MATHEMATIQUES
Variables alétoires à densité

Exercice :

Dans tout l'exercice, les variables aléatoires sont définies sur le même espace probabilisé $(\Omega, \tau, \mathbb{P})$
Si X *est une variable aléatoire, on notera* $\mathbb{E}(X)$ *son espérance et* $\mathbb{V}(X)$ *sa variance.*
Soient a et b deux réels strictement positifs et f la fonction définie sur $\mathbb{R}$ par :

$$f(x) = \begin{cases} 0 & \text{si } x < b \\ a\dfrac{b^a}{x^{a+1}} & \text{si } x \geq b \end{cases}$$

① a) Soit $\alpha \in \mathbb{R}$. Montrer que $\displaystyle\int_b^{+\infty} \frac{dt}{t^\alpha}$ converge si et seulement si $\alpha > 1$ et que :

$$\forall \alpha > 1, \int_b^{+\infty} \frac{dt}{t^\alpha} = \frac{1}{(\alpha - 1)b^{\alpha-1}}$$

b) Montrer que f est une densité de probabilité.
On dit qu'une variable aléatoire suit la loi de Pareto de paramètres a et b si elle admet f pour densité.

② Soit X une variable aléatoire suivant la loi de Pareto de paramètres a et b.
Déterminer la fonction de répartition de X.

③ a) Soit U une variable aléatoire suivant la loi uniforme sur $]0, 1[$.
Montrer que $bU^{-\frac{1}{a}}$ suit la loi de Pareto de paramètres a et b.

b) Écrire une fonction Python `pareto` prenant en argument deux réels strictement positifs a et b et simulant X.

c) Comment pourriez-vous estimer l'espérance de X à l'aide de Python ?

④ a) Montrer que X admet une espérance si, et seulement si, $a > 1$ et que, le cas échéant :

$$\mathbb{E}(X) = \frac{ab}{a-1}$$

b) Montrer que X admet une variance si, et seulement si, $a > 2$ et que, le cas échéant :

$$\mathbb{V}(X) = \frac{ab^2}{(a-1)^2(a-2)}$$

⑤ On suppose désormais que $b = 1$ et pour tout $n \in \mathbb{N}^*$, on pose $W_n = \ln(X_n)$.
Montrer que W_n suit une loi exponentielle dont on précisera le paramètre. En déduire l'espérance et la variance de W_n.

Problème 1 :

Lu dans le rapport de jury : « *De façon générale,* ***la présentation des copies est à améliorer****. Mettre en valeur ses résultats et rendre une copie soignée sont des compétences grandement appréciées par les correcteurs et qu'il n'est pas difficile d'acquérir en s'entraînant. Une bonne utilisation des parenthèses est nécessaire pour marquer la priorité des opérations à effectuer.*
Les questions de cours sont l'occasion pour les candidats de montrer leur sérieux, il ne faut pas les négliger*. Lorsqu'il est explicitement demandé de prouver un résultat, on ne peut pas se contenter de dire qu'il apparaît dans le cours ou de citer son nom.*
Lors de l'utilisation d'un théorème ou d'un résultat démontré dans une question précédente, il est nécessaire de s'assurer que ses hypothèses sont vérifiées. Il est tout à fait possible d'utiliser un résultat d'une question précédente même si l'on n'a pas réussi à la traiter, mais il est souhaitable de préciser de façon explicite à quelle question on fait référence. Évidemment il convient de mettre en majuscules les noms propres.
On ne peut que conseiller aux candidats de bien lire les questions et de prendre le temps de justifier et rédiger les questions traitées plutôt que de se lancer dans un grappillage très rarement fructueux. Le jury note cette année encore un écart entre le niveau des candidats et celui attendu. Il n'est pas nécessaire de faire l'intégralité de l'épreuve (qui est longue pour couvrir une large partie du programme) pour avoir une note excellente. Par contre, ***pour avoir une note correcte, il est nécessaire de connaître son cours, de savoir raisonner, rédiger et calculer****.* »

I. Résultats préliminaires

1. a) Soit $j \in [\![0, p]\!]$. Alors d'une part $\binom{p}{j}\frac{1}{1+j} = \frac{p!}{j!(p-j)!}\frac{1}{1+j} = \frac{p!}{(j+1)!(p-j)!}$.
   Et d'autre part $\binom{p+1}{j+1}\frac{1}{p+1} = \frac{(p+1)!}{(j+1)!(p-j)!} = \frac{p!}{(j+1)!(p-j)!}$.
   Donc $\boxed{\binom{p}{j}\frac{1}{1+j} = \binom{p+1}{j+1}\frac{1}{p+1}.}$

   Lu dans le rapport de jury : « *Question quasiment toujours bien traitée. On ne pouvait bien sûr pas se contenter de citer la formule du pion ou du capitaine [voir l'article Wikipedia « formule du pion » pour deux autres démonstrations de ce résultat...]* »

   b) Soit $j \in [\![0, p-1]\!]$. Alors

$$\begin{aligned}\binom{p}{j+1} + \binom{p}{j} &= \frac{p!}{(j+1)!(p-j-1)!} + \frac{p!}{j!(p-j)!} \\ &= \frac{p!(p-j)}{(j+1)!(p-j)!} + \frac{p!(j+1)}{(j+1)!(p-j)!} \\ &= \frac{p!(p+1)}{(j+1)!(p-j)!} \\ &= \frac{(p+1)!}{(j+1)!(p-j)!} \\ &= \boxed{\binom{p+1}{j+1}.}\end{aligned}$$

   Lu dans le rapport de jury : « *Seuls 45% des candidats ont traité et réussi cette question. On ne peut, là non plus, pas se contenter de citer la formule de Pascal.*
   De très rares candidats proposent une preuve combinatoire. »

2. Une densité de X : $\boxed{x \mapsto \lambda e^{-\lambda x} 1_{\mathbb{R}^+}(x).}$

   Fonction de répartition : $\boxed{x \mapsto (1 - e^{-\lambda x}) 1_{\mathbb{R}^+}(x).}$

   $\boxed{\mathbb{E}(X) = \frac{1}{\lambda} \text{ et } \mathbf{V}(X) = \frac{1}{\lambda^2}.}$

   Lu dans le rapport de jury : « *Il s'agissait d'une question de cours abordée dans quasiment toutes les copies et avec succès dans près de 70% des cas.*
   Les erreurs les plus fréquentes portent sur l'oubli de la fonction indicatrice.
   Attention également aux parenthèses : $1 - e^{-\lambda x} \boldsymbol{1}_{\mathbb{R}_+}(x) \neq (1 - e^{-\lambda x}) \boldsymbol{1}_{\mathbb{R}_+}(x)$. »

3. a) Soit un entier $i \geqslant 2$. La fonction $t \mapsto \dfrac{1}{t}$ est décroissante sur $\mathbb{R}_+^*$ et donc, puisque $i - 1 > 0$: $\forall t \in [i-1, i],\ \dfrac{1}{i} \leqslant \dfrac{1}{t}$.

   Comme $t \mapsto \dfrac{1}{i}$ et $t \mapsto \dfrac{1}{t}$ sont continues sur $[i-1, i]$, on peut intégrer ces fonctions. On obtient par croissance de l'intégrale $\displaystyle\int_{i-1}^{i} \frac{dt}{i} \leq \int_{i-1}^{i} \frac{dt}{t}$, c'est-à-dire

   $$\boxed{\frac{1}{i} \leq \int_{i-1}^{i} \frac{dt}{t}}$$

   De même, on a : $\forall t \in [i, i+1],\ \frac{1}{t} \leqslant \frac{1}{i}$, ce qui donne

   $$\boxed{\int_{i}^{i+1} \frac{dt}{t} \leqslant \frac{1}{i}.}$$

   Lu dans le rapport de jury : « *Question fréquemment traitée, mais moins de 10% des copies ont la totalité des points notamment du fait de l'absence de justifications.*

   — *Il fallait invoquer la décroissance de* $t \longmapsto \dfrac{1}{t}$ *sur* $\mathbb{R}_+^*$ *et utiliser l'hypothèse* $i > 1$ *pour avoir* $i - 1 \in \mathbb{R}_+^*$.
   — *La croissance de l'intégrale devait être mentionnée. Attention, elle conduit à une implication :* $f \leq g$ *sur* $[a, b]$ *n'équivaut pas à* $\displaystyle\int_a^b f \leq \int_a^b g$.
   — *De rares candidats utilisent un schéma, ce qui est une bonne approche, mais ils mentionnent les aires des intégrales sans indiquer à quelle aire correspond* $\dfrac{1}{i}$, *ce qui est dommage et ne permet pas de donner la bonne inégalité. Même avec un schéma, il faut préciser que cela fonctionne grâce à la décroissance de la fonction.*

   Certains candidats essaient d'autres méthodes, comme une étude de fonctions. L'idée est bonne mais il faut se montrer rigoureux, ce qui est trop rarement le cas. Quelques copies tentent d'utiliser les accroissements finis, mais à nouveau une rédaction solide est très rare. »

   b) Soit $n \geqslant 1$. Alors $\ln(n+1) = \ln(n) + \ln\left(1 + \frac{1}{n}\right)$, donc

   $$\frac{\ln(n+1)}{\ln(n)} = 1 + \frac{\ln\left(1 + \frac{1}{n}\right)}{\ln(n)} \xrightarrow[n \to +\infty]{} 1$$

par opérations sur les limites. Donc $\boxed{\ln(n+1) \underset{n\to+\infty}{\sim} \ln(n).}$

Lu dans le rapport de jury : « *Ce résultat classique n'a été correctement traité que dans 25% des cas.*
Le définition correcte de $u \underset{n\to\infty}{\sim} v$ *a été valorisée car sa connaissance est peu maîtrisée :* $u_n \underset{n\to\infty}{\sim} v_n$ *ne signifie pas que* $\lim\limits_{n\to\infty}(u_n - v_n) = 0$ *et* $u_n \underset{n\to\infty}{\sim} v_n$ *n'implique pas forcément* $\ln(u_n) \underset{n\to\infty}{\sim} \ln(v_n)$.
Enfin, obtenir $u_n \underset{n\to\infty}{\sim} 0$ *devrait alerter et écrire* $\lim\limits_{n\to\infty} \ln(n+1) = \ln(n)$ *n'a pas de sens.* »

c) L'encadrement de la question 3.a étant établi pour tout entier $i \geqslant 2$, on le somme pour i de 2 à n, ce qui donne par relation de Chasles

$$\int_2^{n+1} \frac{dt}{t} \leqslant \sum_{i=2}^{n} \frac{1}{i} \leqslant \int_1^n \frac{dt}{t}.$$

En ajoutant 1 et en divisant par $\ln(n)$ (strictement positif dès que $n \geqslant 2$) on obtient donc

$$\frac{\ln(n+1)}{\ln(n)} - \frac{\ln(2)}{\ln(n)} + \frac{1}{\ln(n)} \leqslant \frac{1}{\ln(n)} \sum_{k=1}^{n} \frac{1}{k} \leqslant 1 + \frac{1}{\ln(n)}.$$

D'après la question précédente, $\frac{\ln(n+1)}{\ln(n)} \xrightarrow[n\to+\infty]{} 1$, donc les membres de gauche et de droite de l'encadrement précédent tendent vers 1. Par théorème des gendarmes, il s'ensuit $\boxed{\sum_{k=1}^{n} \frac{1}{k} \underset{n\to+\infty}{\sim} \ln(n).}$

Lu dans le rapport de jury : « *Question classique traitée par 2/3 des copies mais moins de 10% d'entre elles ont eu tous les points.*

— *Il fallait sommer à partir du rang 2 pour éviter une intégrale généralisée divergente en 0.*
— *Obtenir* $\int_0^n \frac{1}{t} dt$ *ou* $\ln(n+1) \leq \cdots \ln(n)$ *devrait interroger.*
— *Le théorème d'encadrement pour les équivalents n'est pas au programme de BCPST. Il fallait donc tout diviser par* $\ln(n)$ *pour* $n > 1$ *et se ramener ainsi à un théorème d'encadrement des limites.*

»

II. Quelques résultats autour de la loi exponentielle

4. Soit un réel x. Alors $[X_{(1)} > x] = \bigcap_{i=1}^{n} [X_i > x]$, donc l'indépendance de $X_1, \dots, X_n$ donne

$$\mathbf{P}(X_{(1)} > x) = \prod_{i=1}^{n} \mathbf{P}(X_i > x).$$

Cependant, toutes les variables X_i suivent la loi $\mathcal{E}(\lambda)$, par conséquent

$$\mathbf{P}(X_{(1)} > x) = \begin{cases} (e^{-\lambda x})^n = e^{-n\lambda x} & \text{si } x \geqslant 0 \\ 1 & \text{si } x < 0 \end{cases}.$$

Ainsi la fonction de répartition de $X_{(1)}$ est donnée par $F_{X_{(1)}}(x) = (1 - e^{-n\lambda x})1_{\mathbb{R}^+}(x)$. On reconnaît la fonction de répartition de la loi exponentielle $\mathcal{E}(n\lambda)$. Or la fonction de répartition caractérise la loi, donc $\boxed{X_{(1)} \hookrightarrow \mathcal{E}(n\lambda).}$

Lu dans le rapport de jury : « *Globalement, la décomposition de l'événement $(X_{(1)} > x)$ est correcte et l'utilisation de l'indépendance des variables aléatoires $X_1, X_2, \cdots, X_n$ est mentionnée. Il ne fallait pas oublier de traiter le cas négatif pour obtenir la loi de $X_{(1)}$.* »

5. Pour tout réel x, on a $[X_{(n)} \leqslant x] = \bigcap_{i=1}^{n} [X_i \leqslant x]$ ce qui donne par indépendance des X_i la fonction de répartition de $X_{(n)}$:

$$F_{X_{(n)}}(x) = \prod_{i=1}^{n} F_{X_i}(x) = F_{X_1}(x)^n = (1 - e^{-\lambda x})^n 1_{\mathbb{R}^+}(x).$$

On constate :
— par composition de fonctions $\mathcal{C}^1$, $F_{X_{(n)}}$ est de classe $\mathcal{C}^1$ sauf peut-être en 0 ;
— $\underset{x\to 0}{lim}(1 - e^{-\lambda x})^n = 0$ donc $F_{X_{(n)}}$ est continue en 0.
Donc $\boxed{X_{(n)} \text{ est une variable à densité.}}$
En outre, $\frac{d}{dx}\left[(1 - e^{-\lambda x})^n\right] = n\lambda e^{-\lambda x}(1 - e^{-\lambda x})^{n-1}$, donc une densité de $X_{(n)}$ est bien donnée par

$$\boxed{x \mapsto n\lambda e^{-\lambda x}(1 - e^{-\lambda x})^{n-1} 1_{\mathbb{R}_+}(x).}$$

Lu dans le rapport de jury : « *Certains prennent le résultat donné et montrent qu'ils ont une densité de probabilité mais ne montrent pas que c'est une densité de $X_{(n)}$.
Le résultat étant donné, il s'agissait de le justifier. Rappelons que X admet une densité si, et seulement si, sa fonction de répartition est continue sur $\mathbb{R}$ et de classe $\mathcal{C}^1$ sur $\mathbb{R}$ sauf éventuellement en un nombre fini de points. Si l'obtention avec justification de la fonction de répartition sur $\mathbb{R}^+$ est souvent correcte, l'obtention sur $\mathbb{R}^-$ est oubliée, ainsi que la régularité.
Attention aux parenthèses : $\prod_{i=1}^{n}(1 - e^{-\lambda x}) \neq \prod_{i=1}^{n} 1 - e^{-\lambda x}$* »

6. Préalablement, observons que pour tout réel $\mu > 0$, l'intégrale $\int_0^{+\infty} xe^{-\mu x}dx$ est convergente et vaut $\dfrac{1}{\mu^2}$. Pour le prouver, il suffit d'observer que $\int_0^{+\infty} \mu xe^{-\mu x}dx$ est l'espérance d'une variable aléatoire de loi $\mathcal{E}(\mu)$, donc est convergente et égale à $\frac{1}{\mu}$.
Ensuite, l'espérance de $X_{(n)}$ est donnée par l'intégrale suivante

$$I = \int_0^{+\infty} n\lambda xe^{-\lambda x}(1 - e^{-\lambda x})^{n-1}dx$$

sous réserve de convergence. Or par binôme de Newton :

$$\begin{aligned} e^{-\lambda x}(1-e^{-\lambda x})^{n-1} &= e^{-\lambda x}\sum_{j=0}^{n-1}\binom{n-1}{j}(-1)^j(e^{-\lambda x})^j \\ &= \sum_{j=0}^{n-1}\binom{n-1}{j}(-1)^j e^{-\lambda(1+j)x} \end{aligned}$$

Par linéarité de l'intégrale, l'intégrale I est donc convergente, l'espérance de $X_{(n)}$ existe et est égale à

$$\begin{aligned} \mathbb{E}(X_{(n)}) &= \sum_{j=0}^{n-1} n\lambda\binom{n-1}{j}(-1)^j\int_0^{+\infty} xe^{-\lambda(1+j)x}dx \\ &= \sum_{j=0}^{n-1} n\lambda\binom{n-1}{j}(-1)^j\frac{1}{(1+j)^2\lambda^2} \\ &= \frac{1}{\lambda}\sum_{j=0}^{n}\binom{n}{j+1}\frac{(-1)^j}{1+j} \end{aligned}$$

en utilisant $\binom{n-1}{j}\frac{1}{j+1} = \binom{n}{j+1}\frac{1}{n}$ (d'après la question 1). On a donc bien

$$\boxed{\mathbb{E}(X_{(n)}) = \tfrac{1}{\lambda}\sum_{j=0}^{n-1}\binom{n}{j+1}\frac{(-1)^j}{j+1}.}$$

Lu dans le rapport de jury : « *L'existence de l'espérance n'est pas assurée, on ne pouvait donc pas la manipuler sans précautions.*
La question était technique : le binôme de Newton a été correctement utilisé mais le calcul de l'intégrale a posé problème :
— *On pouvait reconnaître qu'à une constante multiplicative près, il s'agissait de l'espérance d'une loi exponentielle, mais la constante était souvent fausse et certaines copies sont allées jusqu'à modifier leur question 2 initialement juste pour obtenir le résultat attendu.*
— *On pouvait procéder par intégration par parties généralisée. On souligne alors la nécessité de confirmer la convergence de tous les termes apparaissant dans une telle formule.*
»

7. a) Par binôme de Newton, $\boxed{\sum_{k=0}^{p+1}\binom{p+1}{k}(-1)^k = (1-1)^{p+1} = 0.}$

Lu dans le rapport de jury : « *La question n'a été traitée que par* 60% *des copies, correctement dans* 60% *des cas* »

b) On établit le résultat par récurrence sur p.
— Initialisation : pour $p=1$, le résultat à prouver est

$$\binom{1}{1}\frac{1}{1} = \frac{1}{1}$$

ce qui est vrai.

— Hérédité : supposons le résultat acquis pour un rang $p \in \mathbb{N}^*$ donné. Alors

$$\begin{aligned}
\sum_{j=0}^{p} \binom{p+1}{j+1} \frac{(-1)^j}{j+1} &= \sum_{j=0}^{p-1} \binom{p+1}{j+1} \frac{(-1)^j}{j+1} + \binom{p+1}{p+1} \frac{(-1)^p}{p+1} \\
&= \sum_{j=0}^{p-1} \left[\binom{p}{j+1} + \binom{p}{j} \right] \frac{(-1)^j}{j+1} + \frac{(-1)^p}{p+1} && \text{par la question 2} \\
&= \sum_{j=0}^{p-1} \binom{p}{j+1} \frac{(-1)^j}{j+1} + \sum_{j=0}^{p-1} \binom{p}{j} \frac{(-1)^j}{j+1} + \binom{p}{p} \frac{(-1)^p}{p+1} \\
&= \underbrace{\sum_{j=0}^{p-1} \binom{p}{j+1} \frac{(-1)^j}{j+1}}_{A} + \underbrace{\sum_{j=0}^{p} \binom{p}{j} \frac{(-1)^j}{j+1}}_{B}
\end{aligned}$$

D'une part, $A = \sum_{k=1}^{p} \frac{1}{k}$ par hypothèse de récurrence.
D'autre part, en utilisant la première question :

$$\begin{aligned}
B &= \frac{1}{p+1} \sum_{j=0}^{p} \binom{p+1}{j+1} (-1)^j \\
&= -\frac{1}{p+1} \sum_{k=1}^{p+1} \binom{p+1}{k} (-1)^k \\
&= -\frac{1}{p+1} \left(\sum_{k=0}^{p+1} \binom{p+1}{k} (-1)^k - \binom{p+1}{0} (-1)^0 \right) \\
&= -\frac{1}{p+1} (0-1) && \text{par la question précédente} \\
&= \frac{1}{p+1}
\end{aligned}$$

Finalement, on trouve bien $\sum_{j=0}^{p} \binom{p+1}{j+1} \frac{(-1)^j}{j+1} = \sum_{k=1}^{p} \frac{1}{k} + \frac{1}{p+1} = \sum_{k=1}^{p+1} \frac{1}{k}$, ce qui conclut la preuve de l'hérédité.

— Conclusion : $\boxed{\forall p \in \mathbb{N}^*,\ \sum_{j=0}^{p-1} \binom{p}{j+1} \frac{(-1)^j}{j+1} = \sum_{k=1}^{p} \frac{1}{k}.}$

Lu dans le rapport de jury : « *Question de synthèse traitée dans moins de la moitié des copies. La rédaction correcte de la récurrence a été valorisée de même que les copies amorçant l'hérédité en utilisant la formule de Pascal.* »

8. D'après les question 6 et 7, $\mathbb{E}(X_{(n)}) = \frac{1}{\lambda} \sum_{k=1}^{n} \frac{1}{k}$.

   Mais d'après la question 3.c, $\sum_{k=1}^{n} \frac{1}{k} \underset{n \to +\infty}{\sim} \ln(n)$.

Donc $\boxed{\mathbb{E}(X_{(n)}) \underset{n\to+\infty}{\sim} \frac{\ln(n)}{\lambda}.}$

Lu dans le rapport de jury : « *Question traitée dans 1/3 des copies avec succès dans plus de 70% des cas. Les résultats cohérents avec la question 3.c) ont eu la totalité des points.* »

9. Soit $n \in \mathbb{N}^*$. Alors

$$\begin{aligned} u_{n+1} - u_n &= \sum_{k=1}^{n+1} \frac{1}{k} - \ln(n+1) - \sum_{k=1}^{n} \frac{1}{k} + \ln(n) \\ &= \frac{1}{n+1} - \ln\left(1 + \frac{1}{n}\right) \\ &= \frac{1}{n} \cdot \frac{1}{1+\frac{1}{n}} - \ln\left(1 + \frac{1}{n}\right) \\ &= g\left(\frac{1}{n}\right) \end{aligned}$$

pour $g(x) = \frac{x}{1+x} - \ln(1+x)$. Calculons un développement limité en 0 de g.

$$\begin{aligned} g(x) &= x\,(1 - x + o(x)) - \left(x - \frac{x^2}{2} + o(x^2)\right) \\ &= -\frac{x^2}{2} + o(x^2) \end{aligned}$$

Donc $g(x) \underset{x\to0}{\sim} -\frac{x^2}{2}$. Comme $\frac{1}{n} \xrightarrow[n\to+\infty]{} 0$, on a donc $\boxed{u_{n+1} - u_n \underset{n\to+\infty}{\sim} -\frac{1}{2n^2}.}$

Par conséquent $u_n - u_{n+1}$ est positif à partir d'un certain rang, et le théorème de comparaison pour les séries à termes positifs indique que les séries $\sum(u_n - u_{n+1})$ et $\sum \frac{1}{n^2}$ sont de même natures. Cette dernière série est une série de référence convergente, donc $\sum(u_n - u_{n+1})$ est convergente. Par linéarité, $\boxed{\sum(u_{n+1} - u_n) \text{ est convergente.}}$

Lu dans le rapport de jury : « *Question peu traitée.* »

10. La série $\sum(u_{n+1} - u_n)$ est télescopique, et $\sum_{k=1}^{n-1}(u_{k+1} - u_k) = u_n - u_1$. La convergence de la série $\sum(u_{n+1} - u_n)$ est donc équivalente à la convergence de la suite u. Donc $\boxed{u \text{ est convergente.}}$

III. Loi de Gumbel

11. Premièrement, la fonction exp étant positive et continue, f est également positive et continue. Reste à prouver que $\int_{-\infty}^{+\infty} f$ est convergente et égale à 1.
Vérifions les hypothèses pour effectuer le changement de variable $y = \exp(-x)$.
— $x \mapsto \exp(-x)$ est de classe $\mathcal{C}^1$ et strictement monotone, et on a $dy = -\exp(-x)dx$;
— $\lim_{x\to-\infty} \exp(-x) = +\infty$ et $\lim_{x\to+\infty} \exp(-x) = 0$.
Donc $\int_{-\infty}^{+\infty} f$ est de même nature, et en cas de convergence de même valeur, que

$$\int_{+\infty}^{0} (-\lambda) \exp(-\lambda y) dy = \int_{0}^{+\infty} \lambda \exp(-\lambda y) dy.$$

On reconnaît l'intégrale de la densité de la loi $\mathcal{E}(\lambda)$, qui est donc convergente et égale à 1.
Conclusion : $\boxed{f \text{ est densité de probabilité.}}$

Lu dans le rapport de jury : « *La continuité et la positivité ont souvent été citées. La question a été problématique pour le calcul de l'intégrale impropre.*
— *Les hypothèses d'un changement de variables pour les intégrales impropres ne sont que trop rarement mentionnées. Le fait que le changement de variable soit donné par l'énoncé ne dispense pas de vérifier qu'il est licite surtout sur une intégrale généralisée.*
— *Le changement d'écriture de l'intégrale après changement de variable reste très problématique, ce qui amène à des erreurs pour trouver* 1 *à la fin.*
— *Certains reconnaissent une primitive, ce qui simplifiait beaucoup la question.*
»

12. Calculons la fonction de répartition F_Y de Y. Pour tout réel x, on a

$$\begin{aligned}
F_Y(x) &= \mathbf{P}(Y \leqslant x) \\
&= \mathbf{P}(-\ln X_1 \leqslant x) \\
&= \mathbf{P}(\ln X_1 \geqslant -x) \\
&= \mathbf{P}(X_1 \geqslant e^{-x}) && \text{car } \exp \text{ croît strictement ;} \\
&= e^{-\lambda e^{-x}} && \text{car } X_1 \hookrightarrow \mathcal{E}(\lambda).
\end{aligned}$$

Il apparaît que F_Y est de classe $\mathcal{C}^1$ sur $\mathbb{R}$, donc Y admet une densité égale à F_Y' c'est-à-dire

$$x \mapsto \lambda \exp(-x) \exp[-\lambda \exp(-x)]$$

Donc $\boxed{Y = -\ln X_1 \text{ suit la loi de Gumbel.}}$

Lu dans le rapport de jury : « *La réponse étant donnée, les points sont attribués au raisonnement. Exemple de niveau de rigueur attendu,* ***les points valorisés sont encadrés****.*
On détermine la fonction de répartition de Y. Soit $x \in \mathbb{R}$. On a :

$$\mathbb{P}(Y \leq x) = \mathbb{P}(-\ln(X_1) \leq x) = \mathbb{P}(\ln(X_1) \geq -x)$$

Comme $\boxed{\exp \textit{ est strictement croissante sur } \mathbb{R}}$*, on en déduit que :*

$$\mathbb{P}(Y \leq x) = \mathbb{P}(X_1 \geq e^{-x}) = 1 - \mathbb{P}(X_1 < e^{-x})$$

Or $\boxed{X_1 \textit{ est à densité}}$ *donc :*

$$\mathbb{P}(Y \leq x) = \boxed{1 - F_{X_1}(e^{-x})}$$

Enfin, X_1 suit une loi exponentielle de paramètre λ et $e^{-x} > 0$, donc :

$$\mathbb{P}(Y \leq x) = 1 - (1 - \exp(-\lambda \exp(-x))) = \exp(-\lambda e^{-x})$$

On constate alors que $\boxed{F_Y \textit{ est de classe } \mathcal{C}^1(\mathbb{R})}$ *(grâce aux théorèmes d'opérations). On peut donc conclure que Y est à densité de densité :*

$$F_Y' : x \longmapsto \lambda e^{-x} \exp(-\lambda e^{-x})$$

Autrement dit, Y suit une loi de Gumbel »

13. Soit $x \in \mathbb{R}$.
    Pour tout n assez grand, $1 + \frac{x}{n} > 0$, et

$$\ln\left[\left(1+\frac{x}{n}\right)^n\right] = n\ln\left(1+\frac{x}{n}\right) \underset{n\to+\infty}{\sim} n\frac{x}{n} = x$$

donc $\ln\left[\left(1+\frac{x}{n}\right)^n\right] \xrightarrow[n\to+\infty]{} x$. Par composition de limites $\boxed{\lim_{n\to+\infty}\left(1+\frac{x}{n}\right)^n = \exp(x).}$

Lu dans le rapport de jury : « *La réponse étant donnée, les points sont attribués au raisonnement.* »

14. On a calculé à la question 5 la fonction de répartition de $X_{(n)}$, qui était (pour $\lambda = 1$)

$$x \mapsto (1 - e^{-x})^n 1_{\mathbb{R}^+}(x)$$

Mais pour tout réel x, on a $\mathbf{P}(X_{(n)} - \ln(n) \leqslant x) = \mathbf{P}(X_{(n)} \leqslant x + \ln(n))$. Donc la fonction de répartition de $X_{(n)} - \ln(n)$, que l'on notera H_n, est égale à

$$H_n : x \mapsto (1 - e^{-(x+\ln n)})^n 1_{[-\ln(n),+\infty[}(x) = \left(1 - \frac{e^{-x}}{n}\right)^n 1_{[-\ln(n),+\infty[}(x).$$

Or, pour un réel x donné :
— d'une part $\left(1 - \frac{e^{-x}}{n}\right)^n \xrightarrow[n\to+\infty]{} e^{-e^{-x}}$ par la question précédente,
— d'autre part $1_{[-\ln(n),+\infty[}(x) = 1$ pour tout n assez grand, car $-\ln(n) \xrightarrow[n\to+\infty]{} -\infty$.
Donc $H_n(x) \xrightarrow[n\to+\infty]{} e^{-e^{-x}}$. Mais $x \mapsto e^{-e^{-x}}$ est (dans le cas où $\lambda = 1$) la fonction de répartition de la loi de Gumbel, qu'on avait trouvée à la question 12. C'est donc la fonction de répartition de Y. Par conséquent $\boxed{(X_{(n)} - \ln(n))_{n\in\mathbb{N}^*} \text{ converge en loi vers } Y.}$

Lu dans le rapport de jury : « *Question très peu abordée. A noter que le jury a récompensé ceux qui écrivaient la définition de la convergence en loi.* »

Problème 2 :

Lu dans le rapport de jury : « *Ce problème était un exercice d'analyse plus que de probabilités, beaucoup de questions demandant de calculer des intégrales, parfois sur es intervalles non fermés, bornés. Les définitions du cours de probabilités étaient quand même nécessaires (et ont souvent fait défaut) et ce dès la première question.* »

① ✐ On peut commencer par noter que si $\alpha = 1$, alors $l_{\alpha,r}(t)$ vaut $1/r$ sur l'intervalle $]0, r[$ et 0 sinon. Il s'agit d'une densité de probabilité sur $\mathbb{R}$, celle de la loi uniforme sur $]0, r[$.

Plus généralement, on commence par dire que $l_{\alpha,r}$ est **positive** sur $\mathbb{R}$ si et seulement si $\alpha > 0$.

Ensuite, pour la **continuité**, elle est trivialement continue sur $]-\infty, 0]$ et sur $[r, +\infty[$ puisque nulle sur ces intervalles de $\mathbb{R}$ et continue sur $]0, r[$ puisque $t \longmapsto t^{\alpha-1}$ est continue sur $]0, r[$ pour tout $\alpha \in \mathbb{R}$.
Dès lors, pour tout $\alpha \in \mathbb{R}$, $l_{\alpha,r}$ est **continue par morceaux sur** $\mathbb{R}$ ou encore $l_{\alpha,r}$ **est continue sur** $\mathbb{R} \setminus \{0, r\}$**.**

Si on veut prolonger cette réponse (<u>ce qui n'est pas indispensable</u>), on pourra dire que $l_{\alpha,r}$ n'est jamais continue en r puisque, pour tout $\alpha > 0$, $\lim\limits_{t \to r^-} l_{\alpha,r}(t) = \dfrac{\alpha}{r} \neq 0 = l_{\alpha,r}(r)$ mais que la continuité en 0 dépend des valeurs de α :

— Si $\alpha > 1$ alors $\lim\limits_{t\to 0} t^{\alpha-1} = 0$ et donc $\lim\limits_{t \to 0^+} l_{\alpha,r}(t) = 0 = l_{\alpha,r}(0)$. $\boxed{l_{\alpha,r} \text{ est continue sur } \mathbb{R} \setminus \{r\}}$.
— Si $0 < \alpha < 1$, $\lim\limits_{t\to 0} t^{\alpha-1} = +\infty$ donc $\lim\limits_{t\to 0} l_{\alpha,r} = +\infty \neq l_{\lambda,r}(0)$. $\boxed{l_{\alpha,r} \text{ est continue sur } \mathbb{R} \setminus \{0, r\}}$.
— si $\alpha = 1$, alors $l_{\alpha,r}$ est continue sur $\mathbb{R} \setminus \{0, r\}$ (cf la loi uniforme sur $]0, 1[$...)

Pour montrer que $\displaystyle\int_{-\infty}^{\infty} l_{\alpha,r}(t)dt$ converge et vaut 1, on commence par appliquer la relation de Chasles. Alors :

$$\int_{-\infty}^{+\infty} l_{\alpha,r}(t)dt = \frac{\alpha}{r^\alpha}\int_0^r t^{\alpha-1}dt = \frac{\alpha}{r^\alpha}\int_0^r \frac{1}{t^{1-\alpha}}dt.$$

Cette intégrale est une intégrale généralisée en 0 si $0 < \alpha < 1$ et une intégrale définie si $\alpha \geq 1$. Dans tous les cas, on peut dire que c'est une intégrale de Riemann, qui converge si et seulement si $1 - \alpha < 1 \Leftrightarrow \alpha > 0$.

Ainsi $\displaystyle\int_{-infty}^{+\infty} l_{\alpha,r}(t)dt$ converge si et seulement si $\alpha > 0$.

Prenons donc $\alpha > 0$. On peut dire de façon générale que pour $\varepsilon > 0$:

$$I(\varepsilon) = \int_\varepsilon^r t^{\alpha-1}dt = \left[\frac{t^\alpha}{\alpha}\right]_\varepsilon^r = \frac{r^\alpha - \varepsilon^\alpha}{\alpha}.$$

Puisque $\alpha > 0$, $\lim\limits_{\varepsilon \to 0^+} I(\varepsilon) = \frac{r^\alpha}{\alpha}$.
Ainsi $\displaystyle\int_{-\infty}^{+\infty} l_{\alpha,r}(t)dt$ converge lorsque $\alpha > 0$ et : $\displaystyle\int_{-\infty}^{+\infty} l_{\alpha,r}(t)dt = \frac{\alpha}{r^\alpha} \times \frac{r^\alpha}{\alpha} = 1.$

En conclusion : $\boxed{l_{\alpha,r} \text{ est une densité de probabilité si et seulement si } \alpha > 0.}$

Lu dans le rapport de jury : *« Il y a trois axiomes à vérifier pour démontrer que $l_{\alpha,r}$ est une densité de probabilité : sa continuité par morceaux sur $\mathbb{R}$, le fait qu'elle est positive, et enfin que son intégrale sur $]-\infty,+\infty[$ converge et vaut 1. Les trois points ont été vérifiés de manière totalement aléatoire par les candidats, ce qui montre qu'ils connaissent mal la définition.*
Concernant chacun des points : le plus délicat à vérifier est le calcul de l'intégrale, qui demande de fixer d'abord une borne finie A, puis de la faire tendre vers 0. Beaucoup de candidats n'ont pas vu que pour certaines valeurs de α, la fonction n'était pas continue sur $\mathbb{R}$, avec une limite infinie en 0^+.
Enfin le si et seulement si indiquant une équivalence, a posé de nombreux problèmes de rédaction, les candidats semblant pour bon nombre d'entre eux mal à l'aise avec la logique élémentaire, ou en tous cas ayant du mal à retranscrire convenablement les arguments à l'écrit. »

Maintenant $X \hookrightarrow \mathcal{L}(\alpha, r)$, ce qui signifie que X a pour densité $l_{\alpha,r}$.

② a) Soit $x \in]0, r[$.

$$F(x) = \frac{\alpha}{r^\alpha}\int_0^x t^{\alpha-1}dt = \frac{x^\alpha}{r^\alpha}\text{(on a déjà montré la convergence...)}$$

Et puisque $X(\Omega) =]0, r[$ il est immédiat que $F(x) = 0$ si $x \leq 0$ et $F(x) = 1$ si $x \geq 1$.
En résumé :

$$\boxed{F(x) = \begin{cases} 0 & \text{si } x \leq 0 \\ \dfrac{x^\alpha}{r^\alpha} & \text{si } 0 < x < r \\ 1 & \text{si } r \leq x. \end{cases}}$$

Lu dans le rapport de jury : *« Pour la fonction de répartition, on a à nouveau besoin de calculer une intégrale sur un intervalle non compact. Certains candidats dérivent la densité (qui rappelons-le n'est pas continue lorsque $\alpha \in]0,1[$) pour obtenir le résultat, ce qui évidemment ne donne rien de bon ; mais ce qui est remarquablement incohérent, c'est que dans une question ultérieure, les mêmes vont à nouveau dériver, cette fois-ci la fonction de répartition pour obtenir la densité.*
Dans quelques copies, on trouve :

$$\int_{-\infty}^{t} \alpha\frac{s^{\alpha-1}}{r^\alpha}ds$$

ce qui n'a évidemment aucun sens puisque $l_{\alpha,r}(s)$ n'a l'expression utilisée ci-dessus comme intégrande que lorsque $s \in]0, r[$; et même lorsque le calcul est bien rédigé, la réponse est peu rigoureuse : l'expression de F_X s'obtient en distinguant trois cas, on ne peut pas conclure (et encadrer) $F_X(t) = \left(\frac{t}{r}\right)^\alpha$ sans rien préciser d'autre : cette expression ne saurait quoi qu'il arrive être valable sur $\mathbb{R}$ tout entier, vues les propriétés de la fonction de répartition. »

b) ***Conclusion*** : $\boxed{P(X \leq 0) = F(0) = 0}$

Lu dans le rapport de jury : *« Que vaut $\mathbb{P}(X \leq 0)$? Eh bien elle vaut 0, ce qu'on obtient avec la formule trouvée pour F_X ou bien en remarquant que le support de X est contenu dans $]0,+\infty[$ d'après la fonction densité. »*

c) Posons $W = -\ln\left(\frac{X}{r}\right)$.
Puisque $X(\Omega) =]0, r[$ (*on peut aussi utiliser la question précédente si on veut vraiment suivre la logique de l'énoncé*), W est bien définie et $W(\Omega) =]0, +\infty[$. Soit $t \in]0, +\infty[$.

$$\begin{aligned} F_W(t) &= P(W \leq t) = P\Big(\ln\big(\frac{X}{r}\big) \geq -t\Big) \\ &= P\Big(\frac{X}{r} \geq e^{-t}\Big) = P\big(X \geq re^{-t}\big) = 1 - F(re^{-t}). \end{aligned}$$

Or $t > 0$ donc $re^{-t} \in]0, r[$, donc :

$$\forall t > 0,\ \ F_W(t) = 1 - \big(\frac{re^{-t}}{r}\big)^{\alpha} = 1 - e^{-\alpha t}.$$

On reconnait la loi exponentielle :

$$\boxed{W \hookrightarrow \mathcal{E}(\alpha).}$$

Lu dans le rapport de jury : « *Dans cette question, la conclusion est souvent juste, mais on remarque le même manque de rigueur dans les écritures intermédiaires : la densité de W n'est pas :*

$$f_W(t) = 1 - e^{-\alpha t}$$

mais bien

$$f_W(t) = \begin{cases} 1 - e^{-\alpha t} & si\ t \geq 0 \\ 0 & sinon \end{cases}$$

Toujours concernant la rédaction, on ne peut passer directement de la fonction de répartition à la densité : ayant obtenu F_W, on constate que celle -ci est continûment dérivable par morceaux sur $\mathbb{R}$. On en déduit alors que W est une variable à densité, et que la densité s'obtient en dérivant (chaque morceaux) de F_W »

③ Ici $U_1 \hookrightarrow \mathcal{E}(\lambda)$, $U_2 \hookrightarrow \mathcal{E}(\mu)$, U_1 et U_2 sont indépendantes.

a) Posons $Y = -U_2$. Alors $Y(\Omega) =]-\infty, 0[$. Soit $t < 0$:

$$F_Y(t) = P(-U_2 \leq t) = P(U_2 \geq -t) = 1 - F_{U_2}(-t) = 1 - (1 - e^{\mu t}) = e^{\mu t}.$$

Ainsi :

$$F_Y(t) = \begin{cases} e^{\mu t} & \text{si } t < 0, \\ 1 & \text{si } t \geq 0. \end{cases}$$

La fonction F_Y est de classe C^1 (et donc continue) sur $]-\infty, 0[$ et sur $]0, +\infty[$. De plus F_Y est continue en 0 car $\lim\limits_{t \to 0^-} F_Y(t) = e^0 = 1$, $F_Y(0) = 1$, et $\lim\limits_{t \to 0^+} F_Y(t) = 1$. Ainsi F_Y est continue sur $\mathbb{R}$ et C^1 sur $\mathbb{R}$ privé d'un nombre fini de points. Donc Y est bien une variable à densité.

$$\boxed{\text{Une densité de } Y \text{ est : } f_Y(t) = \mu e^{\mu t} \mathbf{1}_{\mathbb{R}_-^*}(t) = \begin{cases} \mu e^{\mu t} & \text{si } t < 0, \\ 0 & \text{si } t \geq 0. \end{cases}}$$

Lu dans le rapport de jury : « *Il faut ici faire un calcul et ne surtout pas appliquer un faux argument de transformation affine, menant dans quelques copies à la densité négative* *(sic)* $-f_{U_2}$. *Encore une fois, répétons aux candidats de prendre garde au sens de ce qu'ils écrivent, un signe* $-$ *devant une densité devrait faire trembler leur poignet au moment où la pointe de leur stylo s'apprête à le tracer.* »

b) On a : $U_1 - U_2 = U_1 + Y$ avec U_1 et Y indépendantes puisque U_1 et U_2 le sont (**lemme de coalition**). On peut donc appliquer la formule du produit de convolution après avoir noté que $(U_1 + Y)(\Omega) = \mathbb{R}$. D'où : pour tout réel t, en notant $g_{\lambda,\mu}$ la densité de $U_1 - U_2$:

$$g_{\lambda,\mu}(t) = \int_{-\infty}^{+\infty} f_{U_1}(u) f_Y(t-u)du.$$

Soit :

$$\begin{aligned} g_{\lambda,\mu}(t) &= \int_{-\infty}^{\infty} \lambda e^{-\lambda u}\mathbf{1}_{\mathbb{R}_+}(u)\mu e^{\mu(t-u)}\mathbf{1}_{\mathbb{R}_-^*}(t-u)du \\ &= \lambda\mu e^{\mu t}\int_{-\infty}^{\infty} e^{-(\lambda+\mu)u}\mathbf{1}_{\mathbb{R}_+}(u)\mathbf{1}_{\mathbb{R}_-^*}(t-u)du \\ &= \lambda\mu e^{\mu t}\int_{-\infty}^{\infty} e^{-(\lambda+\mu)u}\mathbf{1}_{\mathbb{R}_+\cap]t,+\infty[}(u)du \text{ car } \mathbf{1}_{\mathbb{R}_-^*}(t-u) = 1 \text{ si } t-u<0 \Leftrightarrow u>t \end{aligned}$$

- **Premier cas :** $t > 0$. Alors : $u \geq 0$ et $u > t \Leftrightarrow u > t$. Ainsi :

$$\begin{aligned} g_{\lambda,\mu}(t) &= \int_t^{+\infty} \lambda e^{-\lambda u}\mu e^{\mu(t-u)}du = \lambda\mu e^{\mu t}\int_t^{+\infty} e^{-(\lambda+\mu)u}du \\ &= \lambda\mu e^{\mu t}\lim_{A\to+\infty}\Big[\frac{e^{-(\lambda+\mu)u}}{-(\lambda+\mu)}\Big]_t^A = \lambda\mu e^{\mu t}\lim_{A\to+\infty}\frac{e^{-(\lambda+\mu)t}-e^{-(\lambda+\mu)A}}{\lambda+\mu} \\ &= \lambda\mu e^{\mu t}\frac{e^{-(\lambda+\mu)t}}{\lambda+\mu} = \frac{\lambda\mu}{\lambda+\mu}e^{-\lambda t}. \end{aligned}$$

- **Second cas :** $t \leq 0$. Alors : $u \geq 0$ et $u > t \Leftrightarrow u > 0$. Ainsi :

$$g_{\lambda,\mu}(t) = \int_0^{+\infty} \lambda e^{-\lambda u}\mu e^{\mu(t-u)}du = \lambda\mu e^{\mu t}\lim_{A\to+\infty}\Big[\frac{e^{-(\lambda+\mu)u}}{-(\lambda+\mu)}\Big]_0^A = \lambda\mu e^{\mu t}\frac{1}{\lambda+\mu}.$$

En conclusion :

$$\boxed{g_{\lambda,\mu}(t) = \begin{cases} \dfrac{\lambda\mu}{\lambda+\mu}e^{-\lambda t} & \text{si } t > 0, \\ \dfrac{\lambda\mu}{\lambda+\mu}e^{\mu t} & \text{si } t \leq 0. \end{cases}}$$

Lu dans le rapport de jury : « *Voici la première d'un doublet de questions calculatoires. Ceux qui l'ont abordée sérieusement l'ont en général bien menée, et obtiennent le bon résultat. Mais bien trop souvent la rédaction s'arrête après une ré-écriture de la formule rappelée dans l'énoncé, certains candidats n'allant même pas jusqu'à substituer les expressions de* f_{U_2} *et* f_{-U_2}. *A titre de comparaison, les calculs de ce problème d'analyse sont globalement bien moins maîtrisées que les questions calculatoires d'algèbre linéaire du premier exercice.* »

c) Ici $X \hookrightarrow \mathcal{L}(\alpha, r)$, $Y \hookrightarrow \mathcal{L}(\beta, s)$, X et Y sont indépendantes.
On notera $W_1 = -\ln\big(\dfrac{X}{r}\big)$ et $W_2 = -\ln\big(\dfrac{Y}{s}\big)$.

Alors d'après la question (2.3) : $W_1 \hookrightarrow \mathcal{E}(\alpha)$ et : $W_2 \hookrightarrow \mathcal{E}(\beta)$.
Par ailleurs on pose $Z = -\ln\left(\dfrac{sX}{rY}\right)$.
Pour faire le lien avec les questions qui precèdent on remarque que :

$$Z = -\ln\big(\frac{X}{r}\big) + \ln\big(\frac{Y}{s}\big) = W_1 - W_2.$$

Or W_1 et W_2 sont des variables exponentielles indépendantes, donc on peut appliquer le résultat de (3.2) qui donne la densité de Z :

$$\boxed{g_Z(t) = g_{\alpha,\beta}(t) = \begin{cases} \dfrac{\alpha\beta}{\alpha+\beta} e^{-\alpha t} & \text{si } t > 0, \\ \dfrac{\alpha\beta}{\alpha+\beta} e^{\beta t} & \text{si } t \leq 0. \end{cases}}$$

On a alors (on utilise le fait que toutes les quantités présentes sont strictement positives) :

$$\begin{aligned} P(X<Y) &= P\big(\frac{X}{Y} < 1\big) = P\big(\frac{sX}{rY} < \frac{s}{r}\big) \\ &= P\Big(-\ln\big(\frac{sX}{rY}\big) > -\ln\big(\frac{s}{r}\big)\Big) = \int_{-\ln\left(\frac{s}{r}\right)}^{+\infty} g_Z(t)dt. \end{aligned}$$

Pour continuer il faut connaitre le signe de la borne inférieure de l'intégrale, d'où les deux cas qui suivent :
- Premier cas : $s < r$.
Alors $-\ln\big(\dfrac{s}{r}\big) > 0$ puis :

$$\begin{aligned} P(X<Y) &= \int_{-\ln\left(\frac{s}{r}\right)}^{+\infty} \frac{\alpha\beta}{\alpha+\beta} e^{-\alpha t} dt = \frac{\beta}{\alpha+\beta} \lim_{A\to+\infty} \big[-e^{-\alpha t}\big]_{-\ln\left(\frac{s}{r}\right)}^{A} \\ &= \frac{\beta}{\alpha+\beta} e^{\alpha \ln\left(\frac{s}{r}\right)} = \frac{\beta}{\alpha+\beta}\big(\frac{s}{r}\big)^{\alpha}. \end{aligned}$$

- Second cas : $s \geq r$.
Alors $-\ln\big(\dfrac{s}{r}\big) \leq 0$ puis, avec la relation de Chasles :

$$\begin{aligned} P(X<Y) &= \int_{-\ln\left(\frac{s}{r}\right)}^{0} \frac{\alpha\beta}{\alpha+\beta} e^{\beta t} dt + \int_0^{+\infty} \frac{\alpha\beta}{\alpha+\beta} e^{-\alpha t} dt \\ &= \frac{\alpha}{\alpha+\beta}\big[e^{\beta t}\big]_{-\ln\left(\frac{s}{r}\right)}^{0} + \frac{\beta}{\alpha+\beta} \lim_{A\to+\infty} \big[-e^{-\alpha t}\big]_0^A \\ &= \frac{\alpha}{\alpha+\beta}\big(1 - e^{-\beta\ln\left(\frac{s}{r}\right)}\big) + \frac{\beta}{\alpha+\beta} \\ &= \frac{\alpha}{\alpha+\beta}\big(1 - \big(\frac{r}{s}\big)^{\beta}\big) + \frac{\beta}{\alpha+\beta} \\ &= \frac{\alpha+\beta}{\alpha+\beta} - \frac{\alpha}{\alpha+\beta}\big(\frac{r}{s}\big)^{\beta} \\ &= 1 - \frac{\alpha}{\alpha+\beta}\big(\frac{r}{s}\big)^{\beta}. \end{aligned}$$

Pour conclure on a bien le résultat demandé :

$$\boxed{P(X<Y)=\begin{cases}\dfrac{\beta}{\alpha+\beta}\Big(\dfrac{s}{r}\Big)^{\alpha} & \text{si } s<r\\ 1-\dfrac{\alpha}{\alpha+\beta}\Big(\dfrac{r}{s}\Big)^{\beta} & \text{si } s\geq r\end{cases}}$$

Lu dans le rapport de jury : « *Le point clé de cette question était d'écrire :*

$$-\ln\left(\frac{sX}{rY}\right)=\ln\left(\frac{Y}{s}\right)-\ln\left(\frac{X}{r}\right)$$

ce qui demande de connaître les propriétés algébiques du logarithme. Celles-ci, pourtant vues en classe de Terminale, ne sont toujours pas maîtrisées par un nombre non négligeable de candidats »

④

a) $P_M(R)$ est la probabilité que X soit inférieur à Y en sachant que le sujet est atteint de la maladie H. On sait alors que $\alpha=4, \beta=2, r=400, s=800$. En appliquant le dernier résultat de (3.3) (on est dans le cas $s\geq r$) :

$$P_M(R)=1-\frac{4}{4+2}\Big(\frac{400}{800}\Big)^2=1-\frac{1}{6}=\frac{5}{6}.$$

Pour $P_{\overline{M}}(R)$ on applique encore (3.3) avec $\alpha=2, \beta=3, r=100, s=50$ (on a alors $s<r$) :

$$P_{\overline{M}}(R)=\frac{3}{2+3}\Big(\frac{50}{100}\Big)^2=\frac{3}{20}.$$

En conclusion :

$$\boxed{P_M(R)=\frac{5}{6}.}\qquad\boxed{P_{\overline{M}}(R)=\frac{3}{20}.}$$

Lu dans le rapport de jury : « *La fin du sujet dissimulait deux questions faciles, que tous les candidats n'ont malheureusement pas pensé à aller voir (car lorsque M est réalisé, c'est-à-dire lorsqu le sujet est malade, on a $\alpha=4$, $\beta=2$, $r=400$ et $s=800$, ce qui en particulier indique qu'on est dans le cas $r\leq s$ pour appliquer la formule de la question précédente).*
*Arrêtons-nous sur deux remarques : la première c'est que la calculatrice était autorisée et qu'il est donc absolument impardonnable de faire une erreur de calcul sur ces fractions (et il y en a eu). La seconde, c'est que sur le nombre malgré tout important de candidats ayant abordé la question, beaucoup n'ont pas réussi à substituer les valeurs des quatre paramètres ; certains n'ont même pas su comment remplacer la probabilité conditionnelle $\mathbb{P}_M(R)$. Il est indéniable que ce qui a posé problème à certains, c'est la compréhension de l'énoncé, et le lien entre les phrases et les calculs à faire ? Il paraît **essentiel** de travailler sur ce point : faire le lien entre une situation concrète et une mise en équation est à la base de l'activité scientifique.* »

b) On nous donne $P(M) = \dfrac{1}{40}$ et on cherche $P_R(M)$. Avec la formule de Bayes :

$$P_R(M) = \frac{P(M) \times P_M(R)}{P(R)} = \frac{P(M) \times P_M(R)}{P(M) \times P_M(R) + P(\overline{M}) \times P_{\overline{M}}(R)}.$$

On passe à l'application numérique :

$$P_R(M) = \frac{\frac{1}{40} \times \frac{5}{6}}{\frac{1}{40} \times \frac{5}{6} + \frac{39}{40} \times \frac{3}{20}} = \frac{50}{50 + 39 \times 9} = \frac{50}{401}.$$

$$\boxed{P_R(M) \approx \frac{1}{8} \approx 0.125.}$$

Ainsi il y a seulement une 'chance' sur huit que l'individu soit réellement malade lorsque son test est positif.
Par conséquent, $\boxed{\text{cette méthode de test est très mauvaise.}}$

Lu dans le rapport de jury : « *On doit ici utiliser deux formules. D'abord la définition des probabilités conditionnelles :*

$$\mathbb{P}_R(M) = \frac{\mathbb{P}(M \cap R}{\mathbb{P}(R)} = \frac{\mathbb{P}_M(R)\mathbb{P}(M)}{\mathbb{P}(R)}$$

et la formule des probabilités totales pour calculer :

$$\mathbb{P}(R) = \mathbb{P}(R \cap M) + \mathbb{P}(R \cap \overline{M}) = \mathbb{P}_M(R)\mathbb{P}(M) + \mathbb{P}_{\overline{M}}(R)\mathbb{P}(\overline{M}) = \frac{5}{6} \times \frac{1}{40} + \frac{3}{20} \times \frac{39}{40} = \frac{401}{2400}$$

D'où on déduit que :

$$\mathbb{P}_R(M) \approx 0.12$$

On peut aussi appliquer directement le théorème de Bayes (tel que dans la correction ci-dessus) à condition de ne pas écrire n'importe quoi au dénominateur (on a trouvé, parfois, le dénominateur $\mathbb{P}_M(R) + \mathbb{P}_{\overline{M}}(R)$). Certains candidats ont opté pour cette alternative, et sont parvenus au bon résultat, alors qu'ils n'ont pas su faire la question précédente ; et pourtant, ce faisant, ils ont donc trouvé (à leur insu) les réponses à celle-ci.
On a fait le calcul ci-dessus pour commenter la valeur trouvée (c'était la dernière question du sujet, là aussi quelques candidats n'ont pas vu qu'il y avait deux questions en une) ; le test détecte peut-être cinq malades sur six (c'est le résultat de la question précédente) mais lorsqu'il est positif, il s'agit dans 88% de cas d'un individu qui n'est en fait pas malade : ce test est inutilisable. Un nombre finalement assez restreint de candidats ont osé tirer cette conclusion ; parmi les autres, certains ont été jusqu'à écrire qu'ils « avaient dû faire une erreur quelque part ». Il faut avoir confiance en ses résultats, sans compter qu'ici, le calcul est suffisamment court pour pouvoir le relire et trouver si une erreur s'y est glissée. »

- FIN -