Devoir Surveillé 04

Le samedi 16 décembre 2023
durée 3h00

Documents écrits, électroniques, calculatrices et téléphones portables interdits
La plus grande attention sera apportée à la qualité de la rédaction, syntaxe et orthographe comprise.
Numérotez les copies et questions et soulignez ou encadrez les résultats.
Vous pouvez utiliser les résultats des questions en amont d'une question donnée en la référençant clairement
Les calculs, même longs et fastidieux doivent figurer sur la copie

Si au cours de l'épreuve, vous repérez ce qui vous semble être une erreur d'énoncé, vous le signalerez sur votre copie et poursuivrez la composition.

Problème I

Soit $0 < q < 1$ et $f : [0,+\infty[\to [0,+\infty[$, une fonction vérifiant l'inégalité :

$$\forall x,y \in [0,+\infty[\,,\ |f(x)-f(y)| \leq q.|x-y|.$$

On considère une suite u définie par la donnée de $u_0 \in [0,+\infty[$ et la relation de récurrence :

$$\forall n \in \mathbb{N},\ u_{n+1} = f(u_n).$$

Partie A

Un résultat général

Le but de cette partie est de démontrer que la suite u converge vers une certaine limite $\ell \in [0,+\infty[$, indépendante du point de départ u_0. On pose, pour $n \in \mathbb{N}$, $\delta_n = u_{n+1} - u_n$.

A.1. Justifier que $(u_n)_{n\in\mathbb{N}}$ est à valeurs dans $[0,+\infty[$ et que :

$$\forall n \in \mathbb{N},\ |\delta_{n+1}| \leq q.|\delta_n|.$$

A.2. Montrer que :

$$\forall n \in \mathbb{N},\ |\delta_n| \leq q^n.|\delta_0|.$$

A.3. Montrer que la série $\sum_{n\geq 0} \delta_n$ est convergente et en déduire que la suite $(u_n)_{n\geq 0}$ converge vers une certaine limite ℓ que l'on exprimera en fonction de u_0 et $\sum_{n=0}^{+\infty} \delta_n$.

A.4. Montrer que la limite ℓ vérifie $f(\ell) = \ell$.

Indication:On pourra commencer par démontrer que f est continue sur tout $[0,+\infty[$.

A.5. Montrer, en utilisant l'inégalité satisfaite par f que si ℓ et ℓ' vérifient $f(\ell) = \ell$ et $f(\ell') = \ell'$ alors $\ell = \ell'$.

A.6. Conclure quant au but de cete partie.

Partie B

Un exemple.

Soit $f : [0,+\infty[\to [0,+\infty[$ définie par $\forall x \in [0,+\infty[\,,\ f(x) = e^{-\frac{1}{2}x}$.

B.1. Montrer que f vérifie l'inégalité présentée en préambule de ce problème avec une valeur de q à préciser.

B.2. Un programme Python a calculé les 20 premières valeurs de la suite u en partant de $u_0 = 1$ et donne pour u_{20} et la u_{21} les valeurs :

`0.7034674218283558` et `0.7034674227340658`

Donner une valeur approchée de la solution de l'équation $f(x) = x$ d'inconnue $x \in [0,+\infty[$ en justifiant du niveau de précision.

Problème II

Le but de cet exercice est de construire deux variables aléatoires à densité X et Y, à valeurs dans $[0,+\infty[$ dont tous les moments existent (et sont finis) et qui, pourtant, n'ont pas même loi.

On note $(\Omega, \mathscr{T}, \mathscr{P})$ un espace probabilisé sur lequel toutes les variables aléatoires du texte sont définies. On note, pour une v.a. X à valeurs réelles admettant une espérance, $\mathbb{E}(X)$ la valeur de cette espérance.

Partie A

Quelques intégrales généralisées.

On définit la fonction $g : [0,+\infty[\to \mathbb{R}$ par :

$$\forall x \in [0,+\infty[,\, g(x) = \exp(-x^{\frac{1}{4}}) \sin(x^{\frac{1}{4}}) = e^{-x^{\frac{1}{4}}} . \sin(x^{\frac{1}{4}}).$$

A.1. Soit $n \in \mathbb{N}$. Montrer que les intégrales généralisées

$$\int_0^{+\infty} t^n e^{-t} \sin(t)\, dt \text{ et } \int_0^{+\infty} t^n e^{-t} \cos(t)\, dt$$

sont convergentes. On note S_n et C_n leurs valeurs respectives :

$$S_n = \int_0^{+\infty} t^n e^{-t} \sin(t)\, dt \text{ et } C_n = \int_0^{+\infty} t^n e^{-t} \cos(t)\, dt.$$

On note aussi $V_n = \begin{pmatrix} S_n \\ C_n \end{pmatrix}$.

A.2. Montrer que $S_0 = \frac{1}{2}$. On admettra que $C_0 = S_0 = \frac{1}{2}$.

A.3. Montrer que :

$$\forall n \in \mathbb{N}, \left\{ \begin{array}{ccccc} S_{n+1} & - & C_{n+1} & = & (n+1)S_n \\ S_{n+1} & + & C_{n+1} & = & (n+1)C_n \end{array} \right. .$$

A.4. En déduire que :

$$\forall n \in \mathbb{N}, V_{n+1} = (n+1)M.V_n \text{ où } M = \frac{1}{2}\begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}$$

puis que

$$\forall n \in \mathbb{N}, V_n = n! M^n . V_0.$$

A.5. Calculer M^4 et en déduire que :

$$\forall k \in \mathbb{N}, S_{4k+3} = 0.$$

A.6. En utilisant et justifiant le changement de variable $x = t^4$, montrer que :

$$\forall k \in \mathbb{N}, \int_0^{+\infty} x^k . g(x)\, dx = 0.$$

Partie B

Densités et variables aléatoires.

On considère T_1, T_2, T_3, T_4 des variables aléatoires indépendantes exponentielles de paramètre 1, *i.e.* de loi $\mathscr{E}(1)$.

B.1. Rappeler espérance, variance, et densité de T_1.

⋆⋆⋆

On rappelle la formule de convolution donnant une densité f_S de la somme $S = U + V$ de deux variables alétoires indépendantes U et V à densité, de densités respectives f_U et f_V :

$$\forall s \in \mathbb{R},\, f_S(s) = \int_{-\infty}^{+\infty} f_U(u).f_V(s-u)\, du$$

⋆⋆⋆

B.2. En utilisant (avec justification) cette formule de convolution, donner une densité de $U = T_1 + T_2$. Donner une densité de $V = T_3 + T_4$.

B.3. En utilisant (avec justification) la formule de convolution, donner une densité de $T = U + V$.

B.4. Montrer qu'une densité de $X = T^4$ est la fonction f vérifiant :

$$\forall x \in \mathbb{R},\, f(x) = \begin{cases} 0 & \text{si } x < 0 \\ c.\exp(-x^{\frac{1}{4}}) & \text{sinon} \end{cases}$$

où c est une constante réelle à préciser.

B.5.a. On prolonge g à tout $\mathbb{R}$ en posant $g(x) = 0$ pour $x < 0$. Donner –en justifiant– une valeur possible pour $\varepsilon > 0$ pour que la fonction $\phi = f + \varepsilon.g$ soit une densité de probabilité sur $\mathbb{R}$.

B.5.b. Soit Y une variable aléatoire réelle ayant ϕ pour densité. Justifier que pour tout $k \in \mathbb{N}$,

$$\mathbb{E}(Y^k) = \mathbb{E}(X^k)$$

B.5.c. Comparer $\mathbb{P}(X \leq \frac{1}{16})$ et $\mathbb{P}(Y \leq \frac{1}{16})$. Les variables aléatoires réelles X et Y ont-elles même loi ?

Partie C

Simulations informatiques

On suppose qu'en préambule du script dans lequel les fonctions à écrire sont intégrées, la bibliothèque `numpy.random` a été importée sous l'alias `nr`. On rappelle qu'alors la fonction `nr.rand` retourne un nombre (`float` aléatoire tiré uniformément dans l'intervalle $]0,1[$ et que les appels successifs à cette fonction sont autant de tirages indépendants.

C.1. Ecrire une fonction d'entête `Exp()` retournant un réel positif aléatoire suivant la loi exponentielle $\mathscr{E}(1)$.

C.2. Ecrire une fonction d'entête `X()` simulant la variable X de la partie/question B.4, retournant un réel positif aléatoire suivant la densité f.

Problème III

Ce problème comporte deux parties largement indépendantes.

Lorsque l'on effectue des sondages, de nombreux biais statistiques peuvent apparaitre : on peut par exemple avoir considéré un échantillon non-représentatif de la population, il peut y avoir un biais dans les réponses des personnes sondées... On va s'intéresser dans ce problème à ce que l'on appelle le biais par la taille : il provient du fait que si l'on choisit une personne au hasard dans la population, celle-ci a plus de chances de faire partie d'une catégorie nombreuse de la population.

Le biais par la taille est la source de nombreux « paradoxes » probabilistes, comme le fait que les gagnants du loto vivent en moyenne plus longtemps (parce que les gagnants sont ceux qui ont pu jouer au loto plus longtemps) ou le fait que vos amis ont en moyenne plus d'amis que vous (car les gens qui ont un très grand nombre d'amis font sûrement partie de vos amis). On verra ici comment formaliser le biais par la taille, et l'utiliser dans différents contextes.

Toutes les variables aléatoires intervenant dans le problème sont définies sur un espace probabilisé $(\Omega, \mathscr{T}, \mathbb{P})$. Pour toute variable aléatoire X, on notera $\mathbb{E}(X)$ son espérance (resp. $\mathbb{V}(X)$ sa variance) lorsqu'elles existent.

Pour deux événements $A, B \in \mathscr{T}$, on note $\mathbb{P}_B(A)$ la probabilité (conditionnelle) de A, sachant B, qui est définie sans ambiguité lorsque $\mathbb{P}(B) > 0$.

Partie A

Biais par la taille, exemples discrets.

A.1. On suppose que le nombre d'enfants dans une famille française prise au hasard est une variable aléatoire X. Pour connaître la loi de X, une idée serait d'interroger les élèves d'une école pour connaitre le nombre d'enfants dans leur famille.

On va voir que cette approche introduit un biais, en considérant une situation particulière. Supposons que X suive la loi binomiale de paramètres $n = 10$ et $p = 1/5$. On note $p_k = \mathbb{P}(X = k)$ pour $k \in \{0, 1, \ldots, 10\}$.

A.1.a. (i) Rappeler l'expression de p_k pour $k \in \{0, 1, \ldots, 10\}$. (ii) Que vaut $\mathbb{E}(X)$?

(iii) Donner $\mathbb{V}(X)$, et en déduire $\mathbb{E}\left(X^2\right)$

A.1.b. Soit M_k le nombre de familles à k enfants, $M = \sum_{k=0}^{10} M_k$ le nombre total de familles (donc $p_k = M_k/M$). Soit N_k le nombre total d'enfants (c'est-à-dire dans toute la population) qui font partie d'une famille à k enfants, et $N = \sum_{k=0}^{10} N_k$ le nombre total d'enfants de la population.

(i) Montrer que $N_k = kp_kM$. (ii) Montrer que $N/M = 2$.

(iii) Montrer que la proportion des enfants provenant d'une famille à k enfants est $p_k^* = kp_k/2$.

A.1.c. On choisit un enfant au hasard dans une école, à qui l'on demande combien d'enfants ses parents ont eu (elle incluse). On note Y ce nombre d'enfants.

(i) Pour tout entier k élément de $\{1, 2, \ldots, 10\}$, montrer que $\mathbb{P}(Y = k) = kp_k/2$.

(ii) Montrer que $\mathbb{E}(Y) = \mathbb{E}\left(X^2\right)/\mathbb{E}(X)$.

(iii) En déduire $\mathbb{E}(Y)$ et le comparer à $\mathbb{E}(X)$.

A.2. Soit n un entier naturel et soit X une variable aléatoire à valeurs dans $\{0, \ldots, n\}$, non identiquement nulle. Pour tout entier $i > 0$, on pose :

$$q_i = \frac{i}{\mathbb{E}(X)} \mathbb{P}(X = i).$$

A.2.a. Montrer que la suite finie $(q_i)_{1 \leq i \leq n}$ définit la loi de probabilité d'une variable aléatoire à valeurs dans $\{1, \ldots, n\}$.

On considère une telle variable aléatoire X^*, c'est-à-dire vérifiant que, pour tout i entier naturel non nul $P(X^* = i) = q_i$. On dit que X^* suit la loi de X biaisée par la taille.

A.2.b. Montrer que $\mathbb{E}(X^*) = \mathbb{E}\left(X^2\right)/\mathbb{E}(X)$.

A.2.c. En déduire que $\mathbb{V}(X) = \mathbb{E}(X)\left(\mathbb{E}(X^*) - \mathbb{E}(X)\right)$.

A.2.d. Conclure que $\mathbb{E}(X^*) \geq \mathbb{E}(X)$.

Partie B

Biais par la taille, propriétés

.

Dans cette partie, on démontre de nombreuses propriétés des variables aléatoires biaisées par la taille.

B.1. *Biais par la taille : le cas de variables à densité.*

Soit f une densité de probabilité sur $\mathbb{R}$, nulle sur $]-\infty,0[$ et soit X une variable aléatoire de densité f.

On suppose que X admet une espérance $\mathbb{E}(X)$ finie et strictement positive.

On définit la fonction g par :

$$\forall x \in \mathbb{R}, g(x) = \frac{x}{\mathbb{E}(X)} f(x).$$

B.1.a. Montrer que g définit une densité de variable aléatoire positive. Soit une variable aléatoire X^* dont la densité est g. On dit que X^* suit la loi de X biaisé par la taille.

B.1.b. Soit a un réel strictement positif.

(i) Montrer que la variable aléatoire aX possède pour densité $z \mapsto \frac{1}{a} f\left(\frac{z}{a}\right)$.

(ii) En déduire que $(aX)^*$ et $a \times X^*$ possèdent la même loi.

B.1.c. Une propriété importante. Soit $h : [0,+\infty[\to \mathbb{R}$ une fonction bornée et continue sauf éventuellement en un nombre fini de points. Montrer que $\mathbb{E}(Xh(X))$ est bien défini et que

$$\mathbb{E}(h(X^*)) = \frac{1}{\mathbb{E}(X)} \mathbb{E}(Xh(X))$$

⋆⋆⋆

On pose alors la définition suivante, *que la variable X soit à densité ou non* : si X est une variable aléatoire réelle positive d'espérance $\mathbb{E}(X)$ strictement positive, on dit qu'une variable aléatoire positive X^* *suit la loi de X biaisée par la taille* si on a

$$\mathbb{E}(h(X^*)) = \frac{1}{\mathbb{E}(X)} \mathbb{E}(Xh(X))$$

pour toute fonction $h : [0,+\infty[\to \mathbb{R}$ bornée et continue sauf éventuellement en un nombre fini de points, et, par extension naturelle, pour toute fonction h telle que $X.h(X)$ admette une espérance.

B.2. Dans cette question, on se fixe $f : \mathbb{R} \to \mathbb{R}$ et $g : \mathbb{R} \to \mathbb{R}$ deux fonctions croissantes. Soit X une variable aléatoire telle que les espérances $\mathbb{E}(f(X))$, $\mathbb{E}(g(X))$ et $\mathbb{E}(f(X).g(X))$ sont bien définies.

B.2.a. Montrer que quels que soient les réels x_1 et x_2, on a $(f(x_1) - f(x_2))(g(x_1) - g(x_2)) \geq 0$

B.2.b. Soient X_1, X_2 deux variables aléatoires indépendantes, de même loi que X. Montrer que

$$\mathbb{E}((f(X_1) - f(X_2))(g(X_1) - g(X_2))) = 2\mathbb{E}(f(X).g(X)) - 2\mathbb{E}(f(X)).\mathbb{E}(g(X))$$

B.2.c. En déduire que $\mathbb{E}(f(X).g(X)) \geq \mathbb{E}(f(X)).\mathbb{E}(g(X))$.

B.3. Dans cette question, on suppose que X est une variable aléatoire positive d'espérance strictement positive, et telle que $\mathbb{E}\left(X^{m+1}\right)$ existe pour un entier $m \geq 1$ donné. On suppose aussi que X^* est une variable aléatoire positive qui *suit la loi de X biaisée par la taille*.

B.3.a. Soit p un entier naturel tel que $1 \leq p \leq m$.

(i) Montrer que pour tout réel $x \geq 0$, on a $0 \leq x^p \leq 1 + x^{m+1}$.

(ii) Montrer que $\mathbb{E}(X^p)$ existe.

B.3.b. Montrer que $\mathbb{E}\left(X^{m+1}\right) \geq \mathbb{E}(X).\mathbb{E}(X^m)$.

B.3.c. En déduire que $\mathbb{E}((X^*)^m) \geq \mathbb{E}(X^m)$.

B.4. Pour A un événement, on note $1\!\!1_A$ la variable aléatoire définie par $1\!\!1_A(\omega) = 1$ si $\omega \in A$ et $1\!\!1_A(\omega) = 0$ sinon. Pour tout t réel, on définit la fonction g_t par $\forall x \in \mathbb{R}$, $g_t(x) = 1\!\!1_{]t,+\infty[}(x)$.
B.4.a. Montrer que la fonction g_t est croissante sur $\mathbb{R}$.
B.4.b. Soit X une variable aléatoire positive, admettant une espérance. Montrer que pour tout t réel, $\mathbb{E}(Xg_t(X))$ est bien défini et que :

$$\mathbb{E}(Xg_t(X)) \geq \mathbb{E}(X)\,\mathbb{P}(X > t)\,.$$

B.4.c. Montrer que pour tout t réel, $\mathbb{P}(X^* > t) \geq \mathbb{P}(X > t)$.
B.5. Soit $X_1, \ldots, X_n$ des variables aléatoires positives, indépendantes, non nécessairement de même loi. On suppose qu'elles admettent toutes une espérance strictement positive, et on note $\mu_i = \mathbb{E}(X_i)$. De plus, on pose $\mu = \sum_{i=1}^n \mu_i$, et $S_n = \sum_{i=1}^n X_i$.
B.5.a. Donner $\mathbb{E}(S_n)$.
B.5.b. Soit J une variable aléatoire à valeurs dans $\{1, \ldots, n\}$, de loi définie par $\forall k \in \{1, \ldots, n\}, \mathbb{P}(J = k) = \mu_k/\mu$. Quelle est la loi de J si les variables aléatoires X_i sont de même loi ?

⋆⋆⋆

On considère $X_1^*, \ldots, X_n^*$ des variables aléatoires indépendantes, indépendantes de $X_1, \ldots, X_n$, telles que, pour tout entier i tel que $1 \leq i \leq n, X_i^*$ suive la loi de X_i biaisée par la taille.

On suppose que J à valeurs dans $\{1, \ldots, n\}$, est indépendante de $X_1, X_1^*, \ldots, X_n, X_n^*$, de loi définie par

$$\forall k \in \{1, \ldots, n\}, \mathbb{P}(J = k) = \mu_k/\mu.$$

On considère la variable aléatoire $X_J = \sum_{j=1}^n X_j 1\!\!1_{\{J=j\}}$ et on définit $T_n = S_n - X_J + X_J^*$

Autrement dit, on choisit un indice aléatoire J et, dans la somme $\sum_{i=1}^n X_i$, on remplace X_J par X_J^*
B.5.c. Soit $h : [0, +\infty[\to \mathbb{R}$ bornée et continue sauf éventuellement en un nombre fini de points.

(i) Montrer que :

$$h(T_n) = \sum_{i=1}^n h(T_n)\, 1\!\!1_{\{J=i\}} = \sum_{i=1}^n h(S_n - X_i + X_i^*)\, 1\!\!1_{\{J=i\}}.$$

(ii) En déduire que :

$$\mathbb{E}(h(T_n)) = \sum_{i=1}^n \mathbb{P}(J = i)\,\mathbb{E}(h(S_n - X_i + X_i^*))\,.$$

B.5.d. Pour $i \in \{1, \ldots, n\}$, montrer que pour tout réel s,

$$\mathbb{E}(h(s + X_i^*)) = \frac{1}{\mu_i}\mathbb{E}(X_i.h(s + X_i))\,.$$

On *admettra* qu'on en déduit l'égalité $\mathbb{E}(h(S_n - X_i + X_i^*)) = \frac{1}{\mu_i}\mathbb{E}(X_i h(S_n))$
B.5.e. En déduire que $\mathbb{E}(h(T_n)) = \mathbb{E}(S_n h(S_n))\,/\mathbb{E}(S_n)$.
B.5.f. Conclure que T_n suit la loi de S_n biaisée par la taille.

PYTHON
AGRO-VETO
2020

Listes

`[]`	Créer une liste vide
`[a]*n`	Créer une liste avec n fois l'élément `a`
`L.append(a)`	Ajoute l'élément `a` à la fin de la liste `L`
`L1 + L2`	Concatène les deux listes `L1` et `L2`
`len(L)`	Renvoie le nombre d'éléments de la liste `L`
`L.pop(k)`	Renvoie l'élément d'indice k de `L` et l'enlève de `L`
`L.remove(a)`	Enlève une fois la valeur `a` de la liste `L`
`max(L)`	Renvoie le plus grand élément de la liste `L`
`min(L)`	Renvoie le plus petit élément de la liste `L`
`sum(L)`	Renvoie la somme de tous les éléments de la liste `L`

Numpy

`import numpy as np`

`np.array()`	Transforme une liste en matrice `numpy`
`np.linspace(a,b,n)`	Crée une matrice ligne de n valeurs uniformément réparties entre a et b (inclus)
`np.zeros([n,m])`	Crée la matrice nulle de taille $n \times m$
`np.eye(n)`	Crée la matrice identité de taille n
`np.diag(L)`	Crée la matrice diagonale dont les termes diagonaux sont les éléments de la liste `L`
`np.transpose(M)`	Renvoie la transposée de M
`np.dot(M,P)`	Renvoie le produit matriciel MP
`np.sum(M)`	Renvoie la somme de tous les éléments de M
`np.prod(M)`	Renvoie le produit de tous les éléments de M
`np.max(M)`	Renvoie le plus grand élément de M
`np.min(M)`	Renvoie le plus petit élément de M
`np.shape(M)`	Renvoie dans un couple le format de la matrice M
`np.size(M)`	Renvoie le nombre d'éléments de M

Numpy.linalg

`import numpy.linalg as la`

`la.inv(M)`	Renvoie l'inverse de la matrice M si elle est inversible
`la.eigvals(M)`	Renvoie la liste des valeurs propres de M
`la.eig(M)`	Renvoie un couple `L,P` où `L` est la liste des valeurs propres de M et `P` la matrice de passage associée
`la.matrix_rank(M)`	Renvoie le rang de M

Random

`import numpy.random as rd`

`rd.rand()`	Simule une réalisation d'une variable $X \hookrightarrow \mathcal{U}([0,1])$
`rd.randint(a,b)`	Simule une réalisation d'une variable $X \hookrightarrow \mathcal{U}([\![a,b[\![)$
`rd.gauss(0,1)`	Simule une réalisation d'une variable $X \hookrightarrow \mathcal{N}(0,1)$
`rd.choice(L)`	Choisit aléatoirement un élément de la liste `L`

Math

`import numpy as np`

`np.atan(x)`	Renvoie $\arctan(x)$
`np.floor(x)`	Renvoie $\lfloor x \rfloor$
`np.factorial(n)`	Renvoie $n!$ si $n \in \mathbb{N}$
`np.sqrt(x)`	Renvoie $\sqrt{x}$ si $x \geq 0$
`np.log(x)`	Renvoie $\ln(x)$ si $x > 0$
`np.exp(x)`	Renvoie e^x

Logique

`a == b`	Teste l'égalité « $a = b$ »
`a != b`	Teste « $a \neq b$ »
`a < b`	Teste « $a < b$ »
`a <= b`	Teste « $a \leq b$ »
`a > b`	Teste « $a > b$ »
`a >= b`	Teste « $a \geq b$ »
`not A`	Renvoie la négation de A
`A and B`	Renvoie « A et B »
`A or B`	Renvoie « A ou B »
`True`	Constante booléenne « Vrai »
`False`	Constante booléenne « Faux »

Matplotlib.pyplot

`import matplotlib.pyplot as plt`

`plt.plot(X,Y,'+-r')`	Génère la courbe des points définis par les listes `X` et `Y` (abscisses et ordonnées) avec les options : • symbole : `'.'` point, `'o'` rond, `'h'` hexagone, `'+'` plus, `'x'` croix, `'*'` étoile, ... • ligne : `'-'` trait plein, `'--'` pointillé, `'-.'` alterné, ... • couleur : `'b'` **bleu** , `'r'` **rouge** , `'g'` **vert** , `'c'` **cyan** , `'m'` **magenta** , `'k'` **noir** , ...
`plt.bar(X,Y)`	Génère l'histogramme des points définis par les listes `X` et `Y` (abscisses et ordonnées)
`plt.axis('equal')`	Rend le repère orthonormé
`plt.xlim(xmin,xmax)`	Fixe les bornes de l'axe des abscisses
`plt.ylim(ymin,ymax)`	Fixe les bornes de l'axe des ordonnées
`plt.show()`	Affiche le graphique

Cette liste est non exhaustive. Les candidats sont libres d'utiliser les commandes de leur choix.

Correction DS 04

Correction Ex.–1

Partie A

A.1. La suite $(u_n)_{n\in\mathbb{N}}$ est à valeurs dans $[0,+\infty[$ car l'intervalle $[0,+\infty[$ est stable par f et $u_0 \in [0,+\infty[$. Pour $n \in \mathbb{N}$, on a, en utilisant la relation de récurrence, puis l'inégalité satisfaite par f avec $x = u_{n+1}$ et $y = u_n$:

$$|\delta_{n+1}| = |u_{n+2} - u_{n+1}| = |f(u_{n+1}) - f(u_n)| \leq q.|u_{n+1} - u_n| = q.|\delta_n|.$$

A.2. Montrons par une récurrence de type géométrique que :

$$\forall n \in \mathbb{N},\ (H_n) : |\delta_n| \leq q^n.|\delta_0|.$$

1. (H_0) est $|\delta_0| \leq q^0.|\delta_0|$ qui suit du fait que $q^0 = 1$.
2. Supposons que l'on dispose de (H_n) pour un certain $n \in \mathbb{N}$, on a alors :

$$|\delta_{n+1}| \leq q.|\delta_n| \leq q.q^n.|\delta_0| = q^{n+1}.|\delta_0|$$

ce qui est (H_{n+1}).

A.3. La série $\sum_{n\geq 0} |\delta_n|$ à termes positifs vérfie :

$$\forall n \in \mathbb{N},\ |\delta_n| \leq q^n.|\delta_0|$$

Or $\sum_{n\geq 0} |\delta_0|.q^n$ converge c'est une suite géométrique de raison q avec $0 < q < 1$ et donc, par le théorème de comparaison pour les séries à termes positifs, $\sum_{n\geq 0} |\delta_n|$ est convergente, donc $\sum_{n\geq 0} \delta_n$ est ACV, donc, par le théorème $ACV \Rightarrow CV$, $\sum_{n\geq 0} \delta_n$ CV. Pour alléger les écritures, on pose ensuite

$$S = \sum_{n=0}^{+\infty} \delta_n.$$

On en déduit que la suite des sommes partielles de cette série converge vers S :

$$\sum_{n=0}^{N-1} \delta_n \overset{N\to+\infty}{\to} S$$

Or, par télescopie $\sum_{n=0}^{N-1} \delta_n = u_N - u_0$ et donc

$$u_N = u_0 + \sum_{n=0}^{N-1} \delta_n \overset{N\to+\infty}{\to} u_0 + S$$

On a donc obtenu que la suite $(u_n)_{n\geq 0}$ converge vers une certaine limite $\ell = u_0 + \sum_{n=0}^{+\infty} \delta_n$.

A.4. La fonction f est continue sur tout $[0,+\infty[$. En effet, soit $x_0 \in [0,+\infty[$. Par l'inégalité supposée satisfaite par f,

$$\forall x \in [0,+\infty[\,,\, 0 \leq |f(x)-f(x_0)| \leq q.|x-x_0|.$$

Lorsque $x \to x_0$, le théorème des « gendarmes », implique que $|f(x)-f(x_0)| \to 0$, *i.e.* $f(x) \to f(x_0)$. Ceci montre que f est continue en x_0. Comme $x_0 \in [0,+\infty[$ est quelconque, f est continue sur $[0,+\infty[$.

Remarquons que $\ell \geq 0$ en passant à la limite la famille d'égalités larges :

$$\forall n \in \mathbb{N}, u_n \geq 0.$$

Maintenant, en écrivant la récurrence, on a

$$\forall n \in \mathbb{N}, u_{n+1} = f(u_n).$$

En passant cette famille d'égalités à la limite lorsque $n \to +\infty$:

— $u_{n+1} \to \ell$ (suite décalée)

— $f(u_n) \to f(\ell)$ car f est continue en ℓ (caractérisation séquentielle de la continuité)

Par unicité de la limite, il vient $\ell = f(\ell)$.

A.5. Si ℓ et ℓ' vérifient $f(\ell) = \ell$ et $f(\ell') = \ell'$ alors

$$|\ell-\ell'| = |f(\ell)-f(\ell')| \leq q.|\ell-\ell'|$$

et donc

$$(1-q).|\ell-\ell'| \leq 0$$

Comme $1-q > 0$, il vient $|\ell-\ell'| \leq 0$ et donc $|\ell-\ell'| = 0$ et donc $\ell = \ell'$.

A.6. On a donc démontré que si la fonction $f : [0,+\infty[\to [0,+\infty[$ vérifie l'inégalité présentée en préambule alors,

1. La fonction f admet un unique point fixe $\ell \in [0,+\infty[$, *i.e.* il existe un unique nombre réel $\ell \in [0,+\infty[$ vérifiant $f(\ell) = \ell$ et
2. toute suite u vérifiant : $u_0 \in I, \forall n \in \mathbb{N}, u_{n+1} = f(u_n)$ converge vers ℓ.

Partie B

Un exemple.

Soit $f : [0,+\infty[\to [0,+\infty[$ définie par $\forall x \in [0,+\infty[\,, f(x) = e^{-\frac{1}{2}x}$.

B.1. L'exponentielle étant positive, $f : [0,+\infty[\to [0,+\infty[$. Par ailleurs, f est de classe $\mathscr{C}^1$ sur $[0,+\infty[$ avec

$$\forall x \in [0,+\infty[\,, f'(x) = -\frac{1}{2}\underbrace{e^{-\frac{1}{2}x}}_{0\leq . \leq 1}.$$

On a donc

$$\forall x \in [0,+\infty[\,, |f'(x)| \leq \frac{1}{2}.$$

Si maintenant x et y sont dans $[0,+\infty[$, par le TAF, il existe c entre x et y tel que

$$f(x)-f(y) = f'(c).(x-y)$$

En passant à la valeur absolue et en utilisant le fait que $|f'(c)| \leq \frac{1}{2}$, il vient :

$$\forall x,y \in [0,+\infty[\,, |f(x)-f(y)| \leq \frac{1}{2}.|x-y|.$$

La fonction f vérifie l'inégalité demandée en préambule de ce problème avec une valeur $q = \frac{1}{2}$.

B.2. Les valeurs u_{20} et u_{21} sont

$$\texttt{0.7034674218283558} \text{ et } \texttt{0.7034674227340658}$$

Une valeur approchée de ℓ semble être 0.70346742 à 10^{-8} près. Pourquoi cette précision (ou un peu moins 10^{-6}, par exemple) ? Deux raisons possibles :

1. En reprenant les majorations faites en partie A avec $q = \frac{1}{2}$, on peut voir que

$$|u_N - \ell| = |\sum_{n=N}^{+\infty} \delta_n| \leq |u_1 - u_0| \sum_{n=N}^{+\infty} q^n = \frac{|u_1 - u_0|}{1-q}.q^N$$

   Comme ici $|u_1 - u_0| \leq 1$, que $q = \frac{1}{2}$, que $n = 21$, l'erreur commise par la deuxième valeur est $\leq 2^{-20} \simeq 10^{-6}$ (car $10^3 = 1000 \simeq 2^{10}$).
2. En comprenant comment est organisée la suite (u_n), on comprend que du fait de la décroissance de la fonction f, les deux sous-suites $(u_{2k+1})_{k\in\mathbb{N}}$ et $(u_{2k})_{k\in\mathbb{N}}$ sont adjacentes et donc la limite ℓ est comprise entre deux valeurs consécutives de la suite u_n...

Correction Ex.–2

Partie A

Quelques intégrales généralisées.

On définit la fonction $g : [0,+\infty[\to \mathbb{R}$ par :

$$\forall x \in [0,+\infty[,\, g(x) = \exp(-x^{\frac{1}{4}})\sin(x^{\frac{1}{4}}) = e^{-x^{\frac{1}{4}}}.\sin(x^{\frac{1}{4}})$$

A.1. Soit $n \in \mathbb{N}$. Soient les intégrales généralisées

$$\int_0^{+\infty} t^n e^{-t}\sin(t)\,dt \text{ et } \int_0^{+\infty} t^n e^{-t}\cos(t)\,dt.$$

Traitons uniquement le cas de la première, celui de la deuxième étant similaire en remplaçant sin par cos.

— L'intégrande $t \mapsto t^n e^{-t}\sin(t)$ est continue sur $[0,+\infty[$, l'intégrale généralisée est donc légitime à considérer, elle a une singularité en $+\infty$.

— Lorsque $t \to +\infty$, comme $|\sin(t)| \leq 1$,

$$\forall t \in [0,+\infty[,\, 0 \leq |t^n e^{-t}\sin(t)| \leq |t^n e^{-t}| = t^n e^{-t}$$

et, comme par croissance comparées $t^2.t^n e^{-t} \to 0$, on en déduit (par utilisation de la définition de limite avec $\varepsilon = 1$) qu'il existe $A > 0$ tel que

$$\forall t \geq A,\, 0 \leq |t^n e^{-t}\sin(t)| \leq \frac{1}{t^2}$$

— Comme $\int_A^{+\infty} \frac{1}{t^2}\,dt$ est convergente (elle vaut, au sens des limites, $\left[-\frac{1}{t}\right]_A^{+\infty} = \frac{1}{A}$), par le théorème de comparaison pour les intégrales à intégrande positive, l'intégrale $\int_0^{+\infty} |t^n e^{-t}\sin(t)|\,dt$ est convergente, ce qui signifie que $\int_0^{+\infty} t^n e^{-t}\sin(t)\,dt$ est absolument convergente et donc qu'elle est convergente par le théorème $ACV \Rightarrow CV$.

A.2. Pour mener le calcul de S_0,on peut

1. Soit faire deux intégrations parties en dérivant à chaque fois la fonction trigonométrique : (on laisse ceci au lecteur, ça marche),
2. Soit passer en complexes et calculer directement

$$C_0 + i.S_0 = \int_0^{+\infty} e^{-t}e^{it}\,dt = \int_0^{+\infty} e^{-t+it}\,dt = \frac{1}{-1+i}\left[e^{-t+it}\right]_0^{+\infty} = \frac{1}{1-i} = \frac{1+i}{2}$$

   et finalement $C_0 = S_0 = \frac{1}{2}$.

A.3. Soit $n \in \mathbb{N}$, on a par intégration par parties généralisée :

$$\begin{aligned}(n+1)S_n &= \int_0^{+\infty} \underbrace{(n+1)t^n}_{u'(t)} \underbrace{e^{-t}.\sin(t)}_{v(t)} \, dt \\ &= \left[t^{n+1}e^{-t}\sin(t)\right]_0^{+\infty} - \int_0^{+\infty} t^{n+1}e^{-t}.(-\sin(t)+\cos(t)) \, dt\end{aligned}$$

On a posé dans ce calcul :

$$\forall t \in [0,+\infty[\,,\, u(t) = t^{n+1},\, v(t) = e^{-t}\sin(t)$$

Les fonctions u et v ainsi définies sont de classe $\mathscr{C}^1$ sur $[0,+\infty[$,

$$\forall t \in [0,+\infty[\,,\, u'(t) = (n+1)t^n,\, v'(t) = e^{-t}(-\sin(t)+\cos(t)).$$

Par ailleurs les intégrales intervenant dans la formule sont toutes connues pour être convergentes et enfin le crochet (au sens des limites) vaut 0 du fait de t^{n+1} en 0 et par croissance comparée en $+\infty$.

On a donc

$$\begin{aligned}(n+1)S_n &= \int_0^{+\infty} t^{n+1}e^{-t}.(+\sin(t)-\cos(t)) \, dt \\ &= S_{n+1} - C_{n+1}\end{aligned}$$

Pour C_{n+1} le calcul est le même avec les fonctions u et v définies par :

$$\forall t \in [0,+\infty[\,,\, u(t) = t^{n+1},\, v(t) = e^{-t}\cos(t)$$

et vérifiant

$$\forall t \in [0,+\infty[\,,\, u'(t) = (n+1)t^n,\, v'(t) = -e^{-t}(\sin(t)+\cos(t)).$$

Cela donne

$$\begin{aligned}(n+1)C_n &= \int_0^{+\infty} t^{n+1}e^{-t}.(+\sin(t)+\cos(t)) \, dt \\ &= S_{n+1} + C_{n+1}\end{aligned}$$

Ces deux identités se réécrivent :

$$\forall n \in \mathbb{N}, \left\{\begin{array}{ccccc} S_{n+1} & - & C_{n+1} & = & (n+1)S_n \\ S_{n+1} & + & C_{n+1} & = & (n+1)C_n \end{array}\right. .$$

A.4. Posons $N = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$. L'identité se réécrit matriciellement en :

$$\forall n \in \mathbb{N}, N.V_{n+1} = (n+1)V_n$$

En posant $M = \dfrac{1}{2}\begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}$, on vérifie facilement que $M.N = I_2$ et donc en inversant la relation précédente,

$$\forall n \in \mathbb{N}, V_{n+1} = (n+1)M.V_n$$

Par une récurrence assez triviale (**A Rédiger ! ! sur la copie**), on obtient que

$$\forall n \in \mathbb{N}, V_n = n!M^n.V_0.$$

A.5. On a
$$M^2 = \frac{1}{2}\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, M^4 = -\frac{1}{4}.I_2 = -\frac{1}{4}.N.M \text{ et } M^3 = -\frac{1}{4}.N.$$
Remarquons de plus que
$$M^3.V_0 = -\frac{1}{4}.N.V_0 = -\frac{1}{4}\begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

On a donc, pour $k \in \mathbb{N}$,

$$V_{4k+3} = (4k+3)!M^{4k}M^3.V_0 = \frac{(-1)^{k+1}(4k+3)!}{4^{k+1}}\begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

En extrayant la première composante de cette identité vectorielle, $S_{4k+3} = 0$ et donc :

$$\forall k \in \mathbb{N}, S_{4k+3} = 0.$$

Remarque :on a aussi la valeur de C_{4k+3} et en travaillant de la même manière avec $n = 4k$, $n = 4k+1$, $n = 4k+2$, on peut disposer de toutes les valeurs des C_n, S_n. Elles ne nous servent à rien dans la suite.
A.6. Posons, pour $k \in \mathbb{N}$,

$$I_k = \int_0^{+\infty} x^k.g(x)\ dx = \int_0^{+\infty} x^k.e^{-x^{\frac{1}{4}}} \sin(x^{\frac{1}{4}})\ dx$$

La fonction $t \mapsto t^4 = x$ est de classe $\mathscr{C}^1$ sur $[0, +\infty[$, *strictement monotone*, lorsque $t = 0$, $x = 0$, lorsque $t \to +\infty$, $x \to +\infty$ et enfin $dx = 4t^3\ dt$, par changement de variable généralisé, on en déduit que

$$\int_0^{+\infty} t^{4k} \sin((t^4)^{\frac{1}{4}}) \exp(-(t^4)^{\frac{1}{4}}) 4t^3\ dt = \int_0^{+\infty} x^k. \sin(x^{\frac{1}{4}}) \exp(-x^{\frac{1}{4}})\ dx = I_k$$

On reconnait donc que $I_k = 4S_{4k+3} = 0$ et finalement :

$$\forall k \in \mathbb{N}, \int_0^{+\infty} x^k.g(x)\ dx = 0.$$

Partie B

Densités et variables aléatoires.

On considère T_1, T_2, T_3, T_4 des variables aléatoires indépendantes, à densité de loi $\mathscr{E}(1)$.
B.1. Avec $\lambda = 1$:

$$\mathbb{E}(T_1) = \frac{1}{\lambda} = 1, \mathbb{V}(T_1) = \frac{1}{\lambda^2} = 1$$

et pour densité f_1 avec

$$\forall t \in \mathbb{R}, f_1(t) = \begin{cases} e^{-t} & \text{si } t \geq 0 \\ 0 & \text{sinon} \end{cases} = e^{-t}.1\!\!1_{\{t \geq 0\}}$$

B.2. On peut utiliser la formule de convolution pour donner une densité f_U de $U = T_1 + T_2$ car ces deux v.a. à densité sont indépendantes (hypothèses).

On a, pour $u \in \mathbb{R}$,

$$f_U(u) = \int_{-\infty}^{+\infty} f_1(t).f_1(u-t)\, dt = e^{-u} \int_{-\infty}^{+\infty} 1\!\!1_{\{t \geq 0\}}.1\!\!1_{\{u-t \geq 0\}}\, dt$$

En distinguant les cas $u < 0$ et $u \geq 0$, il vient :

— Pour $u < 0$, $f_U(u) = \int_{-\infty}^{+\infty} 0\, dt = 0$;

— Pour $u \geq 0$, $f_U(u) = e^{-u} \displaystyle\int_{-\infty}^{+\infty} 1\!\!1_{\{0 \leq t \leq u\}}\, dt = u.e^{-u}$;

et donc

$$\forall u \in \mathbb{R},\ f_U(u) = u.e^{-u} 1\!\!1_{\{u \geq 0\}}$$

Une densité de $V = T_3 + T_4$ est aussi f_U car T_3 et T_4 ont même loi que T_1 et T_2 et sont indépendantes.
B.3. On peut utiliser la formule de convolution pour donner une densité f_U de $T = U + V$ car ces deux v.a. à densité sont indépendantes (par le lemme des coalitions U est fonction de T_1, T_2, V de T_3 T_4 avec T_1, T_2, T_3, T_4 indépendantes). On a, pour $t \in \mathbb{R}$,

$$f_T(t) = \int_{-\infty}^{+\infty} f_U(u).f_1(t-u)\, du = e^{-t} \int_{-\infty}^{+\infty} u(t-u) 1\!\!1_{\{u \geq 0\}}.1\!\!1_{\{t-u \geq 0\}}\, du$$

En distinguant les cas $t < 0$ et $t \geq 0$, il vient :

— Pour $t < 0$, $f_T(t) = \int_{-\infty}^{+\infty} 0\, du = 0$;

— Pour $t \geq 0$, $f_T(t) = e^{-t} \displaystyle\int_{-\infty}^{+\infty} u.(t-u) 1\!\!1_{\{0 \leq u \geq t\}}\, du = \frac{1}{6}.t^3.e^{-t}$;

et donc

$$\forall t \in \mathbb{R},\ f_T(t) = \frac{1}{6}.t^3.e^{-t} 1\!\!1_{\{t \geq 0\}}.$$

B.4. Soit $X = T^4$. Calculons une densité de X par la méthode de la formule de transfert générique. Soit $h : \mathbb{R} \to \mathbb{R}$ une fonction telle que $h(X)$ admette une espérance. Alors $h(T^4)$ admet une espérance et donc, par la formule de transfert pour T (avec intégrale ACV) :

$$\begin{aligned}
\mathbb{E}(h(X)) &= \mathbb{E}(h(T^4)) = \int_0^{+\infty} h(t^4)\frac{1}{6}.t^3.e^{-t}\, dt \\
&= \frac{1}{24} \int_0^{+\infty} h(t^4) e^{-(t^4)^{\frac{1}{4}}}\ 4t^3 dt \\
&= \frac{1}{24} \int_0^{+\infty} h(x) e^{-x^{\frac{1}{4}}}\ dx.
\end{aligned}$$

Pour la dernière ligne, on a effectué le changement de variable $x = t^4$ déjà utilisé en A.6, ce qui conserve la nature ACV des intégrales.

Une densité de X est donc la fonction f vérifiant :

$$\forall x \in \mathbb{R},\, f(x) = \begin{cases} 0 & \text{si } x < 0 \\ c.\exp(-x^{\frac{1}{4}}) & \end{cases} \quad \text{où } c = \frac{1}{24}.$$

B.5.a. En choisissant $\varepsilon = \frac{1}{24}$, on voit que la fonction $\phi = f + \varepsilon.g$ est une densité de probabilité sur $\mathbb{R}$. En effet,

1. pour $x < 0$, $\phi(x) + 0$ et pour tout $x \in [0, +\infty[$,

$$\phi(x) = f(x) + \frac{1}{24}g(x) = \frac{1}{24}\exp(-x^{\frac{1}{4}}).(1 + \sin(x^{\frac{1}{4}}) \geq 0$$

   car $1 + \sin(y) \geq 0$.

2. $$\int_{-\infty}^{+\infty} \phi(x)\, dx = \int_0^{+\infty} f(x)\, dx + \frac{1}{24}\int_0^{+\infty} g(x)\, dx = 1 + 0 = 1$$

B.5.b. Soit Y une variable aléatoire réelle ayant ϕ pour densité. Pour tout $k \in \mathbb{N}$,

$$\mathbb{E}(Y^k) = \int_{-\infty}^{+\infty} x^k \phi(x)\, dx = \int_0^{+\infty} x^k.f(x)\, dx + \frac{1}{24}\underbrace{\int_0^{+\infty} x^k.g(x)\, dx}_{=0} = \mathbb{E}(X^k)$$

B.5.c. On a

$$\mathbb{P}(Y \leq \frac{1}{16}) - \mathbb{P}(X \leq \frac{1}{16}) = \int_0^{\frac{1}{6}} \phi(x)\, dx - \int_0^{\frac{1}{6}} f(x)\, dx = \varepsilon \int_0^{\frac{1}{6}} g(x)\, dx.$$

Si x varie sur l'intervalle $\left]0, \frac{1}{6}\right]$, $x^{\frac{1}{4}}$ varie sur $\left]0, \frac{1}{2}\right]$ et donc $g(x) = e^{-x^{\frac{1}{4}}}\sin(x^{\frac{1}{4}})$ est > 0. Son intégrale (de fonction continue) est donc > 0 et on a

$$\mathbb{P}(Y \leq \frac{1}{16}) - \mathbb{P}(X \leq \frac{1}{16}) > 0,\ \textit{i.e.}\ \mathbb{P}(Y \leq \frac{1}{16}) > \mathbb{P}(X \leq \frac{1}{16})$$

Les variables aléatoires réelles X et Y n'ont donc pas même loi.

Partie C
Simulations informatiques

C.1. C'est un classique !

```
def Exp() :
    return -np.log(nr.rand())
```

C.2. C'est une simple formule à implémenter : faire attention à faire 4 tirages indépendants !

```
def X() :
    t1 = Exp();     t2 = Exp()
    t3 = Exp();     t4 = Exp()
    return (t1 + t2 + t3 + t4)**4
```

Correction Ex.–3

Partie A

Biais par la taille, exemples discrets.

A.1. Supposons que X suive la loi binomiale de paramètres $n = 10$ et $p = 1/5$. On note $p_k = \mathbb{P}(X = k)$ pour $k \in \{0, 1, \ldots, 10\}$.

A.1.a. (i) On a, pour $k \in \{0, 1, \ldots, n\}$, $p_k = \binom{n}{k} p^k (1-p)^{n-k}$ (ii) et $\mathbb{E}(X) = np = 2$.

(iii) Finalement $\mathbb{V}(X) = n.p.(1-p) = \frac{8}{5}$ et donc $\mathbb{E}\left(X^2\right) = \mathbb{V}(X) + \mathbb{E}(X)^2 = \frac{28}{5}$.

A.1.b.

(i) Si le nombre de familles à k enfants est M_k, le nombre d'enfants apprantenant à une telle famille est donc $N_k = kM_k$ et donc $N_k = kp_kM$, vu que $p_k = M_k/M$, comme souligné dans le texte (c'est un élément de la modélisation).

(ii) On a donc

$$N = \sum_{k=0}^{10} N_k = \sum_{k=0}^{10} kp_kM = M.\mathbb{E}(X) = 2M.$$

On a donc $N/M = 2$.

(iii) La proportion des enfants provenant d'une famille à k enfants est donc :

$$p_k^* := \frac{N_k}{N} = \frac{kp_kM}{N} = kp_k/\mathbb{E}(X) = kp_k/2.$$

(i) La proportion d'enfants au sein d'une école « au hasard », ce que sous-entend le tirage uniforme d'un enfant dans la population, qui proviennent d'une famille de k enfants est p_k^* et donc, pour $k \in \{1, \ldots, n\}$, la proportion d'enfants répondant k est p_k^*. On a donc, pour $k \in \{1, 2, \ldots, 10\}$,

$$\mathbb{P}(Y = k) = p_k^* = kp_k/2.$$

Il est piquant de remarquer que $\mathbb{P}(Y = 0) = 0$ car on ne peut rencontrer aucun enfant provenant d'une famille n'ayant pas eu d'enfants.

(ii) En appliquant la formule de l'espérance pour Y, puis en remarquant que l'on a affaire à la formule de transfert pour $\mathbb{E}(X^2)$, on a donc

$$\mathbb{E}(Y) = \sum_{k=0}^{10} k.p_k^* = \frac{1}{\mathbb{E}(X)} \sum_{k=0}^{10} k^2.p_k = \mathbb{E}\left(X^2\right)/\mathbb{E}(X).$$

(iii) On a donc (question A.1.a.iii) $\mathbb{E}(Y) = \frac{28}{5}.\frac{1}{2} = \frac{14}{5} = 2,8 \geq \mathbb{E}(X)$.

On a obtenu cette comparaison d'un point de vue numérique, ce qui suffit dans cette partie, on peut remarquer qu'un argument plus théorique est peut-être plus satisfaisant (on remarque $\mathbb{V}(X) \geq 0$) :

$$\mathbb{E}(Y) - \mathbb{E}(X) = \frac{\mathbb{E}(X^2) - \mathbb{E}(X)^2}{\mathbb{E}(X)} = \frac{\mathbb{V}(X)}{\mathbb{E}(X)} \geq 0$$

A.2. Soit n un entier naturel et soit X une variable aléatoire à valeurs dans $\{0, \ldots, n\}$, non identiquement nulle. Pour tout entier $i > 0$, on pose $q_i = \frac{i}{\mathbb{E}(X)} \mathbb{P}(X = i)$.

A.2.a. On a (en remarquant que $0.\mathbb{P}(X = 0) = 0$, puis que l'on a affaire à $\mathbb{E}(X)$) :

$$\sum_{i=1}^{n} q_i = \frac{1}{\mathbb{E}(X)} \sum_{i=0}^{n} i.\mathbb{P}(X = i) = \frac{\mathbb{E}(X)}{\mathbb{E}(X)} = 1.$$

Le fait que X soit positive, non identiquement nulle garantit que $\mathbb{E}(X) > 0$ (et donc que sa présence au dénominateur est légitime).

La suite finie $(q_i)_{1\leq i\leq N}$ de nombres positifs définit donc bien une loi de probabilité. On peut donc considérer une variable aléatoire X^* vérifiant, pour tout i entier naturel non nul

$$P(X^* = i) = \frac{i}{\mathbb{E}(X)}\mathbb{P}(X = i)$$

A.2.b. La variable aléatoire X^* est (presque-sûrement) à valeurs dans $\{1,\dots,n\}$ car $\sum_{k=1}^n \mathbb{P}(X^* = k) = 1$.

On a donc, par la formule de l'espérance pour X^*, puis par la formule de transfert pour X,

$$\mathbb{E}(X^*) = \sum_{i=1}^{n} i.\mathbb{P}(X^* = i) = \frac{1}{\mathbb{E}(X)}\sum_{i=1}^{n} i^2.\mathbb{P}(X = i) = \frac{\mathbb{E}(X^2)}{\mathbb{E}(X)}.$$

A.2.c. Et donc, calcul déjà exhibé en A.1.c.iii, par KOENIG–HUYGHENS,

$$\mathbb{V}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \mathbb{E}(X)\left[\mathbb{E}(X^*) - \mathbb{E}(X)\right].$$

A.2.d. Comme $\mathbb{V}(X) \geq 0$, $\mathbb{E}(X) > 0$, on en conclut que $\mathbb{E}(X^*) - \mathbb{E}(X) \geq 0$, c'est à dire que $\mathbb{E}(X^*) \geq \mathbb{E}(X)$.

Partie B

Biais par la taille, propriétés

.

B.1. *Biais par la taille : le cas de variables à densité.*

B.1.a. La fonction f est nulle sur $[0,+\infty[$, positive (au sens large) ailleurs. On en déduit que $\mathbb{P}(X \geq 0) = 1$ et, comme X est à densité, $\mathbb{P}(X = 0) = 1$ et dnc $X > 0$ presque-sûrement.

On en déduit que la fonction g définie par $\forall x,\ g(x) = \frac{x.f(x)}{\mathbb{E}(X)}$ est partout positive, nulle sur $[0,+\infty[$.

Par ailleurs son intégrale généralisée est

$$\int_0^{+\infty} g(x)\,dx = \frac{1}{\mathbb{E}(X)}\int_0^{+\infty} x.f(x)\,dx = \frac{1}{\mathbb{E}(X)}\mathbb{E}(X) = 1$$

(Cette intégrale généralisée est tout aussi légitime à considérer que celle donnant $\mathbb{E}(X)$)

Soit une variable aléatoire X^* dont la densité est g. On dit que X^* suit la loi de X biaisé par la taille.

B.1.b. Soit a un réel strictement positif. Posons $Z = a.X$.

(i) Soit $h : \mathbb{R} \to \mathbb{R}$ une fonction telle que $\mathbb{E}(h(Z))$ existe. Alors $\mathbb{E}(h(a.X))$ et donc, on a l'égalité, l'intégrale à droite étant absolument convergente,

$$\mathbb{E}(h(Z))) = \int_{-\infty}^{+\infty} h(a.x)f(x)\,dx$$

En appliquant le changement de variable linéaire $z = a.x$ dans cette intégrale (noter que, comme $a > 0$, ce changement de variable est bijectif, strictement croissant de $]-\infty,+\infty[$ sur $]-\infty,+\infty[$. On a aussi « $dx = \frac{1}{a}\,dz$ »), on obtient :

$$\mathbb{E}(h(Z))) = \int_{-\infty}^{+\infty} h(z)\frac{1}{a}f\left(\frac{z}{a}\right)\,dx$$

Cela démontre que la variable aléatoire $Z = aX$ possède pour densité $f_a : z \mapsto \frac{1}{a}f\left(\frac{z}{a}\right)$.

(ii) La variable aléatoire Z^* (dont la densité possède les propriétés requises pour entrer dans la possibilité de définition) a donc pour densité g_a définie par

$$\forall z \in \mathbb{R},\ g_a(z) = \frac{1}{\mathbb{E}(Z)}z.f_a(z)$$

Comme $\mathbb{E}(Z) = a.\mathbb{E}(X)$, on a donc

$$\forall z \in \mathbb{R}, g_a(z) = \frac{1}{\mathbb{E}(X)} \frac{1}{a} . \frac{z}{a} . f\left(\frac{z}{a}\right) = \frac{1}{a} g\left(\frac{z}{a}\right)$$

Comme g_a est une densité de $a \times X^*$ (même calcul de densité par changement de variable linéaire), on en déduit que $(aX)^*$ et $a \times X^*$ possèdent la même loi.
B.1.c. Soit $h : [0, +\infty[\to \mathbb{R}$ une fonction bornée et continue sauf éventuellement en un nombre fini de points. On a, par la formule de transfert pour X que $\mathbb{E}(Xh(X))$ est bien définie si

$$\int_{-\infty}^{+\infty} xh(x).f(x)\, dx \text{ est absolument convergente.}$$

Comme h est bornée et donc majorée, en valeur absolue, par une certaine constante $C > 0$, l'intégrande de cette intégrale (qui est légitime du fait des propriétés de continuité h et de f) est majorée en valeur absolue par :

$$\forall x \in \mathbb{R}, |xh(x).f(x)| \leq C.|x|.f(x)$$

Comme X admet une espérance, l'intégrale $\int_{-\infty}^{+\infty} x.f(x)\, dx$ est ACV et donc par le TCIIP et l'inégalité d'intégrande exhibée précédemment, l'intégrale $\int_{-\infty}^{+\infty} xh(x).f(x)\, dx$ est absolument convergente.

On a donc, comme $\forall x, x.f(x) = \mathbb{E}(X).g(x)$,

$$\mathbb{E}(X.h(X)) = \int_{-\infty}^{+\infty} xh(x).f(x)\, dx = \mathbb{E}(X). \int_{-\infty}^{+\infty} h(x).g(x)\, dx = \mathbb{E}(X).\mathbb{E}(h(X^*)).$$

On peut réécrire cette identité en

$$\mathbb{E}(h(X^*)) = \frac{1}{\mathbb{E}(X)} \mathbb{E}(Xh(X))$$

⋆⋆⋆

B.2. Soient $f : \mathbb{R} \to \mathbb{R}$ et $g : \mathbb{R} \to \mathbb{R}$ deux fonctions croissantes.

Soit X une variable aléatoire telle que les espérances $\mathbb{E}(f(X))$, $\mathbb{E}(g(X))$ et $\mathbb{E}(f(X).g(X))$ sont bien définies.
B.2.a. Soient des nombres réels x_1 et x_2, comme f, resp. g, est croissante, $(f(x_1) - f(x_2))$, resp. $(g(x_1) - g(x_2))$ est du même signe que $x_1 - x_2$ et comme lorsque deux nombres sont de même signe, on en déduit que leur produit est positif, on obtient alors que :

$$(f(x_1) - f(x_2))(g(x_1) - g(x_2)) \geq 0.$$

B.2.b. Soient X_1, X_2 deux variables aléatoires indépendantes, de même loi que X. On a, en développant, (égalité de variables aléatoires) :

$$(f(X_1) - f(X_2))(g(X_1) - g(X_2)) = f(X_1)g(X_1) - f(X_2).g(X_1) - f(X_1).g(X_2) + f(X_2)g(X_2)$$

On remarque ensuite que, du fait que X_1, X_2 ont même loi que X alors

$$\mathbb{E}(f(X_1)g(X_1)) = \mathbb{E}(f(X_2)g(X_2)) = \mathbb{E}(f(X)g(X))$$

De plus, par indépendance de X_1 et X_2, par lemme des coalitions, $f(X_1)$ et $g(X_2)$ sont indépendantes et donc

$$\mathbb{E}(f(X_1)g(X_2)) = \mathbb{E}(f(X_1))\mathbb{E}(g(X_2)) = \mathbb{E}(f(X))\mathbb{E}(g(X))$$

et de même

$$\mathbb{E}(f(X_2)g(X_1)) = \mathbb{E}(f(X_2))\mathbb{E}(g(X_1)) = \mathbb{E}(f(X))\mathbb{E}(g(X))$$

En prenant l'espérance de la formule développée, en utilisant la linéarité de l'espérance et les identités ci-dessus, il reste

$$\begin{aligned}\mathbb{E}\left(\left(f\left(X_1\right)-f\left(X_2\right)\right)\left(g\left(X_1\right)-g\left(X_2\right)\right)\right) &= \mathbb{E}\left(f\left(X_1\right)g\left(X_1\right)\right)-\mathbb{E}\left(f\left(X_2\right).g\left(X_1\right)\right)\\ &\quad -\mathbb{E}\left(f\left(X_1\right).g\left(X_2\right)\right)+\mathbb{E}\left(f\left(X_2\right)g\left(X_2\right)\right)\\ &= \mathbb{E}\left(f\left(X\right)g\left(X\right)\right)-\mathbb{E}\left(f\left(X\right)\right).\mathbb{E}\left(g\left(X\right)\right)\\ &\quad -\mathbb{E}\left(f\left(X\right)\right).\mathbb{E}\left(g\left(X\right)\right)+\mathbb{E}\left(f\left(X\right)g\left(X\right)\right)\\ &= 2\mathbb{E}\left(f\left(X\right)g\left(X\right)\right)-2\mathbb{E}\left(f\left(X\right)\right).\mathbb{E}\left(g\left(X\right)\right)\end{aligned}$$

B.2.c. Comme l'espérance d'une variable aléatoire positive est positive, l'inégalité de B.2.a et le calcul ci-dessus montrent que

$$2\mathbb{E}\left(f\left(X\right)g\left(X\right)\right)-2\mathbb{E}\left(f\left(X\right)\right).\mathbb{E}\left(g\left(X\right)\right)\geq 0$$

Ceci est évidemment équivalent à

$$\mathbb{E}\left(f(X).g(X)\right)\geq\mathbb{E}\left(f(X)\right).\mathbb{E}\left(g(X)\right).$$

B.3. Dans cette question, on suppose que X est une variable aléatoire positive d'espérance strictement positive, et telle que $\mathbb{E}\left(X^{m+1}\right)$ existe pour un entier $m\geq 1$ donné.

B.3.a. Soit p un entier naturel tel que $1\leq p\leq m$.

(i) Soit $x\geq 0$,

— Si $x\leq 1$, $0\leq x^p\leq 1\leq 1+x^{m+1}$,

— si $x\geq 1$, $0\leq x^p\leq x^{m+1}\leq 1+x^{m+1}$

et donc dans tous les cas, $0\leq x^p\leq 1+x^{m+1}$.

(ii) Comme $X\geq 0$, l'inégalité précédente montre que $0\leq X^p\leq 1+X^{m+1}$. Comme on suppose que X^{m+1} admet une espérance alors on en déduit que $1+X^{m+1}$ admet une espérance et la majoration indiquée (variante du TCIIP pour les v.a positives) montre que X^p admet une espérance.

On a par ailleurs (ce qui n'est pas demandé) :

$$0\leq\mathbb{E}\left(X^p\right)\leq 1+\mathbb{E}\left(X^{m+1}\right).$$

B.3.b. Comme les fonctions $f:x\mapsto x$ et $g:x\mapsto x^m$ sont croissantes sur $\mathbb{R}^+$, la question B.2.c montre que :

$$\mathbb{E}\left(X^{m+1}\right)\geq\mathbb{E}\left(X\right).\mathbb{E}\left(X^m\right).$$

B.3.c. En appliquant la définition du fait que X^* suit la loi de X biaisée par la taille, avec $h:x\mapsto x^m$ (la variable $X.h(X)=X^{m+1}$ admet bien une espérance), on obtient que

$$\mathbb{E}(h(X^*))=\frac{1}{\mathbb{E}(X)}\mathbb{E}(X.h(X))\geq\frac{1}{\mathbb{E}(X)}\mathbb{E}(X)\mathbb{E}(X^m)$$

ce qui se réécrit en :

$$\mathbb{E}\left(\left(X^*\right)^m\right)\geq\mathbb{E}\left(X^m\right).$$

B.4. Pour tout t réel, on définit la fonction g_t par $g_t(x)=\mathbb{1}_{]t,+\infty[}(x)=\begin{cases}1 & \text{si } x>t\\ 0 & \text{si } x\leq t\end{cases}$.

B.4.a. Soient $x\leq y$ deux nombres réels. On a la distinction de cas :

— Soit $x\leq y\leq t$ auquel cas $g_t(x)=0\leq 0=g_t(y)$;

— Soit $x\leq t<y$ auquel cas $g_t(x)=0<1=g_t(y)$;

— Soit $t<x\leq y$ auquel cas $g_t(x)=1\leq 1=g_t(y)$;

Dans tous les cas $g_t(x) \leq g_t(y)$. On en conclut que $x \mapsto g_t(x)$ est croissante sur $\mathbb{R}$.

On peut être encore plus convaincant en traçant le graphe de g_t. (laissé à la lectrice)

B.4.b. Soit X une variable aléatoire positive, admettant une espérance. Soit $t \in \mathbb{R}$. Comme $g_t \leq 1$, on a $0 \leq X.g_t(X) \leq X$. Comme on suppose que X admet une espérance alors (variante du TCIIP pour l'espérance de v.a. positives), $X.g_t(X)$ admet une espérance et[1] :

$$\mathbb{E}(X.g_t(X)) \leq \mathbb{E}(X).$$

Maintenant on applique B.2.c aux fonctions croissantes $f : x \mapsto x$ et $g = g_t$ pour obtenir que

$$\mathbb{E}\left(Xg_t(X)\right) \geq \mathbb{E}\left(X\right).\mathbb{E}\left(g_t(X)\right)$$

On remarque que $\mathbb{P}(A) = \mathbb{E}(1\!\!1_A)$ pour tout événement A et donc $\mathbb{E}\left(g_t(X)\right) = \mathbb{P}\left(X > t\right)$, ce qui entraîne que :

$$\mathbb{E}\left(Xg_t(X)\right) \geq \mathbb{E}\left(X\right)\mathbb{P}\left(X > t\right).$$

B.4.c. Soit $t \in \mathbb{R}$. En appliquant la définition du fait que X^* suit la loi de X biaisée par la taille, avec $h : x \mapsto g_t(x)$ (la variable $X.h(X) = X.g_t(X)$ admet bien une espérance), puis le résultat tout juste obtenu on obtient que

$$\mathbb{E}(g_t(X^*)) = \frac{1}{\mathbb{E}(X)}\mathbb{E}(X.g_t(X)) \geq \frac{1}{\mathbb{E}(X)}\mathbb{E}\left(X\right)\mathbb{P}\left(X > t\right) = \mathbb{P}\left(X > t\right),$$

ce qui se réécrit (toujours l'identité $\mathbb{P}(A) = \mathbb{E}(1\!\!1_A)$) en :

$$\mathbb{P}\left(X^* > t\right) \geq \mathbb{P}\left(X > t\right).$$

B.5. Soit $X_1, \ldots, X_n$ des variables aléatoires positives, indépendantes, non nécessairement de même loi. On suppose qu'elles admettent toutes une espérance strictement positive, et on note $\mu_i = \mathbb{E}\left(X_i\right)$. De plus, on pose $\mu = \sum_{i=1}^{n} \mu_i$, et $S_n = \sum_{i=1}^{n} X_i$.

B.5.a. Par linéarité de l'espérance,

$$\mathbb{E}\left(S_n\right) = \mathbb{E}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathbb{E}\left(X_i\right) = \sum_{i=1}^{n} \mu_i = \mu.$$

B.5.b. Si les variables aléatoires X_i sont de même loi, elles ont toutes même espérance $\mu_1 = \mathbb{E}(X_1)$ et $n.\mu_1 = \sum_{k=0}^{n} \mathbb{E}(X_k) = \mu$. On en déduit que pour tout k, $\mu_k = \mathbb{E}(X_k) = \mu_1 = \frac{\mu}{n}$.

Soit J une variable aléatoire à valeurs dans $\{1, \ldots, n\}$, de loi définie par $\forall k \in \{1, \ldots, n\}, \mathbb{P}\left(J = k\right) = \mu_k/\mu = \frac{1}{n}$. La v.a. J est donc uniforme sur l'ensemble fini $\{1, \ldots, n\}$.

⋆⋆⋆

On considère $X_1^*, \ldots, X_n^*$ des variables aléatoires indépendantes, indépendantes de $X_1, \ldots, X_n$, telles que, pour tout entier i tel que $1 \leq i \leq n, X_i^*$ suive la loi de X_i biaisée par la taille.

On suppose que J à valeurs dans $\{1, \ldots, n\}$, de loi définie par $\forall k \in \{1, \ldots, n\}, \mathbb{P}\left(J = k\right) = \mu_k/\mu$ est indépendante de $X_1, X_1^*, \ldots, X_n, X_n^*$.

On considère la variable aléatoire $X_J = \sum_{j=1}^{n} X_j 1\!\!1_{\{J=j\}}$ et on définit $T_n = S_n - X_J + X_J^*$

Autrement dit, on choisit un indice aléatoire J et, dans la somme $\sum_{i=1}^{n} X_i$, on remplace X_J par X_J^*

B.5.c. Soit $h : [0, +\infty[\to \mathbb{R}$ bornée et continue sauf éventuellement en un nombre fini de points.

(i) Par esssence, J étant à valeurs dans $\{1, \ldots, n\}$, $(\{J = i\}, i \in \{1, \ldots, n\})$ est un système complet d'événements deux à deux incompatibles (scei), c'est à dire, en langage ensembliste, une partition de Ω.

Soit $\omega \in \Omega$, il existe alors $j \in \{1, \ldots, n\}$ tel que $\omega \in \{J = j\}$ (il suffit de prendre $j = J(\omega)$) et on a alors :

$$1\!\!1_{\{J=j\}}(\omega) = 1 \text{ et } \forall i \in \{1, \ldots, n\} \setminus \{j\},\ 1\!\!1_{\{J=i\}}(\omega) = 0,$$

1. non demandé dans l'énoncé

et donc

$$\begin{aligned}
\left(\sum_{i=1}^{n} h(T_n)\, 1\!\!1_{\{J=i\}}\right)(\omega) &= \sum_{i=1}^{n} h(T_n(\omega))\, 1\!\!1_{\{J=i\}}(\omega) = h(T_n(\omega)) \\
&= h(S_n(\omega) - X_{J(\omega)}(\omega) + X^*_{J(\omega)}(\omega)) = h(S_n(\omega) - X_j(\omega) + X^*_j(\omega)) \\
&= \sum_{i=1}^{n} h(S_n(\omega) - X_i(\omega) + X^*_i(\omega))\, 1\!\!1_{\{J=i\}}(\omega)
\end{aligned}$$

Ces identités sont réalisées pour tout $\omega \in \Omega$ et donc, au sens des variables aléatoires

$$\left(\sum_{i=1}^{n} h(T_n)\, 1\!\!1_{\{J=i\}}\right) = h(T_n) = \sum_{i=1}^{n} h(S_n - X_i + X^*_i)\, 1\!\!1_{\{J=i\}}$$

(ii) En prenant l'espérance de cette identité, par linéarité, on obtient :

$$\mathbb{E}(h(T_n)) = \sum_{i=1}^{n} \mathbb{E}\left(1\!\!1_{\{\{J=i\}\}}.h(S_n - X_i + X^*_i)\right).$$

Comme, pour chaque i, $1\!\!1_{\{\{J=i\}\}}$ et $h(S_n - X_i + X^*_i)$ sont indépendantes par le lemme des coalitions car J est indépendante de $X_1, X^*_1, \ldots, X_n, X^*_n$, on obtient alors :

$$\mathbb{E}(h(T_n)) = \sum_{i=1}^{n} \mathbb{E}\left(1\!\!1_{\{\{J=i\}\}}\right).\mathbb{E}(h(S_n - X_i + X^*_i)) = \sum_{i=1}^{n} \mathbb{P}(J=i)\,\mathbb{E}(h(S_n - X_i + X^*_i)).$$

Pour la toute dernière identité, nous avons appliqué la formule $\mathbb{P}(A) = \mathbb{E}(1\!\!1_A)$.

B.5.d. Soit $i \in \{1, \ldots, n\}$, et s un réel, on applique la définition du fait que X^*_i suit la « loi de X_i biaisée par la taille » à la fonction $\tilde{h}$ définie par

$$\forall x \in \mathbb{R},\ \tilde{h}(x) = h(s+x)$$

pour obtenir

$$\mathbb{E}(h(s + X^*_i)) = \frac{1}{\mathbb{E}(X_i)}\mathbb{E}(X_i.h(s+X_i)) = \frac{1}{\mu_i}\mathbb{E}(X_i.h(s+X_i)).$$

On *admettra* qu'on en déduit l'égalité $\mathbb{E}(h(S_n - X_i + X^*_i)) = \frac{1}{\mu_i}\mathbb{E}(X_i h(S_n))$

B.5.e. En sommant cette dernière identité avec les pondérations $\mathbb{P}(J=i)$, on obtient, en utilisant B.5.c.ii, on obtient que

$$\begin{aligned}
\mathbb{E}(h(T_n)) &= \sum_{i=1}^{n} \mathbb{P}(J=i)\,\mathbb{E}(h(S_n - X_i + X^*_i)) \\
&= \sum_{i=1}^{n} \mathbb{P}(J=i)\,\frac{1}{\mu_i}\mathbb{E}(X_i h(S_n)) \\
&= \sum_{i=1}^{n} \mathbb{E}(X_i h(S_n)) = \mathbb{E}\left(\sum_{i=1}^{n} X_i h(S_n)\right) = \mathbb{E}(S_n.h(S_n)).
\end{aligned}$$

B.5.f. Comme h est quelconque (bornée, continue sur $\mathbb{R}$, sauf en un nombre fini de points de discontinuité), il s'agit de la définition du fait que T_n suit la loi de S_n biaisée par la taille.