

Devoir surveillé — Compléments de probabilités : couples et n-uplets
de variables aléatoires réelles
ECG maths approfondies — 240 min

Durée : 240 min — Barème : 20 points. Calculatrice selon les consignes de l'énoncé. La rédaction et la clarté comptent.

Exercice 1 (3 points) — Le temps de réponse d'un serveur

Une entreprise héberge son application sur un serveur. On appelle *temps de réponse* la durée, exprimée en secondes, qui sépare l'arrivée d'une requête de l'envoi de la réponse correspondante. Ce temps de réponse est modélisé par une variable aléatoire X dont une densité est la fonction f définie sur $\mathbb{R}$ par

$$f(x) = \begin{cases} \frac{c}{(1+x)^3} & \text{si } x \geq 0, \\ 0 & \text{si } x < 0, \end{cases}$$

où c désigne un réel.

On donne $\sqrt{2} \approx 1,4142$.

- (0,5 point) Déterminer la valeur de c pour laquelle f est une densité de probabilité. On justifiera la convergence de l'intégrale utilisée.
- (0,5 point) Déterminer la fonction de répartition F_X de X en tout réel x .
- (0,25 point) Calculer la probabilité que le temps de réponse dépasse une seconde.
- (0,25 point) Calculer $P(1 < X \leq 3)$.
- (0,5 point) Déterminer la médiane de X , c'est-à-dire l'unique réel m tel que $F_X(m) = \frac{1}{2}$. On donnera la valeur exacte, puis une valeur approchée à 10^{-4} près.
- (0,5 point) Montrer que X admet une espérance et que $E(X) = 1$.
- (0,5 point) Montrer que X n'admet pas de moment d'ordre 2. En déduire que X n'admet pas de variance, puis commenter ce résultat en une ou deux phrases dans le contexte du serveur.

Corrigé

1. (0,5 point) Une fonction f est une densité de probabilité lorsqu'elle est positive sur $\mathbb{R}$, continue sur $\mathbb{R}$ sauf éventuellement en un nombre fini de points, et lorsque l'intégrale $\int_{-\infty}^{+\infty} f(x) dx$ converge et vaut 1.

Positivité. Pour $x \geq 0$, on a $(1+x)^3 > 0$, donc $f(x)$ est du signe de c : la positivité impose $c > 0$ (la valeur $c = 0$ est à écarter, l'intégrale serait nulle).

Continuité. La fonction f est continue sur $]-\infty, 0[$ (elle y est nulle) et sur $]0, +\infty[$ (quotient de fonctions continues dont le dénominateur ne s'annule pas). Le seul point de discontinuité éventuel est 0, ce qui est autorisé.

Convergence et calcul de l'intégrale. Comme f est nulle sur $]-\infty, 0[$, il reste à étudier $\int_0^{+\infty} \frac{c}{(1+x)^3} dx$. Pour tout réel $A > 0$,

$$\int_0^A \frac{c}{(1+x)^3} dx = c \left[\frac{-1}{2(1+x)^2} \right]_0^A = c \left(\frac{1}{2} - \frac{1}{2(1+A)^2} \right).$$

Comme $\frac{1}{2(1+A)^2} \xrightarrow{A \rightarrow +\infty} 0$, l'intégrale **converge** et

$$\int_0^{+\infty} \frac{c}{(1+x)^3} dx = \frac{c}{2}.$$

Conclusion : la condition $\frac{c}{2} = 1$ donne $c = 2$, valeur bien strictement positive.

2. (0,5 point) Par définition, $F_X(x) = P(X \leq x) = \int_{-\infty}^x f(t) dt$.

Si $x < 0$: la fonction f est nulle sur $]-\infty, x]$, donc $F_X(x) = 0$.

Si $x \geq 0$: la relation de Chasles et la nullité de f sur les négatifs donnent

$$F_X(x) = \int_0^x \frac{2}{(1+t)^3} dt = \left[\frac{-1}{(1+t)^2} \right]_0^x = 1 - \frac{1}{(1+x)^2}.$$

Conclusion :

$$F_X(x) = \begin{cases} 0 & \text{si } x < 0, \\ 1 - \frac{1}{(1+x)^2} & \text{si } x \geq 0. \end{cases}$$

Contrôle. Cette fonction est bien continue sur $\mathbb{R}$ (elle vaut 0 en 0 par les deux expressions), croissante, de limite 0 en $-\infty$ et de limite 1 en $+\infty$. ✓

3. (0,25 point) L'événement demandé est $(X > 1)$, contraire de $(X \leq 1)$:

$$P(X > 1) = 1 - F_X(1) = 1 - \left(1 - \frac{1}{(1+1)^2} \right) = \frac{1}{4} = 0,25.$$

Conclusion : une requête sur quatre demande plus d'une seconde de traitement.

4. (0,25 point) La fonction de répartition donne directement la probabilité d'un intervalle :

$$P(1 < X \leq 3) = F_X(3) - F_X(1) = \left(1 - \frac{1}{16} \right) - \left(1 - \frac{1}{4} \right) = \frac{1}{4} - \frac{1}{16} = \frac{3}{16} = 0,1875.$$

5. (0,5 point) Existence et unicité. Sur $[0, +\infty[$, la fonction F_X est continue et strictement croissante (sa dérivée f y est strictement positive), avec $F_X(0) = 0$ et $\lim_{x \rightarrow +\infty} F_X(x) = 1$. Elle réalise donc une bijection de $[0, +\infty[$ sur $[0, 1[$, et l'équation $F_X(m) = \frac{1}{2}$ possède une unique solution.

Résolution.

$$\begin{aligned} 1 - \frac{1}{(1+m)^2} = \frac{1}{2} &\iff \frac{1}{(1+m)^2} = \frac{1}{2} \\ &\iff (1+m)^2 = 2 \\ &\iff 1+m = \sqrt{2}, \end{aligned}$$

la dernière équivalence venant de $1+m > 0$. D'où

$$m = \sqrt{2} - 1 \approx 1,4142 - 1 = 0,4142.$$

Conclusion : la médiane vaut $m = \sqrt{2} - 1 \approx 0,4142$ seconde. Une requête sur deux est traitée en moins de 0,41 seconde.

6. (0,5 point) Existence. Par définition, X admet une espérance lorsque l'intégrale $\int_{-\infty}^{+\infty} x f(x) dx$ converge absolument, c'est-à-dire ici lorsque $\int_0^{+\infty} \frac{2x}{(1+x)^3} dx$ converge (l'intégrande est positif sur $[0, +\infty[$, la convergence absolue équivaut donc à la convergence).

La fonction $x \mapsto \frac{2x}{(1+x)^3}$ est continue et positive sur $[0, +\infty[$, et

$$\frac{2x}{(1+x)^3} \underset{x \rightarrow +\infty}{\sim} \frac{2x}{x^3} = \frac{2}{x^2}.$$

Or $\int_1^{+\infty} \frac{dx}{x^2}$ est une intégrale de Riemann d'exposant $2 > 1$: elle converge. Par critère d'équivalence pour des fonctions positives, l'intégrale étudiée converge. **L'espérance $E(X)$ existe.**

Calcul. Décomposons l'intégrande en écrivant $2x = 2(1+x) - 2$:

$$\frac{2x}{(1+x)^3} = \frac{2}{(1+x)^2} - \frac{2}{(1+x)^3}.$$

Pour tout réel $A > 0$,

$$\begin{aligned}\int_0^A \frac{2x}{(1+x)^3} dx &= \left[\frac{-2}{1+x} + \frac{1}{(1+x)^2} \right]_0^A \\ &= \left(\frac{-2}{1+A} + \frac{1}{(1+A)^2} \right) - (-2+1) \\ &= 1 - \frac{2}{1+A} + \frac{1}{(1+A)^2}.\end{aligned}$$

En faisant tendre A vers $+\infty$, les deux derniers termes tendent vers 0, donc

$$E(X) = 1. \quad \square$$

Conclusion : le temps de réponse moyen du serveur est de 1 seconde.

Contrôle de cohérence. La médiane (0,41 s) est nettement inférieure à la moyenne (1 s) : la loi est étalée vers la droite, ce que confirme la décroissance lente de la densité. ✓

7. (0,5 point) Le moment d'ordre 2 n'existe pas. Par le théorème de transfert, $E(X^2)$ existe

si et seulement si l'intégrale $\int_0^{+\infty} \frac{2x^2}{(1+x)^3} dx$ converge. L'intégrande est continu et positif sur $[0, +\infty[$, et

$$\frac{2x^2}{(1+x)^3} \underset{x \rightarrow +\infty}{\sim} \frac{2x^2}{x^3} = \frac{2}{x}.$$

Or $\int_1^{+\infty} \frac{dx}{x}$ est une intégrale de Riemann d'exposant 1 : elle **diverge**. Par critère d'équivalence pour des fonctions positives, l'intégrale étudiée diverge.

Confirmation par le calcul explicite. En écrivant $2x^2 = 2(1+x)^2 - 4(1+x) + 2$, on obtient

$$\frac{2x^2}{(1+x)^3} = \frac{2}{1+x} - \frac{4}{(1+x)^2} + \frac{2}{(1+x)^3},$$

puis, pour $A > 0$,

$$\int_0^A \frac{2x^2}{(1+x)^3} dx = \left[2\ln(1+x) + \frac{4}{1+x} - \frac{1}{(1+x)^2} \right]_0^A = 2\ln(1+A) + \frac{4}{1+A} - \frac{1}{(1+A)^2} - 3,$$

quantité qui tend vers $+\infty$ à cause du terme logarithmique.

Conclusion : $E(X^2)$ **n'existe pas**. Comme la variance se définit par $V(X) = E(X^2) - (E(X))^2$, et qu'elle n'a de sens que si le moment d'ordre 2 existe, la variable X **n'admet pas de variance**. $\square$

Commentaire. Le serveur possède un temps de réponse moyen parfaitement défini (1 seconde), mais aucune dispersion mesurable par un écart-type : les requêtes très lentes sont assez fréquentes pour faire diverger la moyenne des carrés. En pratique, un tel service ne se pilote pas avec un couple moyenne et écart-type, ni avec l'inégalité de Bienaymé-Tchebychev qui exige une variance, mais avec des quantiles (« 99% des requêtes en moins de t secondes »), qui existent toujours.

Erreur classique. L'existence de $E(X)$ n'entraîne jamais celle de $E(X^2)$. Chaque moment se justifie séparément, et l'implication ne vaut que dans l'autre sens.

Exercice 2 (4 points) — Le montant d'un sinistre

Une compagnie d'assurance range les sinistres qu'elle indemnise en trois catégories : *légers*, *moyens* et *graves*. On tire au hasard un dossier de sinistre dans le portefeuille et l'on note :

- A l'événement « le sinistre est léger », de probabilité $P(A) = 0,6$;
- B l'événement « le sinistre est moyen », de probabilité $P(B) = 0,3$;
- C l'événement « le sinistre est grave », de probabilité $P(C) = 0,1$.

On note M le montant de l'indemnisation du sinistre tiré, exprimé en milliers d'euros. Conditionnellement à la catégorie du sinistre, la loi de M est donnée par :

- sachant A , la variable M suit la loi uniforme $\mathcal{U}([0, 2])$;
- sachant B , la variable M suit la loi exponentielle $\mathcal{E}(\frac{1}{5})$;
- sachant C , la variable M suit la loi exponentielle $\mathcal{E}(\frac{1}{40})$.

On **admet** la formule de l'espérance totale pour une variable à densité : si M admet une espérance et si (A, B, C) est un système complet d'événements de probabilités non nulles, alors

$$E(M) = P(A) E(M | A) + P(B) E(M | B) + P(C) E(M | C),$$

où $E(M | A)$ désigne l'espérance de M pour la loi conditionnelle sachant A , et de même pour B et C .

On donne $e^{-0,25} \approx 0,778801$ et $e^{-2} \approx 0,135335$.

1. (0,25 point) Justifier que (A, B, C) est un système complet d'événements de probabilités non nulles, puis écrire, pour chacune des trois catégories, la fonction de répartition conditionnelle du montant.
2. (0,75 point) Déterminer la fonction de répartition F_M de M en tout réel x , en distinguant les cas $x < 0$, $0 \leq x \leq 2$ et $x > 2$.
3. (0,5 point) Justifier que M est une variable à densité, puis déterminer une densité f_M de M .
4. (0,5 point) Calculer $E(M)$ à l'aide de la formule de l'espérance totale rappelée ci-dessus. Interpréter le résultat.
5. (0,75 point) Retrouver la valeur de $E(M)$ par intégration directe de $x f_M(x)$, en justifiant la convergence des intégrales utilisées.
6. (0,5 point) Calculer la probabilité qu'un sinistre pris au hasard dépasse 10 milliers d'euros. On donnera la valeur exacte, puis une valeur approchée à 10^{-4} près.
7. (0,5 point) Un dossier est tiré au hasard et l'on constate que son montant dépasse 10 milliers d'euros. Quelle est la probabilité qu'il s'agisse d'un sinistre grave ? Valeur approchée à 10^{-4} près.
8. (0,25 point) Quelle part du coût moyen d'un sinistre les sinistres graves représentent-ils ? Commenter en deux ou trois phrases le poids des sinistres rares dans le coût du portefeuille.

Corrigé

1. (0,25 point) Un sinistre appartient à une catégorie et à une seule : les événements A , B et C sont deux à deux incompatibles et leur réunion est Ω . Leurs probabilités valent 0,6, 0,3 et 0,1, toutes non nulles, et leur somme vaut bien

$$0,6 + 0,3 + 0,1 = 1.$$

La famille (A, B, C) est donc un **système complet d'événements** de probabilités non nulles, et tout conditionnement par A , B ou C est licite.

Fonctions de répartition conditionnelles. D'après le cours :

Sachant A (loi uniforme sur $[0, 2]$), pour tout réel x ,

$$P_A(M \leq x) = \begin{cases} 0 & \text{si } x < 0, \\ \frac{x}{2} & \text{si } 0 \leq x \leq 2, \\ 1 & \text{si } x > 2. \end{cases}$$

Sachant B (loi exponentielle de paramètre $\frac{1}{5}$), pour tout réel x ,

$$P_B(M \leq x) = 0 \quad \text{si } x < 0, \quad P_B(M \leq x) = 1 - e^{-x/5} \quad \text{si } x \geq 0.$$

Sachant C (loi exponentielle de paramètre $\frac{1}{40}$), pour tout réel x ,

$$P_C(M \leq x) = 0 \quad \text{si } x < 0, \quad P_C(M \leq x) = 1 - e^{-x/40} \quad \text{si } x \geq 0.$$

2. (0,75 point) La formule des probabilités totales, appliquée au système complet (A, B, C) et à l'événement $(M \leq x)$, donne pour tout réel x :

$$F_M(x) = 0,6 P_A(M \leq x) + 0,3 P_B(M \leq x) + 0,1 P_C(M \leq x).$$

Cas $x < 0$. Les trois probabilités conditionnelles sont nulles, donc $F_M(x) = 0$.

Cas $0 \leq x \leq 2$.

$$\begin{aligned} F_M(x) &= 0,6 \times \frac{x}{2} + 0,3 (1 - e^{-x/5}) + 0,1 (1 - e^{-x/40}) \\ &= 0,3x + 0,4 - 0,3e^{-x/5} - 0,1e^{-x/40}. \end{aligned}$$

Cas $x > 2$. La première probabilité conditionnelle vaut 1 :

$$\begin{aligned} F_M(x) &= 0,6 + 0,3 (1 - e^{-x/5}) + 0,1 (1 - e^{-x/40}) \\ &= 1 - 0,3e^{-x/5} - 0,1e^{-x/40}. \end{aligned}$$

Contrôle en $x = 0$. La deuxième expression donne $0 + 0,4 - 0,3 - 0,1 = 0$, ce qui prolonge bien la première. ✓

Contrôle en $x = 2$. La deuxième expression donne $0,6 + 0,4 - 0,3e^{-0,4} - 0,1e^{-0,05}$, soit exactement la valeur de la troisième expression en 2. ✓

Contrôle en $+\infty$. Les deux exponentielles tendent vers 0, donc $F_M(x) \rightarrow 1$. ✓

3. (0,5 point) M est une variable à densité. Les trois expressions obtenues sont de classe $\mathcal{C}^1$ sur les intervalles ouverts $]-\infty, 0[$, $]0, 2[$ et $]2, +\infty[$, et les deux contrôles de la question précédente montrent que F_M est continue en 0 et en 2, donc continue sur $\mathbb{R}$. La fonction F_M est ainsi continue sur $\mathbb{R}$ et de classe $\mathcal{C}^1$ sur $\mathbb{R}$ sauf éventuellement aux deux points 0 et 2 : c'est exactement la définition d'une variable à densité.

Densité. Une densité s'obtient en dérivant F_M là où c'est possible, et en attribuant une valeur arbitraire (ici 0) aux deux points exceptionnels. En dérivant :

- sur $]0, 2[$: $(0,3x)' = 0,3$, $(-0,3e^{-x/5})' = 0,06e^{-x/5}$, $(-0,1e^{-x/40})' = 0,0025e^{-x/40}$;
- sur $]2, +\infty[$: les deux mêmes termes exponentiels, sans le terme constant.

D'où

$$f_M(x) = \begin{cases} 0 & \text{si } x \leq 0, \\ 0,3 + 0,06e^{-x/5} + 0,0025e^{-x/40} & \text{si } 0 < x < 2, \\ 0,06e^{-x/5} + 0,0025e^{-x/40} & \text{si } x \geq 2. \end{cases}$$

Lecture. La densité présente un saut de 0,3 en $x = 2$: c'est la disparition brutale de la contribution des sinistres légers, dont le montant ne dépasse jamais 2 milliers d'euros. Une densité n'a aucune raison d'être continue, seule la fonction de répartition doit l'être.

4. (0,5 point) Les espérances conditionnelles se lisent dans le cours :

$$E(M | A) = \frac{0+2}{2} = 1, \quad E(M | B) = \frac{1}{\frac{1}{5}} = 5, \quad E(M | C) = \frac{1}{\frac{1}{40}} = 40.$$

La formule de l'espérance totale admise dans l'énoncé donne alors

$$E(M) = 0,6 \times 1 + 0,3 \times 5 + 0,1 \times 40 = 0,6 + 1,5 + 4 = 6,1.$$

Conclusion : $E(M) = 6,1$ milliers d'euros, soit 6 100 euros par sinistre en moyenne.

Interprétation. Ce montant moyen est plus de trois fois supérieur au plafond des sinistres légers, qui représentent pourtant 60% des dossiers : la moyenne est tirée vers le haut par une petite minorité de dossiers très coûteux.

5. (0,75 point) **Mise en place.** Par définition, $E(M) = \int_{-\infty}^{+\infty} x f_M(x) dx$, sous réserve de convergence absolue. La fonction $x \mapsto x f_M(x)$ est nulle sur $]-\infty, 0]$ et positive sur $]0, +\infty[$: la convergence absolue équivaut à la convergence. En notant $g(x) = 0,06 e^{-x/5} + 0,0025 e^{-x/40}$, la relation de Chasles donne

$$\int_0^{+\infty} x f_M(x) dx = \int_0^2 0,3 x dx + \int_0^{+\infty} x g(x) dx.$$

Convergence. Pour tout $\lambda > 0$, l'intégrale $\int_0^{+\infty} x \lambda e^{-\lambda x} dx$ converge et vaut $\frac{1}{\lambda}$: c'est le calcul de l'espérance de la loi $\mathcal{E}(\lambda)$, fait en cours. Par conséquent

$$\int_0^{+\infty} x e^{-\lambda x} dx = \frac{1}{\lambda} \int_0^{+\infty} x \lambda e^{-\lambda x} dx = \frac{1}{\lambda^2},$$

et les deux intégrales impropres intervenant dans $\int_0^{+\infty} x g(x) dx$ convergent (avec $\lambda = \frac{1}{5}$, puis $\lambda = \frac{1}{40}$). L'espérance existe donc bien.

Calcul des trois morceaux.

$$\int_0^2 0,3 x dx = 0,3 \left[\frac{x^2}{2} \right]_0^2 = 0,3 \times 2 = 0,6 ;$$

$$0,06 \int_0^{+\infty} x e^{-x/5} dx = 0,06 \times 5^2 = 0,06 \times 25 = 1,5 ;$$

$$0,0025 \int_0^{+\infty} x e^{-x/40} dx = 0,0025 \times 40^2 = 0,0025 \times 1600 = 4.$$

Total.

$$E(M) = 0,6 + 1,5 + 4 = 6,1. \quad \square$$

Conclusion : on retrouve exactement la valeur de la question 4., ce qui valide à la fois la densité obtenue à la question 3. et l'emploi de la formule de l'espérance totale. ✓

Remarque. Les trois termes de cette intégration directe sont, dans l'ordre, $0,6 \times 1$, $0,3 \times 5$ et $0,1 \times 40$: le calcul reproduit terme à terme la formule de l'espérance totale. Ce n'est pas un hasard, c'est la démonstration de cette formule sur cet exemple.

6. (0,5 point) Comme $10 > 2$, on utilise la troisième expression de F_M :

$$P(M > 10) = 1 - F_M(10) = 0,3 e^{-10/5} + 0,1 e^{-10/40} = 0,3 e^{-2} + 0,1 e^{-0,25}.$$

Numériquement,

$$0,3 \times 0,135335 = 0,040601 \quad \text{et} \quad 0,1 \times 0,778801 = 0,077880,$$

d'où

$$P(M > 10) \approx 0,040601 + 0,077880 = 0,118481, \quad \text{soit} \quad P(M > 10) \approx 0,1185.$$

Conclusion : environ 11,85% des sinistres coûtent plus de 10 000 euros.

7. (0,5 point) On cherche $P_{(M>10)}(C)$, alors que les données fournissent le conditionnement inverse : c'est la situation de la **formule de Bayes**. Le conditionnement est licite car $P(M > 10) \approx 0,1185 \neq 0$.

Numérateur. Sachant C , le montant suit la loi $\mathcal{E}(\frac{1}{40})$, donc

$$P(C \cap (M > 10)) = P(C) P_C(M > 10) = 0,1 \times e^{-0,25} \approx 0,077880.$$

Quotient.

$$P_{(M>10)}(C) = \frac{P(C) P_C(M > 10)}{P(M > 10)} \approx \frac{0,077880}{0,118481} \approx 0,6573.$$

Conclusion : sachant qu'un sinistre dépasse 10 milliers d'euros, il est grave avec une probabilité d'environ 0,6573, soit près de deux chances sur trois.

Contrôle. Le même calcul pour la catégorie « moyen » donne $\frac{0,040601}{0,118481} \approx 0,3427$, et un sinistre léger ne peut pas dépasser 2 milliers d'euros, donc $P_{(M>10)}(A) = 0$. La somme vaut $0,6573 + 0,3427 + 0 = 1$. ✓

Lecture assurantielle. Les sinistres graves ne représentent que 10% des dossiers, mais dès qu'un montant dépasse 10 000 euros, la probabilité qu'il s'agisse d'un sinistre grave passe de 0,1 à 0,66 : l'observation du montant est très informative sur la catégorie.

8. (0,25 point) Dans la décomposition de la question 4., la contribution des sinistres graves au coût moyen est

$$P(C) E(M | C) = 0,1 \times 40 = 4,$$

sur un coût moyen total de 6,1 milliers d'euros, soit une part de

$$\frac{4}{6,1} \approx 0,6557, \quad \text{c'est-à-dire environ } 65,6 \%.$$

Commentaire. Un sinistre sur dix seulement est grave, mais ces dossiers pèsent près des deux tiers du coût total du portefeuille. Une tarification calée sur le sinistre « typique », dont le montant est petit, conduirait donc à une prime très insuffisante. C'est la queue de la distribution, rare mais lourde, qui détermine l'équilibre financier de l'assureur, et c'est elle qu'il faut provisionner ou réassurer.

Exercice 3 (4 points) — Les bornes de recharge de véhicules électriques

Une station de recharge pour véhicules électriques met plusieurs bornes à la disposition des automobilistes. La durée de recharge d'un véhicule, exprimée en minutes, est modélisée par une variable aléatoire de loi exponentielle de paramètre $\lambda = \frac{1}{45}$. Les durées de recharge des différents véhicules sont supposées mutuellement indépendantes.

On note T la durée de recharge d'un véhicule quelconque, de sorte que $T \hookrightarrow \mathcal{E}\left(\frac{1}{45}\right)$.

On donne les valeurs approchées suivantes :

$$e^{-2/3} \approx 0,513417 ; \quad e^{-4/3} \approx 0,263597 ; \quad e^{-8/3} \approx 0,069483 ; \quad e^{-10/3} \approx 0,035674 ; \\ 0,9305174 \approx 0,749717 ; \quad \sqrt{2} \approx 1,4142.$$

- (0,25 point) Rappeler la fonction de répartition F_T de T , ainsi que $E(T)$ et $V(T)$.
- (0,25 point) Calculer $P(T > 60)$ et $P(30 \leq T \leq 60)$ à 10^{-4} près.
- (0,5 point) Montrer que, pour tous réels $s > 0$ et $t > 0$,

$$P_{(T>s)}(T > s+t) = P(T > t).$$

Interpréter ce résultat pour un automobiliste qui attend une borne libre.

- (0,75 point) À un instant donné, les 4 bornes de la station sont occupées et l'on note T_1, T_2, T_3, T_4 les durées de recharge restantes des quatre véhicules, supposées mutuellement indépendantes et de même loi que T . On pose $D = \min(T_1, T_2, T_3, T_4)$, instant de la première libération d'une borne. Déterminer la loi de D , puis $E(D)$.
- (0,75 point) On pose $G = \max(T_1, T_2, T_3, T_4)$, instant à partir duquel les quatre bornes sont toutes libres. Déterminer la fonction de répartition de G , puis une densité de G . Calculer enfin $P(G > 120)$ à 10^{-4} près.
- (0,75 point) Un automobiliste effectuant un long trajet doit enchaîner deux recharges, de durées R_1 et R_2 indépendantes et de même loi que T . On pose $S = R_1 + R_2$. Déterminer une densité de S à l'aide du produit de convolution.
- (0,5 point) Montrer que λS suit la loi $\gamma(2)$. En déduire $E(S)$ et $V(S)$, puis $\sigma(S)$ à 10^{-2} près.
- (0,25 point) Calculer $P(S > 150)$ à 10^{-4} près et interpréter.

Corrigé

- (0,25 point) Pour la loi exponentielle de paramètre $\lambda = \frac{1}{45}$, le cours donne

$$F_T(x) = \begin{cases} 0 & \text{si } x < 0, \\ 1 - e^{-x/45} & \text{si } x \geq 0, \end{cases}$$

ainsi que

$$E(T) = \frac{1}{\lambda} = 45 \quad \text{et} \quad V(T) = \frac{1}{\lambda^2} = 45^2 = 2025.$$

Une recharge dure donc 45 minutes en moyenne, avec un écart-type $\sigma(T) = 45$ minutes également.

- (0,25 point) Pour tout $x \geq 0$, $P(T > x) = 1 - F_T(x) = e^{-x/45}$. Donc

$$P(T > 60) = e^{-60/45} = e^{-4/3} \approx 0,2636.$$

Pour l'intervalle, on utilise la fonction de répartition (la variable étant à densité, les inégalités

strictes ou larges donnent la même probabilité) :

$$P(30 \leq T \leq 60) = F_T(60) - F_T(30) = e^{-30/45} - e^{-60/45} = e^{-2/3} - e^{-4/3}.$$

Numériquement,

$$P(30 \leq T \leq 60) \approx 0,513417 - 0,263597 = 0,249820 \approx 0,2498.$$

Conclusion : une recharge sur quatre environ dépasse une heure, et une sur quatre environ dure entre 30 minutes et 1 heure.

3. (0,5 point) Le conditionnement est licite. Pour $s > 0$, $P(T > s) = e^{-\lambda s} > 0$, donc la probabilité conditionnelle $P_{(T>s)}$ est bien définie.

Calcul. Comme $s + t > s$, on a l'inclusion $(T > s + t) \subset (T > s)$, donc

$$(T > s + t) \cap (T > s) = (T > s + t).$$

Par définition d'une probabilité conditionnelle,

$$\begin{aligned} P_{(T>s)}(T > s + t) &= \frac{P((T > s + t) \cap (T > s))}{P(T > s)} \\ &= \frac{P(T > s + t)}{P(T > s)} \\ &= \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} \\ &= e^{-\lambda t} = P(T > t). \quad \square \end{aligned}$$

Interprétation. C'est l'absence de mémoire de la loi exponentielle. Un véhicule branché depuis 30 minutes a exactement la même probabilité de rester encore plus d'une heure qu'un véhicule qui vient de se brancher : ici $e^{-4/3} \approx 0,2636$ dans les deux cas. L'automobiliste qui patiente n'a donc, avec ce modèle, aucune raison de penser qu'une borne occupée depuis longtemps va se libérer plus vite qu'une autre.

4. (0,75 point) Support. Les quatre durées étant positives, D l'est aussi : $P(D < 0) = 0$.

Passage par l'événement contraire. Le minimum de quatre nombres dépasse x si et seulement si chacun d'eux dépasse x . Donc, pour tout réel $x \geq 0$,

$$(D > x) = \bigcap_{i=1}^4 (T_i > x).$$

Les variables T_1, T_2, T_3, T_4 étant **mutuellement indépendantes**, il en va de même des événements ci-dessus, d'où

$$P(D > x) = \prod_{i=1}^4 P(T_i > x) = (e^{-x/45})^4 = e^{-4x/45}.$$

Fonction de répartition. Pour $x \geq 0$, $F_D(x) = 1 - e^{-4x/45}$, et $F_D(x) = 0$ pour $x < 0$. On reconnaît la fonction de répartition d'une loi exponentielle :

$$D \hookrightarrow \mathcal{E}\left(\frac{4}{45}\right).$$

Espérance.

$$E(D) = \frac{1}{\frac{4}{45}} = \frac{45}{4} = 11,25.$$

Conclusion : la première borne se libère au bout de 11,25 minutes en moyenne, soit quatre fois plus vite qu'une borne isolée. Le minimum de n variables exponentielles indépendantes de même paramètre λ suit la loi $\mathcal{E}(n\lambda)$: c'est un résultat à connaître.

5. (0,75 point) Fonction de répartition. Le maximum de quatre nombres est inférieur ou égal à x si et seulement si chacun d'eux l'est. Donc, pour tout réel x ,

$$(G \leq x) = \bigcap_{i=1}^4 (T_i \leq x),$$

puis, par indépendance mutuelle,

$$F_G(x) = \prod_{i=1}^4 F_{T_i}(x) = (F_T(x))^4.$$

D'où

$$F_G(x) = \begin{cases} 0 & \text{si } x < 0, \\ (1 - e^{-x/45})^4 & \text{si } x \geq 0. \end{cases}$$

Densité. La fonction F_G est continue sur $\mathbb{R}$ (elle vaut 0 en 0 par les deux expressions) et de classe $\mathcal{C}^1$ sur $\mathbb{R}$ sauf éventuellement en 0 : G est bien une variable à densité. En dérivant sur $]0, +\infty[$, avec la dérivée d'une puissance de fonction,

$$f_G(x) = 4(1 - e^{-x/45})^3 \times \frac{1}{45} e^{-x/45} = \frac{4}{45} e^{-x/45} (1 - e^{-x/45})^3,$$

et $f_G(x) = 0$ pour $x \leq 0$.

Calcul de $P(G > 120)$. On a $\frac{120}{45} = \frac{8}{3}$, donc

$$P(G > 120) = 1 - F_G(120) = 1 - (1 - e^{-8/3})^4.$$

Avec $e^{-8/3} \approx 0,069483$, il vient $1 - 0,069483 = 0,930517$, puis

$$0,930517^2 \approx 0,865862, \quad 0,930517^4 \approx 0,865862^2 \approx 0,749717.$$

D'où

$$P(G > 120) \approx 1 - 0,749717 = 0,250283 \approx 0,2503.$$

Conclusion : dans un quart des cas environ, il faut attendre plus de deux heures pour que les quatre bornes soient toutes libres, alors qu'une recharge isolée ne dépasse deux heures que dans 6,9% des cas. Attendre le dernier des quatre véhicules coûte beaucoup plus cher qu'attendre un véhicule seul.

6. (0,75 point) Les variables R_1 et R_2 sont indépendantes et à densité, de densité commune

$$f(x) = \lambda e^{-\lambda x} \quad \text{si } x > 0, \quad f(x) = 0 \quad \text{sinon,}$$

avec $\lambda = \frac{1}{45}$. Le cours autorise l'emploi du **produit de convolution** pour la densité de la somme de deux variables indépendantes à densité (la légitimité du procédé est admise) :

$$f_S(x) = \int_{-\infty}^{+\infty} f(t) f(x-t) dt.$$

Réduction du domaine d'intégration. Le facteur $f(t)$ est nul si $t \leq 0$, et le facteur $f(x-t)$ est nul si $x-t \leq 0$, c'est-à-dire si $t \geq x$. L'intégrande n'est donc non nul que pour $0 < t < x$. Si $x \leq 0$: il n'existe aucun tel t , donc $f_S(x) = 0$. C'était prévisible, S étant une somme de variables positives.

Si $x > 0$:

$$\begin{aligned} f_S(x) &= \int_0^x \lambda e^{-\lambda t} \times \lambda e^{-\lambda(x-t)} dt \\ &= \lambda^2 \int_0^x e^{-\lambda t - \lambda x + \lambda t} dt \\ &= \lambda^2 e^{-\lambda x} \int_0^x dt \\ &= \lambda^2 x e^{-\lambda x}. \end{aligned}$$

Conclusion : avec $\lambda = \frac{1}{45}$ et $45^2 = 2025$,

$$f_S(x) = \begin{cases} \frac{x}{2025} e^{-x/45} & \text{si } x > 0, \\ 0 & \text{si } x \leq 0. \end{cases}$$

Remarque. La densité de S s'annule en 0, puis croît avant de décroître : contrairement à une durée de recharge isolée, une durée totale très courte est peu probable, puisqu'il faudrait que les deux recharges soient courtes en même temps.

7. (0,5 point) Loi de λS . Posons $U = \lambda S$. Comme $\lambda > 0$, on a, pour tout réel u ,

$$F_U(u) = P(\lambda S \leq u) = P\left(S \leq \frac{u}{\lambda}\right) = F_S\left(\frac{u}{\lambda}\right).$$

Cette fonction est nulle pour $u \leq 0$, et dérivable pour $u > 0$, de dérivée

$$f_U(u) = \frac{1}{\lambda} f_S\left(\frac{u}{\lambda}\right) = \frac{1}{\lambda} \times \lambda^2 \times \frac{u}{\lambda} e^{-u} = u e^{-u}.$$

Or la loi $\gamma(\nu)$ a pour densité $x \mapsto \frac{1}{\Gamma(\nu)} x^{\nu-1} e^{-x}$ sur $]0, +\infty[$. Pour $\nu = 2$, et puisque $\Gamma(2) = 1! = 1$, cette densité vaut $x \mapsto x e^{-x}$: c'est exactement f_U . Donc

$$\lambda S \hookrightarrow \gamma(2). \quad \square$$

Espérance et variance. Deux chemins mènent au résultat.

Par la linéarité et l'indépendance. L'espérance est linéaire, donc

$$E(S) = E(R_1) + E(R_2) = 45 + 45 = 90.$$

Les variables R_1 et R_2 étant **indépendantes**, la variance de la somme est la somme des variances :

$$V(S) = V(R_1) + V(R_2) = 2025 + 2025 = 4050.$$

Par la loi $\gamma(2)$. Le cours donne $E(\gamma(\nu)) = \nu$ et $V(\gamma(\nu)) = \nu$, donc $E(\lambda S) = 2$ et $V(\lambda S) = 2$, c'est-à-dire $E(S) = \frac{2}{\lambda} = 90$ et $V(S) = \frac{2}{\lambda^2} = 4050$. ✓

Écart-type.

$$\sigma(S) = \sqrt{4050} = \sqrt{2 \times 2025} = 45\sqrt{2} \approx 45 \times 1,4142 = 63,639,$$

soit $\sigma(S) \approx 63,64$ minutes.

8. (0,25 point) Calculons $P(S > 150)$ à partir de la densité. Pour $A > 150$, une intégration par parties avec $u(t) = t$ et $v'(t) = \frac{1}{2025} e^{-t/45}$, donc $v(t) = -\frac{1}{45} e^{-t/45}$, donne

$$\int_{150}^A \frac{t}{2025} e^{-t/45} dt = \left[-\frac{t}{45} e^{-t/45} - e^{-t/45} \right]_{150}^A.$$

Par croissances comparées, les deux termes tendent vers 0 quand A tend vers $+\infty$, donc l'intégrale converge et

$$P(S > 150) = \left(\frac{150}{45} + 1\right) e^{-150/45} = \left(\frac{10}{3} + 1\right) e^{-10/3} = \frac{13}{3} e^{-10/3}.$$

Numériquement,

$$P(S > 150) \approx \frac{13}{3} \times 0,035674 \approx 4,333333 \times 0,035674 \approx 0,154587,$$

soit $P(S > 150) \approx 0,1546$.

Conclusion : l'automobiliste passe plus de deux heures et demie en recharge dans environ 15,5% des trajets, pour une durée totale moyenne de 90 minutes. Prévoir le temps de trajet à partir de la seule moyenne serait donc trop optimiste une fois sur six environ.

Méthode à retenir. Pour un minimum ou un maximum, on ne cherche jamais la densité directement : on passe par la fonction de répartition, en écrivant $(\min > x)$ ou $(\max \leq x)$ comme une intersection, puis on utilise l'indépendance. Pour une somme, on utilise le produit de convolution.

Exercice 4 (4 points) — Les volumes de commandes d'un entrepôt

Un entrepôt logistique prépare chaque jour des commandes. Le volume total préparé au cours d'une journée, exprimé en mètres cubes, est modélisé par une variable aléatoire V suivant la loi normale $\mathcal{N}(500, 900)$. L'écart-type vaut donc $\sigma(V) = 30$.

On note Φ la fonction de répartition de la loi normale centrée réduite, et l'on donne :

$$\begin{aligned}\Phi(1) &= 0,8413 ; & \Phi(1,83) &= 0,9664 ; & \Phi(1,96) &= 0,9750 ; \\ \Phi(2) &= 0,9772 ; & \Phi(2,33) &= 0,9901 ; & \sqrt{30} &\approx 5,4772.\end{aligned}$$

On rappelle que, pour tout réel t , $\Phi(-t) = 1 - \Phi(t)$.

- (0,5 point)* Calculer la probabilité que le volume préparé au cours d'une journée dépasse 560 mètres cubes.
- (0,5 point)* Calculer la probabilité que ce volume soit compris entre 470 et 530 mètres cubes.
- (0,75 point)* L'entrepôt souhaite dimensionner sa capacité de stockage journalière de manière à n'être saturé qu'un jour sur cent au plus. Déterminer la plus petite capacité entière convenable.
- (0,75 point)* On observe 30 jours d'activité. Pour $i \in \llbracket 1, 30 \rrbracket$, on note V_i le volume préparé le i -ième jour. Les variables $V_1, \dots, V_{30}$ sont supposées mutuellement indépendantes et de même loi que V . On pose

$$S = \sum_{i=1}^{30} V_i.$$

Déterminer la loi de S , en précisant le résultat de cours utilisé, puis donner $\sigma(S)$ à 10^{-2} près.

- (0,5 point)* Calculer $P(S > 15\,300)$ à 10^{-4} près.
- (0,5 point)* On pose $\bar{V} = \frac{S}{30}$, volume moyen préparé par jour sur la période. Déterminer la loi de $\bar{V}$.
- (0,5 point)* Déterminer un intervalle centré en 500 dans lequel $\bar{V}$ se trouve avec la probabilité 0,95. Comparer sa demi-largeur à celle de l'intervalle analogue pour une seule journée, et commenter.

Corrigé

1. (0,5 point) Centrage et réduction. Comme $V \hookrightarrow \mathcal{N}(500, 900)$ avec $\sigma(V) = \sqrt{900} = 30$, la variable

$$Z = \frac{V - 500}{30}$$

suit la loi normale centrée réduite $\mathcal{N}(0, 1)$.

Traduction de l'événement.

$$V > 560 \iff \frac{V - 500}{30} > \frac{560 - 500}{30} \iff Z > 2.$$

Calcul.

$$P(V > 560) = P(Z > 2) = 1 - \Phi(2) = 1 - 0,9772 = 0,0228.$$

Conclusion : le volume dépasse 560 mètres cubes environ 2,3% des jours, soit à peu près un jour sur 44.

2. (0,5 point) Les bornes sont symétriques autour de 500 : $470 = 500 - 30$ et $530 = 500 + 30$. Après centrage et réduction,

$$470 \leq V \leq 530 \iff -1 \leq Z \leq 1.$$

Donc

$$\begin{aligned} P(470 \leq V \leq 530) &= \Phi(1) - \Phi(-1) \\ &= \Phi(1) - (1 - \Phi(1)) \\ &= 2\Phi(1) - 1 \\ &= 2 \times 0,8413 - 1 = 0,6826. \end{aligned}$$

Conclusion : environ 68,3% des journées, le volume s'écarte de moins d'un écart-type de la moyenne. On retrouve la règle classique de la loi normale.

3. (0,75 point) Mise en inéquation. Notons c la capacité cherchée. Être saturé signifie que le volume dépasse la capacité : la contrainte s'écrit

$$P(V > c) \leq 0,01.$$

Après centrage et réduction,

$$P(V > c) = 1 - \Phi\left(\frac{c - 500}{30}\right),$$

donc la contrainte équivaut à

$$\Phi\left(\frac{c - 500}{30}\right) \geq 0,99.$$

Lecture de la table. La fonction Φ est strictement croissante, et la table fournit $\Phi(2,33) = 0,9901 \geq 0,99$. Il suffit donc, avec la précision de la table, que

$$\frac{c - 500}{30} \geq 2,33 \iff c \geq 500 + 2,33 \times 30 = 500 + 69,9 = 569,9.$$

Conclusion : la plus petite capacité entière convenable est $c = 570$ mètres cubes.

Contrôle. Avec $c = 570$, on a $\frac{570 - 500}{30} \approx 2,33$, donc $P(V > 570) \approx 1 - 0,9901 = 0,0099 \leq 0,01$.
✓

Interprétation. Pour ne saturer qu'un jour sur cent, il faut prévoir 14% de capacité de plus que le volume moyen. Dimensionner sur la moyenne conduirait à saturer un jour sur deux.

4. (0,75 point) Résultat utilisé. Le cours admet la **stabilité de la loi normale par somme de variables indépendantes** : si $X \hookrightarrow \mathcal{N}(m_1, \sigma_1^2)$ et $Y \hookrightarrow \mathcal{N}(m_2, \sigma_2^2)$ sont indépendantes, alors $X + Y \hookrightarrow \mathcal{N}(m_1 + m_2, \sigma_1^2 + \sigma_2^2)$. Une récurrence immédiate étend ce résultat à une somme de n variables mutuellement indépendantes.

Application. Les variables $V_1, \dots, V_{30}$ sont mutuellement indépendantes et suivent toutes la loi $\mathcal{N}(500, 900)$, donc S suit une loi normale. Ses paramètres se calculent par linéarité de l'espérance et, grâce à l'indépendance, par additivité de la variance :

$$E(S) = \sum_{i=1}^{30} E(V_i) = 30 \times 500 = 15\,000,$$

$$V(S) = \sum_{i=1}^{30} V(V_i) = 30 \times 900 = 27\,000.$$

Conclusion : $S \hookrightarrow \mathcal{N}(15\,000, 27\,000)$.

Écart-type.

$$\sigma(S) = \sqrt{27\,000} = \sqrt{900 \times 30} = 30\sqrt{30} \approx 30 \times 5,4772 = 164,316,$$

soit $\sigma(S) \approx 164,32$ mètres cubes.

Attention. L'hypothèse d'indépendance sert **uniquement** pour la variance et pour la normalité de la somme. L'espérance, elle, s'additionne toujours.

5. (0,5 point) **Centrage et réduction.** La variable $\frac{S - 15\,000}{30\sqrt{30}}$ suit la loi $\mathcal{N}(0, 1)$, et

$$\frac{15\,300 - 15\,000}{30\sqrt{30}} = \frac{300}{30\sqrt{30}} = \frac{10}{\sqrt{30}} = \frac{10\sqrt{30}}{30} = \frac{\sqrt{30}}{3} \approx \frac{5,4772}{3} \approx 1,8257.$$

Lecture de la table. En arrondissant à 1,83, valeur fournie :

$$P(S > 15\,300) = 1 - \Phi\left(\frac{\sqrt{30}}{3}\right) \approx 1 - \Phi(1,83) = 1 - 0,9664 = 0,0336.$$

Conclusion : la probabilité que l'entrepôt traite plus de 15 300 mètres cubes en 30 jours est d'environ 0,0336, soit 3,4%.

Lecture. Le seuil 15 300 ne dépasse la moyenne mensuelle que de 2%, mais cet écart représente près de deux écarts-types de S : sur un cumul de 30 jours, un dépassement même modeste en pourcentage devient improbable.

6. (0,5 point) La variable $\bar{V} = \frac{1}{30}S$ est l'image de S par une transformation affine (de coefficient $\frac{1}{30}$ et de terme constant nul). Or la définition d'une loi normale ($X = \sigma Z + m$ avec $Z \hookrightarrow \mathcal{N}(0, 1)$), jointe au centrage-réduction de S , montre qu'une transformation affine non constante d'une variable normale suit encore une loi normale. Donc $\bar{V}$ suit une loi normale, de paramètres

$$E(\bar{V}) = \frac{1}{30}E(S) = \frac{15\,000}{30} = 500,$$

$$V(\bar{V}) = \left(\frac{1}{30}\right)^2 V(S) = \frac{27\,000}{900} = 30.$$

Conclusion : $\bar{V} \hookrightarrow \mathcal{N}(500, 30)$, avec $\sigma(\bar{V}) = \sqrt{30} \approx 5,4772$ mètres cubes.

Lecture. La moyenne quotidienne sur un mois a la même espérance qu'une journée isolée, mais un écart-type $\sqrt{30}$ fois plus petit : 5,48 contre 30.

7. (0,5 point) **Mise en équation.** On cherche un réel $h > 0$ tel que

$$P(500 - h \leq \bar{V} \leq 500 + h) = 0,95.$$

En posant $Z = \frac{\bar{V} - 500}{\sqrt{30}}$, qui suit la loi $\mathcal{N}(0, 1)$, l'événement se réécrit $|Z| \leq \frac{h}{\sqrt{30}}$, et

$$P\left(|Z| \leq \frac{h}{\sqrt{30}}\right) = 2\Phi\left(\frac{h}{\sqrt{30}}\right) - 1.$$

Résolution.

$$2\Phi\left(\frac{h}{\sqrt{30}}\right) - 1 = 0,95 \iff \Phi\left(\frac{h}{\sqrt{30}}\right) = 0,975.$$

La table donne $\Phi(1,96) = 0,9750$, donc $\frac{h}{\sqrt{30}} = 1,96$, c'est-à-dire

$$h = 1,96\sqrt{30} \approx 1,96 \times 5,4772 \approx 10,7353.$$

Conclusion : l'intervalle cherché est

$$[500 - 10,74; 500 + 10,74] = [489,26; 510,74].$$

Comparaison avec une journée isolée. Le même raisonnement appliqué à V , dont l'écart-type vaut 30, donnerait une demi-largeur de

$$1,96 \times 30 = 58,8,$$

soit l'intervalle $[441,2 ; 558,8]$.

Commentaire. La demi-largeur passe de 58,8 à 10,74 mètres cubes, dans un rapport

$$\frac{58,8}{10,7353} \approx 5,48 \approx \sqrt{30}.$$

Moyenner sur 30 jours divise la dispersion par $\sqrt{30}$, pas par 30 : c'est pourquoi une activité très irrégulière au jour le jour peut être remarquablement stable au mois. En revanche, il faudrait quadrupler la durée d'observation pour diviser encore la dispersion par deux.

Exercice 5 (5 points) — Les relevés d'une station météorologique

Partie A. La hauteur de pluie d'un mois

Une station météorologique enregistre chaque mois la hauteur de pluie tombée, exprimée en millimètres. Cette hauteur est modélisée par une variable aléatoire H dont une densité est la fonction f définie sur $\mathbb{R}$ par

$$f(x) = \begin{cases} \frac{x}{\theta^2} e^{-x^2/(2\theta^2)} & \text{si } x > 0, \\ 0 & \text{si } x \leq 0, \end{cases}$$

où θ désigne un réel strictement positif fixé.

- (0,5 point) Vérifier que f est bien une densité de probabilité.
- (0,5 point) Déterminer la fonction de répartition F_H de H en tout réel x .
- (0,75 point) On pose $Y = H^2$. Montrer que Y suit une loi exponentielle dont on précisera le paramètre.
- (0,5 point) En déduire $E(H^2)$ et $V(H^2)$ en fonction de θ .
- (0,75 point) Soit U une variable aléatoire suivant la loi uniforme sur $]0, 1[$. Déterminer une fonction g telle que $g(U)$ suive la même loi que H , et le démontrer. On attend une réponse mathématique : aucun langage de programmation n'est demandé.

Partie B. Les relevés de plusieurs mois

On relève désormais les hauteurs de pluie de n mois, avec $n \geq 1$. On note $H_1, \dots, H_n$ ces hauteurs, supposées mutuellement indépendantes et de même loi que H . On pose

$$R_n = \max(H_1, \dots, H_n) \quad \text{et} \quad T_n = \sum_{i=1}^n H_i^2.$$

- (0,75 point) Déterminer la fonction de répartition de R_n , puis une densité de R_n .
- (0,5 point) Soit $s > 0$. Exprimer $P(R_n > s)$ en fonction de s , n et θ . Calculer cette probabilité à 10^{-4} près pour $\theta = 40$, $n = 12$ et $s = 100$, puis interpréter.
- (0,75 point) Calculer $E(T_n)$ et $V(T_n)$ en fonction de n et θ , en précisant les hypothèses utilisées. Donner les valeurs numériques pour $\theta = 40$ et $n = 12$, ainsi que $\sigma(T_n)$ arrondi à l'unité.

On donne les valeurs approchées suivantes :

$$e^{-3,125} \approx 0,043937 ; \quad 0,956063^{12} \approx 0,583228 ; \quad \sqrt{12} \approx 3,4641.$$

Corrigé

Partie A

1. (0,5 point) Il faut vérifier trois points : la positivité, la continuité sauf en un nombre fini de points, et la convergence de l'intégrale vers 1.

Positivité. Pour $x > 0$, on a $\frac{x}{\theta^2} > 0$ et $e^{-x^2/(2\theta^2)} > 0$, donc $f(x) > 0$. Pour $x \leq 0$, $f(x) = 0$. La fonction f est donc positive sur $\mathbb{R}$.

Continuité. La fonction f est continue sur $]-\infty, 0[$ et sur $]0, +\infty[$ comme produit et composée de fonctions continues. En 0, $f(0) = 0$ et $\lim_{x \rightarrow 0^+} \frac{x}{\theta^2} e^{-x^2/(2\theta^2)} = 0$: la fonction est même continue sur $\mathbb{R}$ tout entier.

Convergence et valeur de l'intégrale. Posons, pour x réel, $\Psi(x) = -e^{-x^2/(2\theta^2)}$. Cette

fonction est dérivable, et

$$\Psi'(x) = -\left(-\frac{2x}{2\theta^2}\right) e^{-x^2/(2\theta^2)} = \frac{x}{\theta^2} e^{-x^2/(2\theta^2)} = f(x) \quad \text{pour } x > 0.$$

C'est donc une primitive de f sur $]0, +\infty[$. Pour tout réel $A > 0$,

$$\int_0^A f(x) dx = \left[-e^{-x^2/(2\theta^2)}\right]_0^A = 1 - e^{-A^2/(2\theta^2)}.$$

Comme $-\frac{A^2}{2\theta^2} \xrightarrow{A \rightarrow +\infty} -\infty$, l'exponentielle tend vers 0, donc l'intégrale **converge** et

$$\int_{-\infty}^{+\infty} f(x) dx = \int_0^{+\infty} f(x) dx = 1.$$

Conclusion : f est bien une densité de probabilité. $\square$

2. (0,5 point) Pour $x \leq 0$, la fonction f est nulle sur $]-\infty, x]$, donc $F_H(x) = 0$. Pour $x > 0$, le calcul précédent, mené jusqu'à x au lieu de A , donne

$$F_H(x) = \int_0^x f(t) dt = 1 - e^{-x^2/(2\theta^2)}.$$

Conclusion :

$$F_H(x) = \begin{cases} 0 & \text{si } x \leq 0, \\ 1 - e^{-x^2/(2\theta^2)} & \text{si } x > 0. \end{cases}$$

Contrôle. F_H est continue en 0 (les deux expressions y valent 0), croissante, de limite 1 en $+\infty$. $\checkmark$

3. (0,75 point) On cherche la loi de $Y = H^2$ par la **méthode de la fonction de répartition**, seule méthode légitime pour une transformée.

Cas $y \leq 0$. Un carré est positif, donc l'événement $(H^2 \leq y)$ est impossible pour $y < 0$. Pour $y = 0$, il vaut $(H = 0)$, de probabilité nulle puisque H est une variable à densité. Dans les deux cas, $F_Y(y) = 0$.

Cas $y > 0$. L'inégalité $H^2 \leq y$ équivaut à $-\sqrt{y} \leq H \leq \sqrt{y}$, donc

$$F_Y(y) = P(-\sqrt{y} \leq H \leq \sqrt{y}) = F_H(\sqrt{y}) - F_H(-\sqrt{y}).$$

Or $-\sqrt{y} < 0$, donc $F_H(-\sqrt{y}) = 0$, et $\sqrt{y} > 0$, donc

$$F_Y(y) = 1 - e^{-(\sqrt{y})^2/(2\theta^2)} = 1 - e^{-y/(2\theta^2)}.$$

Identification. La fonction obtenue est nulle sur $]-\infty, 0]$ et vaut $1 - e^{-\mu y}$ sur $]0, +\infty[$ avec $\mu = \frac{1}{2\theta^2} > 0$: c'est la fonction de répartition de la loi exponentielle de paramètre μ . Deux variables ayant la même fonction de répartition ont la même loi, donc

$$Y = H^2 \hookrightarrow \mathcal{E}\left(\frac{1}{2\theta^2}\right). \quad \square$$

4. (0,5 point) Le cours donne, pour la loi exponentielle de paramètre μ , l'existence de l'espérance et de la variance, avec $E = \frac{1}{\mu}$ et $V = \frac{1}{\mu^2}$. Ici $\mu = \frac{1}{2\theta^2}$, donc

$$E(H^2) = E(Y) = 2\theta^2 \quad \text{et} \quad V(H^2) = V(Y) = (2\theta^2)^2 = 4\theta^4.$$

Conclusion : $E(H^2) = 2\theta^2$ et $V(H^2) = 4\theta^4$.

Remarque. Le moment d'ordre 2 de H s'obtient donc sans aucun calcul d'intégrale, en identifiant la loi de H^2 . C'est bien plus rapide que le théorème de transfert appliqué à $x \mapsto x^2$, qui demanderait une intégration par parties.

5. (0,75 point) Recherche de g par inversion. Sur $]0, +\infty[$, la fonction F_H est continue, strictement croissante (sa dérivée f y est strictement positive), avec $\lim_{x \rightarrow 0^+} F_H(x) = 0$ et $\lim_{x \rightarrow +\infty} F_H(x) = 1$. Elle réalise donc une bijection de $]0, +\infty[$ sur $]0, 1[$, ce qui autorise à résoudre $F_H(x) = u$ pour $u \in]0, 1[$:

$$\begin{aligned} 1 - e^{-x^2/(2\theta^2)} = u &\iff e^{-x^2/(2\theta^2)} = 1 - u \\ &\iff -\frac{x^2}{2\theta^2} = \ln(1 - u) \\ &\iff x^2 = -2\theta^2 \ln(1 - u) \\ &\iff x = \theta \sqrt{-2 \ln(1 - u)}, \end{aligned}$$

la dernière équivalence étant licite car $x > 0$, et le radicande étant strictement positif puisque $1 - u \in]0, 1[$ entraîne $\ln(1 - u) < 0$.

Candidat. Posons donc

$$g(u) = \theta \sqrt{-2 \ln(1 - u)} \quad \text{pour } u \in]0, 1[, \quad \text{et} \quad Z = g(U).$$

Démonstration. La variable Z est strictement positive, donc $P(Z \leq x) = 0$ pour $x \leq 0$. Soit maintenant $x > 0$. Les deux membres de l'inégalité étant positifs et la fonction carré étant croissante sur $[0, +\infty[$:

$$\begin{aligned} P(Z \leq x) &= P(\theta \sqrt{-2 \ln(1 - U)} \leq x) \\ &= P\left(-2 \ln(1 - U) \leq \frac{x^2}{\theta^2}\right) \\ &= P\left(\ln(1 - U) \geq -\frac{x^2}{2\theta^2}\right) \\ &= P(1 - U \geq e^{-x^2/(2\theta^2)}) \\ &= P(U \leq 1 - e^{-x^2/(2\theta^2)}), \end{aligned}$$

où l'on a utilisé la croissance stricte de l'exponentielle, puis multiplié par -1 en renversant l'inégalité.

Le réel $t = 1 - e^{-x^2/(2\theta^2)}$ appartient à $]0, 1[$, et pour U uniforme sur $]0, 1[$ on a $P(U \leq t) = t$. Donc

$$P(Z \leq x) = 1 - e^{-x^2/(2\theta^2)} = F_H(x).$$

Les variables Z et H ont la même fonction de répartition sur $\mathbb{R}$: elles suivent la même loi. $\square$

Conclusion : la variable $\theta \sqrt{-2 \ln(1 - U)}$ simule une hauteur de pluie mensuelle.

Variante. Si U est uniforme sur $]0, 1[$, alors $1 - U$ l'est aussi. La variable $\theta \sqrt{-2 \ln U}$ convient donc également, et c'est la forme retenue en pratique.

Partie B

6. (0,75 point) Fonction de répartition. Pour tout réel x , le maximum est inférieur ou égal à x si et seulement si chacune des n hauteurs l'est :

$$(R_n \leq x) = \bigcap_{i=1}^n (H_i \leq x).$$

Les variables $H_1, \dots, H_n$ étant **mutuellement indépendantes**, il vient

$$F_{R_n}(x) = \prod_{i=1}^n P(H_i \leq x) = (F_H(x))^n,$$

puisqu'elles ont toutes la même loi que H . D'où

$$F_{R_n}(x) = \begin{cases} 0 & \text{si } x \leq 0, \\ \left(1 - e^{-x^2/(2\theta^2)}\right)^n & \text{si } x > 0. \end{cases}$$

Densité. La fonction F_{R_n} est continue sur $\mathbb{R}$: les deux expressions valent 0 en 0. Elle est de classe $\mathcal{C}^1$ sur $] -\infty, 0[$ et sur $] 0, +\infty[$, donc sur $\mathbb{R}$ sauf éventuellement en 0 : la variable R_n est bien à densité. En dérivant sur $] 0, +\infty[$ la puissance n -ième de F_H :

$$f_{R_n}(x) = n (F_H(x))^{n-1} f(x) = n \left(1 - e^{-x^2/(2\theta^2)}\right)^{n-1} \frac{x}{\theta^2} e^{-x^2/(2\theta^2)},$$

et $f_{R_n}(x) = 0$ pour $x \leq 0$.

7. (0,5 point) Expression générale. Pour $s > 0$, en passant à l'événement contraire,

$$P(R_n > s) = 1 - F_{R_n}(s) = 1 - \left(1 - e^{-s^2/(2\theta^2)}\right)^n.$$

Application numérique. Avec $\theta = 40$, on a $2\theta^2 = 2 \times 1600 = 3200$, puis avec $s = 100$:

$$\frac{s^2}{2\theta^2} = \frac{10\,000}{3200} = 3,125.$$

La valeur fournie donne $e^{-3,125} \approx 0,043937$, donc

$$1 - e^{-3,125} \approx 0,956063,$$

puis, avec $n = 12$ et la seconde valeur fournie,

$$P(R_{12} > 100) \approx 1 - 0,956063^{12} \approx 1 - 0,583228 = 0,416772,$$

soit

$$P(R_{12} > 100) \approx 0,4168.$$

Interprétation. Un mois donné dépasse 100 millimètres de pluie avec la probabilité $e^{-3,125} \approx 0,0439$, soit environ 4,4%. Mais sur douze mois, la probabilité qu'au moins un mois dépasse ce seuil monte à 41,7% : un événement rare à l'échelle du mois devient courant à l'échelle de l'année. C'est le mécanisme, essentiel en hydrologie, du passage de p à $1 - (1 - p)^n$.

8. (0,75 point) Loi de chaque terme. Pour tout $i \in \llbracket 1, n \rrbracket$, la variable H_i suit la même loi que H , donc la question 3. s'applique à l'identique :

$$H_i^2 \hookrightarrow \mathcal{E}\left(\frac{1}{2\theta^2}\right), \quad \text{d'où} \quad E(H_i^2) = 2\theta^2 \quad \text{et} \quad V(H_i^2) = 4\theta^4.$$

Indépendance des carrés. Les variables $H_1, \dots, H_n$ étant mutuellement indépendantes, il en va de même de leurs images par une même fonction : les variables $H_1^2, \dots, H_n^2$ sont mutuellement indépendantes (résultat admis du cours, conséquence du lemme des coalitions).

Espérance. La linéarité de l'espérance ne demande aucune hypothèse d'indépendance :

$$E(T_n) = \sum_{i=1}^n E(H_i^2) = n \times 2\theta^2 = 2n\theta^2.$$

Variance. Ici l'indépendance mutuelle est **indispensable** : elle seule permet d'écrire que la variance d'une somme est la somme des variances.

$$V(T_n) = \sum_{i=1}^n V(H_i^2) = n \times 4\theta^4 = 4n\theta^4.$$

Application numérique. Avec $\theta = 40$ et $n = 12$, on a $2\theta^2 = 3200$ et $4\theta^4 = (2\theta^2)^2 = 3200^2 = 10\,240\,000$. Donc

$$E(T_{12}) = 12 \times 3200 = 38\,400,$$

$$V(T_{12}) = 12 \times 10\,240\,000 = 122\,880\,000,$$

puis

$$\sigma(T_{12}) = \sqrt{122\,880\,000} = 3200\sqrt{12} \approx 3200 \times 3,4641 = 11\,085,12,$$

soit $\sigma(T_{12}) \approx 11\,085$.

Les hauteurs étant en millimètres, la somme des carrés s'exprime en millimètres carrés : $E(T_{12}) = 38\,400 \text{ mm}^2$ et $\sigma(T_{12}) \approx 11\,085 \text{ mm}^2$.

Contrôle de cohérence. Le rapport de l'écart-type à l'espérance vaut

$$\frac{\sigma(T_{12})}{E(T_{12})} = \frac{3200\sqrt{12}}{12 \times 3200} = \frac{1}{\sqrt{12}} \approx 0,2887.$$

Pour un seul mois, ce rapport vaut $\frac{2\theta^2}{2\theta^2} = 1$, puisque l'espérance et l'écart-type d'une loi exponentielle coïncident. Cumuler douze mois divise donc la dispersion relative par $\sqrt{12}$. ✓

Méthode à retenir. Trois réflexes structurent tout cet exercice : la loi d'une transformée se cherche par la fonction de répartition (questions **3.** et **5.**) ; le maximum d'un échantillon se traite en écrivant $(\max \leq x)$ comme une intersection (question **6.**) ; l'espérance d'une somme se calcule toujours, mais la variance d'une somme n'est la somme des variances que sous hypothèse d'indépendance (question **8.**).