A(\omega) = 1 \quad \text{et} \quad \mathbb{1}_B(\omega) = 1 \\ &\iff \omega \in A \quad \text{et} \quad \omega \in B \\ &\iff \omega \in A \cap B \\ &\iff \mathbb{1}_{A \cap B}(\omega) = 1 \end{aligned}$$

Ayant raisonné par équivalences, on a bien

$$\mathbb{1}_A \times \mathbb{1}_B = \mathbb{1}_{A \cap B}.$$

3. On calcule en utilisant les deux questions précédentes

$$\begin{aligned} \text{Cov}(\mathbb{1}_A, \mathbb{1}_B) &= E(\mathbb{1}_A \cdot \mathbb{1}_B) - E(\mathbb{1}_A)E(\mathbb{1}_B) \\ &= E(\mathbb{1}_{A \cap B}) - E(\mathbb{1}_A)E(\mathbb{1}_B) \\ &= P(A \cap B) - P(A)P(B). \end{aligned}$$

Soit  $I$  un intervalle de  $\mathbb{R}$ . On introduit la *fonction indicatrice* de  $I$  avec la formule

$$\chi_I(x) = \begin{cases} 1, & \text{si } x \in I \\ 0, & \text{si } x \notin I \end{cases}$$

Soient  $X$  une variable aléatoire réelle de densité  $f$  et  $s \in \mathbb{R}$ . On remarque alors que

$$\mathbb{1}_{[X \geq s]} = \chi_{[s; +\infty[}(X).$$

4. On observe que

$$\mathbb{1}_{[X \geq s]}Y = \mathbb{1}_{[X \geq s]}g(X) = \chi_{[s; +\infty[}(X)g(X).$$

Ainsi, par le théorème de transfert, cette variable aléatoire admet une espérance si et seulement si l'intégrale

$$\int_{-\infty}^{+\infty} \chi_{[s; +\infty[}(x)g(x)f(x)dx$$

converge.

Comme  $\chi_{[s; +\infty[}(x) = 0$  si  $x < s$  et 1 si  $x \geq s$ , cette même intégrale est égale à

$$\int_s^{+\infty} g(x)f(x)dx.$$

**Correction 2.- 1. Question de cours :** Inégalité de Markov.

Soit  $X$  une variable aléatoire positive (ou nulle) admettant une espérance. Alors, pour tout  $t > 0$ , on a

$$P(X > t) \leq \frac{E(X)}{t}.$$

Un fabricant cherche à optimiser le prix de vente  $p$  de son produit. Il suppose que chaque client potentiel est prêt à payer un prix aléatoire  $X$  pour acheter le produit. Si  $X$  est supérieur ou égal à  $p$ , la vente a lieu le fabricant encaisse  $p$  euros ; sinon la vente n'a pas lieu et le fabricant n'encaisse aucun chiffre d'affaires.

On suppose que  $X$  admet une densité de probabilité  $f$  continue sur  $\mathbb{R}_+$  et une espérance finie. On note  $F$  la fonction de répartition et  $Z_p$  le chiffre d'affaires réalisé si le prix de son produit est fixé à la valeur  $p$  et on définit la fonction  $g$  sur  $\mathbb{R}_+$  par  $g(p) = E(Z_p)$ .

2. Le fabricant gagne  $p$  euros si la vente à lieu c'est à dire si  $X \geq p$ . Sinon, il ne gagne rien. Ainsi, clairement

$$Z_p = p \cdot \mathbb{1}_{(X \geq p)}$$

et

$$\begin{aligned} g(p) = E(Z_p) &= E(p\mathbb{1}_{(X \geq p)}) = pE(\mathbb{1}_{(X \geq p)}) \\ &= pP(X \geq p) = p(1 - P(X < p)) \\ &= p(1 - F(p)). \end{aligned}$$

3. La fonction de répartition  $F$  tend vers 1 en  $+\infty$  donc  $1 - F(p)$  tend vers 0 mais on a une forme indéterminée pour le produit avec  $p$ . Il faut donc regarder avec un peu de subtilité les quantités en présence :

$$\begin{aligned} g(p) = pP(X \geq p) &= p \int_p^{+\infty} f(t)dt = \int_p^{+\infty} ptf(t)dt \\ &\leq \int_p^{+\infty} tf(t)dt \end{aligned}$$

ce qui est l'intégrale correspondant au "reste" de l'espérance de  $X$ . En effet, on sait que  $X$  admet une espérance (et que  $f$  est nulle sur  $\mathbb{R}_-$ ) donc

$$\lim_{p \rightarrow +\infty} \int_0^p tf(t)dt = E(X) = \int_0^{+\infty} tf(t)dt$$

et

$$\int_p^{+\infty} tf(t)dt = E(X) - \int_0^p tf(t)dt \xrightarrow{p \rightarrow +\infty} 0.$$

Par encadrement on a bien que  $g(p) \rightarrow 0$ ,  $p \rightarrow \infty$ . Ce n'était pas une question si facile !

4. On calcule

$$g(0) = 0(1 - F(0)) = 0.$$

Par ailleurs,  $F$  étant primitive de  $f$ , elle est dérivable et, produit,  $g$  aussi. On a

$$g'(p) = 1 - F(p) - pf(p).$$

On ne peut pas dire grand chose sur cette fonction dérivée ; en particulier on ne peut pas trouver son signe pour en déduire des variations sur  $g$ . On raisonne donc avec des arguments plus fins et *qualitatifs*.

- Une fonction continue sur un intervalle fermé et borné atteint ses bornes et admet donc un maximum. Le problème ici est que  $\mathbb{R}_+$  n'est pas un intervalle fermé et borné... Mais l'idée est de s'y ramener en disant que "à l'infini" cela tend vers 0 donc ce n'est pas là que sera le maximum.



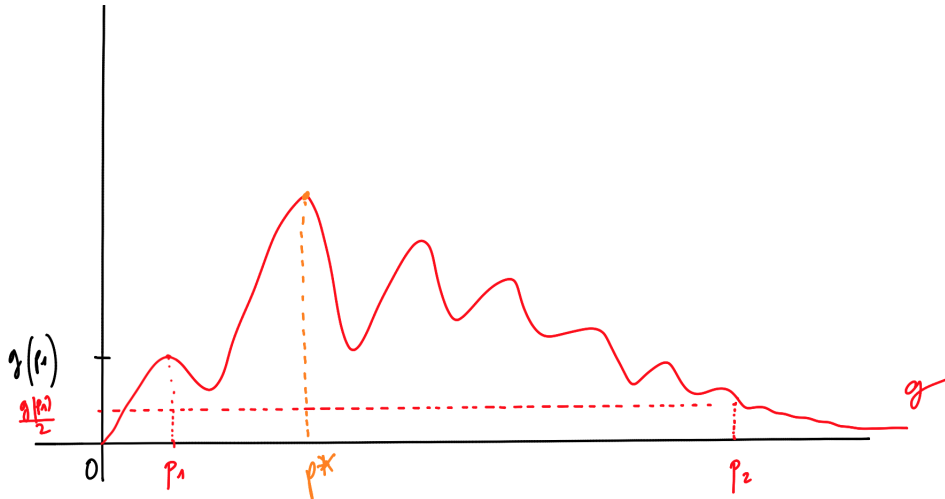
- $g$  est positive et non identiquement nulle. Il existe donc  $p_1 > 0$  tel que  $g(p_1) > 0$ .
- Comme  $g(p) \rightarrow 0, p \rightarrow +\infty$ , il existe une valeur  $p_2 > p_1$  telle que,

$$\forall p \geq p_2, \quad g(p) \leq \frac{g(p_1)}{2}.$$

- Mais alors,  $g$  est continue sur  $[0; p_2]$  et admet donc un maximum sur cet intervalle en un point noté  $p^*$ . En particulier,  $g(p^*) \geq g(p_1)$  et par construction, pour tout  $p \in [0; p_2]$ ,  $g(p) \leq g(p^*)$  et pour tout  $p \geq p_2$ ,  $g(p) \leq g(p_1)/2 < g(p^*)$ . On a bien

$$\forall p \in \mathbb{R}_+, \quad g(p) \leq g(p^*),$$

et  $g$  admet bien un maximum (strictement positif) sur  $\mathbb{R}_+$ .



5. Naturellement, comme  $g$  présente un extremum en  $p^*$ , la dérivée de  $g$  s'annule en  $p^*$  (c'est un *point critique*). On a donc

$$p^* f'(p^*) = 1 - F(p^*).$$

6. Dans le cas de la loi exponentielle de paramètre  $\lambda > 0$ , on a, pour  $p \geq 0$ ,

$$g(p) = pe^{-\lambda p}$$

et l'équation vérifiée par  $p^*$  est

$$e^{-\lambda p^*} = p^* \lambda e^{-\lambda p^*} \iff p^* = \frac{1}{\lambda}.$$

Le chiffre d'affaire maximal est donc l'évaluation du chiffre d'affaires  $g$  en cette valeur de  $p^*$ . Plus précisément, celui-ci vaut

$$g\left(\frac{1}{\lambda}\right) = \frac{1}{\lambda} e^{-\lambda \frac{1}{\lambda}} = \frac{1}{e\lambda}.$$

### Correction 3.-

On étudie le jeu suivant :  $X_1$  et  $X_2$  étant deux variables indépendantes de loi uniforme sur  $[0; 1]$ , un joueur observe d'abord  $X_1$ . S'il décide d'en rester là, il gagne la valeur observée. S'il décide de continuer, il observe et gagne  $X_2$ . On note  $G$  la v.a. correspondant au gain du joueur.

1. Si le joueur continue systématiquement, son gain est égal à  $X_2$  et  $E(G) = E(X_2) = 1/2$  car  $X_2 \hookrightarrow \mathcal{U}([0; 1])$ .

2. Dans un deuxième temps, il décide de continuer et d'observer  $X_2$  si et seulement si  $X_1 \leq s$ , où  $s \in [0; 1]$  est un seuil qu'il se fixe à l'avance.

- a. Si  $X_1 < s$  (et donc  $\mathbb{1}_{X_1 < s} = 1$  et  $\mathbb{1}_{X_1 \leq s} = 0$ ) alors on continue et le gain sera égal à  $X_2$ , c'est à dire à  $\mathbb{1}_{X_1 < s}X_2$ . Sinon, le gain sera égal à  $X_1$  ou à  $\mathbb{1}_{X_1 \leq s}X_1$ . On a bien

$$G = \mathbb{1}_{X_1 \geq s}X_1 + \mathbb{1}_{X_1 < s}X_2.$$

- b. Sans difficulté

```

1 import numpy.random as rd
2
3 def simul_gain():
4     s=int(input("s=?"))
5     G=rd.rand()
6     if G<s : # si X1 < s, on rejoue
7         G=rd.rand()
8     return G

```

- c. Dans ce cas  $E(G) = E(\mathbb{1}_{X_1 \geq s}X_1) + E(\mathbb{1}_{X_1 < s}X_2)$ . Mais,  $X_1$  et  $X_2$  étant indépendantes, le lemme des coalitions permet d'affirmer que  $\mathbb{1}_{X_1 < s}$  et  $X_2$  sont indépendantes. Donc

$$E(\mathbb{1}_{X_1 < s}X_2) = E(\mathbb{1}_{X_1 < s})E(X_2) = P(X_1 < s)E(X_2) = \frac{s}{2}.$$

De plus, par le théorème de transfert et le résultat établi en préambule,

$$\begin{aligned} E(\mathbb{1}_{X_1 \geq s}X_1) &= \int_s^1 t dt = \left[ \frac{t^2}{2} \right]_s^1 \\ &= \frac{1 - s^2}{2}. \end{aligned}$$

Au final,

$$E(G) = \frac{1 - s^2 + s}{2}.$$

- d. On regarde où la fonction  $s \mapsto 1 - s^2 + s$  est maximale. Sa dérivée vaut  $1 - 2s$  et s'annule en  $s^* = 1/2$  où on voit sans difficulté que  $E(G)$  est maximale. Ce résultat est d'ailleurs assez intuitif. Dans ce cas, l'espérance vaut

$$E(G) = \frac{5}{8} > \frac{1}{2},$$

ce qui est donc un peu mieux que la première stratégie.

3. Dans cette optique très avantageuse, on aurait  $G = \max(X_1, X_2)$ . On obtient alors  $E(G)$  en déterminant la loi du max, question très classique. On commence par la fonction de répartition

$$\begin{aligned} F_G(x) = P(\max(X_1, X_2) \leq x) &= P([X_1 \leq x] \cap [X_2 \leq x]) \\ &= P(X_1 \leq x)P(X_2 \leq x) \quad (\text{par indépendance de } X_1 \text{ et } X_2) \\ &= F_{X_1}(x)^2 \\ &= \begin{cases} 0, & \text{si } x < 0 \\ x^2, & \text{si } 0 \leq x < 1 \\ 1, & \text{si } x \geq 1 \end{cases} \end{aligned}$$

Cette fonction est bien la fonction de répartition d'une variable à densité (continue partout sur  $\mathbb{R}$  et de classe  $\mathcal{C}^1$  sur chacun des morceaux) donc  $G$  est bien une v.a à densité. Une densité est donnée par

$$f_G(x) = \begin{cases} 2x, & \text{si } 0 \leq x \leq 1 \\ 0, & \text{sinon} \end{cases}$$

Il suit que

$$E(G) = \int_0^1 2x^2 dx = \left[ \frac{2x^3}{3} \right]_0^1 = \frac{2}{3}.$$

C'est clairement la stratégie la plus avantageuse !