Dans ce sujet on s'intéresse à un problème d'arrêt optimal c'est-à-dire au problème du choix d'un moment pour entreprendre une action spécifique, afin de maximiser un gain attendu ou de minimiser un coût attendu.

Des problèmes d'arrêts optimaux peuvent être trouvés dans les domaines des statistiques, de l'économie et des mathématiques financières (par exemple dans la tarification des options américaines).

On peut modéliser le problème étudié comme suit. Supposons que nous recevions une suite finie de nombres réels, un par un. Ces réels sont des réalisations de variables aléatoires indépendantes de lois connues à l'avance mais pas nécessairement identiques. Nous ne pouvons garder qu'un seul nombre de la suite. À chaque observation, nous pouvons soit sélectionner le nombre actuel, soit pousser notre chance et passer à l'observation suivante. Notre objectif est de maximiser la probabilité de sélectionner le nombre maximal de la suite.

Dans tout le sujet on considère un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$. Toutes les variables aléatoires réelles et événements qui interviennent dans cet énoncé sont définies sur cet espace.

On rappelle que si A désigne un événement, $\mathbb{1}_A$ est la variable aléatoire qui vaut 1 sur A et 0 sur $\overline{A}$.

La partie 2 utilise des résultats de la partie 1. La partie 3 est indépendante des deux premières et seules les deux dernières questions 20 et 21 de la partie 4 utilisent des résultats établis dans les parties précédentes.

Un aide-mémoire **Python** se trouve à la fin de l'énoncé. Pour les scripts et fonctions **Python**, on supposera que les instructions suivantes ont été exécutées :

```
import numpy as np, numpy.random as rd, matplotlib.pyplot as plt
```

Partie I - Des résultats généraux

1. Montrer que pour tout t réel, $\mathrm{e}^t \geqslant 1+t$ et que, pour tout $t > -1$, $\ln(1+t) \leqslant t$.
2. On définit la fonction f sur $[0;1]$ par $f(t) = (1+t)\mathrm{e}^{-t} - t$.
   - 2.a. Étudier les variations de f et montrer qu'il existe un unique $\alpha \in]0;1[$ tel que $f(\alpha) = 0$ et que pour tout $t \in [0;1]$, $f(t) > 0 \iff t < \alpha$.
   - 2.b. Écrire un programme **Python** qui renvoie une valeur approchée de α à 10^{-3} près.
   - 2.c. Montrer que pour tout $t \in [0;1]$, $\mathrm{e}^t \leqslant 1+2t$. En déduire que $\alpha \geqslant \dfrac{1}{\sqrt{2}}$.
3. Soit U une variable aléatoire à densité, à valeurs dans $[0;1[$, qui suit la loi uniforme et $p \in]0;\alpha[$. On pose $\beta = f(p)$. On définit les variables aléatoires X et Y par :

$$X = \mathbb{1}_{[\beta < U \leqslant \beta + p]} \text{ et pour tout } \omega \in \Omega,\ Y(\omega) = \min\left\{k \in \mathbb{N} \ /\ U(\omega) \leqslant \sum_{i=0}^{k} \frac{p^i}{i!}\mathrm{e}^{-p}\right\}$$

   - 3.a. Rappeler la valeur de $\displaystyle\sum_{i=0}^{+\infty} \frac{p^i}{i!}\mathrm{e}^{-p}$.
     En déduire que la variable aléatoire Y est bien définie et déterminer la loi de X.
   - 3.b. Écrire une fonction **Python**, minimum(x,p) qui renvoie le minimum de l'ensemble $\left\{k \in \mathbb{N} \ /\ x \leqslant \displaystyle\sum_{i=0}^{k} \frac{p^i}{i!}\mathrm{e}^{-p}\right\}$ lorsque $x \in [0;1[$ et $p \in]0;1[$. En déduire une fonction **simulY(p)** qui réalise une simulation de Y.
   - 3.c. Montrer que Y suit la loi de Poisson de paramètre p.
   - 3.d. Soit k un entier, $k \geqslant 2$. Montrer que si $[Y = k]$ est réalisé alors $[X = 0]$ l'est.
     En déduire que $\mathbb{P}\big([X = 0] \cap [Y = k]\big) = \dfrac{p^k}{k!}\mathrm{e}^{-p}$.
   - 3.e. Montrer que $[X = 0] \cap [Y = 1] = \varnothing$. En déduire que

$$\mathbb{P}\big([X=1]\cap[Y=1]\big) = p\mathrm{e}^{-p} \quad ; \quad \mathbb{P}\big([X=0]\cap[Y=0]\big) = \beta \quad ; \quad \mathbb{P}\big([X=1]\cap[Y=0]\big) = p\big(1-\mathrm{e}^{-p}\big)$$

   - 3.f. Montrer que $\mathbb{P}\big([X \neq Y]\big) = 1 + p - (1+2p)\mathrm{e}^{-p}$, puis que $\mathbb{P}\big([X \neq Y]\big) \leqslant 2p^2$.
4. *Inégalité de Boole* - Soit $(B_k)_{k\in\mathbb{N}^*}$ une suite d'évènements. Démontrer :

$$\forall n \in \mathbb{N}^*,\ \mathbb{P}\left(\bigcup_{k=1}^{n} B_k\right) \leqslant \sum_{k=1}^{n} \mathbb{P}(B_k)$$

5. Soient X et Y deux variables aléatoires à valeurs dans $\mathbb{N}$. On pose

$$\delta(X,Y) = \sum_{k=0}^{+\infty} \big|\mathbb{P}\big([X=k]\big) - \mathbb{P}\big([Y=k]\big)\big| \quad \text{et} \quad d(X,Y) = \mathbb{P}\big([X \neq Y]\big)$$

5.a. Justifier que la série définissant $\delta(X, Y)$ est bien convergente.

5.b. Soit $k \in \mathbb{N}$ tel que $\mathbb{P}([X = k]) \geqslant \mathbb{P}([Y = k])$. Montrer que $\left|\mathbb{P}([X = k]) - \mathbb{P}([Y = k])\right| \leqslant \mathbb{P}([X = k] \cap [Y \neq k])$.

5.c. En déduire que pour tout $k \in \mathbb{N}$,

$$\left|\mathbb{P}([X = k]) - \mathbb{P}([Y = k])\right| \leqslant \mathbb{P}([X = k] \cap [Y \neq k]) + \mathbb{P}([X \neq k] \cap [Y = k])$$

5.d. En conclure que $\delta(X, Y) \leqslant 2d(X, Y) \leqslant 2$.

Partie 2 – Une inégalité d'après Hodge et Le Cam

On conserve les notations de la partie 1.

On considère une suite $(U_k)_{k\in\mathbb{N}^*}$ de variables aléatoires à densité indépendantes qui suivent la loi uniforme sur $[0, 1[$.

Soit $n \in \mathbb{N}^*$ et $p_1, \ldots, p_n$ des réels appartenant à $]0; 1[$. Pour tout $k \in [\![1; n]\!]$, on définit, comme X et Y dans la question 3. de la partie 1, X_k et Y_k avec U_k et p_k à la place de U et p. On pose $\lambda_n = \sum_{k=1}^{n} p_k$, $S_n = \sum_{k=1}^{n} X_k$ et $T_n = \sum_{k=1}^{n} Y_k$.

On souhaite établir l'inégalité : $\delta(S_n, T_n) \leqslant 4 \sum_{k=1}^{n} p_k^2$ (LC).

6. Montrer que si l'un au moins des p_k est supérieur ou égal à α alors (LC) est vérifiée.

► **On suppose dans la suite de cette partie que pour tout $k \in [\![1; n]\!]$, $p_k < \alpha$.**

7. Justifier brièvement que $X_1, \ldots, X_n$ (respectivement $Y_1, \ldots, Y_n$) sont indépendantes.

8. Quelle est la loi de T_n ? Si les p_k sont tous égaux à $\frac{\lambda}{n}$ (avec λ un réel strictement positif), quelle est la loi de S_n ? Quelle est alors la limite en loi de la suite $(S_n)_{n\in\mathbb{N}^*}$?

9. **9.a.** Montrer que $[S_n \neq T_n] \subset \bigcup_{k=1}^{n} [X_k \neq Y_k]$.

   9.b. En déduire que pour tout $n \in \mathbb{N}^*$, $\delta(S_n, T_n) \leqslant 4 \sum_{k=1}^{n} p_k^2$.

10. *Un cas particulier* – On suppose dans cette question que tous les p_k sont égaux à $\frac{\lambda}{n}$ (avec λ un réel strictement positif). Montrer que pour tout $n \in \mathbb{N}^*$,

$$\sum_{k=0}^{n} \left| \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} - \frac{\lambda^k}{k!} e^{-\lambda} \right| \leqslant \frac{4\lambda^2}{n}$$

11. *Une application* – Soit n un entier, $n \geqslant 2$, on réalise n expériences de Bernoulli indépendantes avec les probabilités de succès respectives $\frac{1}{n+1}, \frac{1}{n+2}, \ldots, \frac{1}{2n}$. On note S_n le nombre de succès de l'expérience totale et $s_n = \sum_{k=1}^{n} \frac{1}{n+k}$.

    11.a. Montrer que pour tout $k \in \mathbb{N}$, $\left|\mathbb{P}([S_n = k]) - \frac{s_n^k}{k!} e^{-s_n}\right| \leqslant \frac{4}{n}$.

    11.b. Établir que pour tout $k \in [\![1; n]\!]$,

$$\int_k^{k+1} \frac{1}{n+t} dt \leqslant \frac{1}{n+k} \leqslant \int_{k-1}^{k} \frac{1}{n+t} dt$$

    11.c. En déduire un encadrement de s_n puis que $\lim_{n\to+\infty} s_n = \ln(2)$.

    11.d. Conclure que $(S_n)_{n\geqslant 2}$ converge en loi vers une variable aléatoire S qui suit la loi de Poisson de paramètre $\ln(2)$.

Partie 3 – Étude du maximum d'une fonction

Pour tout $x \in [0; +\infty[$, on pose $h(x) = \begin{cases} \dfrac{e^x - 1}{x} & \text{si } x > 0 \\ 1 & \text{si } x = 0 \end{cases}$.

12. Montrer que h est de classe $\mathscr{C}^1$ sur $\mathbb{R}^+$ et préciser la valeur de $h'(0)$.

On définit alors la fonction g sur $\mathbb{R}^+$ par $g(x) = e^{-x} \int_0^x h(t) dt$.

13. **13.a.** Montrer que pour tout $x \geqslant 0$, $g'(x) = e^{-x} \left(1 - \int_0^x \big(h(t) - h'(t)\big) dt\right)$.

    13.b. Montrer que pour tout $t > 0$, $h(t) - h'(t) = \frac{e^t - 1 - t}{t^2}$.

    13.c. En déduire que pour tout $t \geqslant 0$, $h(t) - h'(t) > 0$ et $\lim_{x\to+\infty} \int_0^x \big(h(t) - h'(t)\big) dt = +\infty$.

    13.d. Dans un même tableau, représenter le signe de g' et les variations de g en justifiant que g possède un maximum qui est atteint en un unique réel noté $\gamma > 0$ que l'on fera apparaître dans ce tableau.

14. **14.a.** Montrer que pour tout $t > 0$, $\dfrac{1}{2} \leqslant \dfrac{e^t - 1 - t}{t^2} \leqslant \dfrac{1}{2}e^t$.

14.b. En déduire que pour tout $x \geqslant 0$, $e^{-x}\dfrac{3 - e^x}{2} \leqslant g'(x) \leqslant e^{-x}\left(1 - \dfrac{x}{2}\right)$.

14.c. En conclure que $\gamma \in [\ln(3); 2]$.

15. **15.a.** Soit $n \in \mathbb{N}^*$ et $x \in [0; n]$, montrer que :

$$\sum_{k=0}^{n} \frac{x^k}{k!} \leqslant e^x \leqslant \sum_{k=0}^{n} \frac{x^k}{k!} + \frac{x^n}{n!} \sum_{k=n+1}^{+\infty} \left(\frac{x}{n+1}\right)^{k-n} \leqslant \sum_{k=0}^{n} \frac{x^k}{k!} + \frac{x^{n+1}}{n!}$$

et en déduire que $\displaystyle\sum_{k=1}^{n} \frac{x^{k-1}}{k!} \leqslant h(x) \leqslant \sum_{k=1}^{n} \frac{x^{k-1}}{k!} + \frac{x^n}{n!}$.

15.b. En conclure que pour tout $n \in \mathbb{N}^*$ et $x \in [0; n]$,

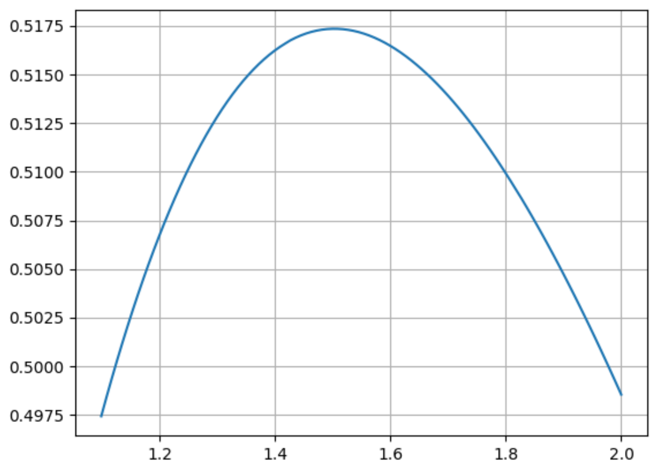
$$e^{-x}\left(\sum_{k=1}^{n} \frac{x^k}{k!k}\right) \leqslant g(x) \leqslant e^{-x}\left(\sum_{k=1}^{n} \frac{x^k}{k!k}\right) + e^{-x}\frac{x^{n+1}}{(n+1)!}$$

puis que pour tout $x \geqslant 0$, $g(x) = e^{-x}\displaystyle\sum_{k=1}^{+\infty} \frac{x^k}{k!k}$.

16. Compléter le programme suivant pour qu'il trace la partie de la courbe de g comprise entre les abscisses $\ln(3)$ et 2, les valeurs de g étant calculées à 10^{-4} près :

```
X = np.linspace(np.log(3),2,100)
Y = []
for x in X:
    n = 2; s = x + x**2/4; d = x**3/6
    while d * np.exp(-x) > 0.0001:
        n = n + 1
        s = s + ...
        d = d * x / ...
    Y.append(s * ...)
plt.plot(X,Y)
plt.grid()
plt.show()
```

On obtient le graphique suivant :



Partie 4 – Le problème du meilleur choix

On considère $(X_k)_{k\in\mathbb{N}^*}$ une suite de variables aléatoires indépendantes à densité.

Soit $s \in \mathbb{R}$, pour tout $n \in \mathbb{N}^*$ on définit des variables aléatoires, $a_{n,s}$, $Y_{n,s}$, Z_n et $K_{n,s}$ par, pour tout $\omega \in \Omega$:

$$a_{n,s}(\omega) = \min\left(\{k \in [\![1; n]\!] \ / \ X_k(\omega) > s\} \cup \{n\}\right) \quad ; \quad Y_{n,s}(\omega) = X_{a_{n,s}(\omega)}(\omega) \quad ; \quad Z_n(\omega) = \max_{k\in[\![1;n]\!]}\left(X_k(\omega)\right)$$

et $K_{n,s}(\omega)$ est égal au nombre d'indices $k \in [\![1; n]\!]$ tels que $X_k(\omega) > s$.
On cherche à choisir s pour maximiser $r_n = \mathbb{P}\left([Y_{n,s} = Z_n]\right)$.
On pose pour tout $k \in [\![1; n]\!]$, $p_k = \mathbb{P}\left([X_k > s]\right)$ et on suppose que $p_k \neq 1$.

17. *Une minoration dans le cas général.*

17.a. Montrer que $\mathbb{P}\big([Y_{n,s} = Z_n]\big) \geqslant \mathbb{P}\big([K_{n,s} = 1]\big)$.

17.b. On pose $\theta = \mathbb{P}\big([Z_n \leqslant s]\big)$. Montrer que $\mathbb{P}\big([K_{n,s} = 1]\big) = \theta \displaystyle\sum_{k=1}^{n} \frac{p_k}{1-p_k}$.

17.c. En déduire que $\mathbb{P}\big([Y_{n,s} = Z_n]\big) \geqslant -\theta \ln(\theta)$.

17.d. En déduire l'existence d'au moins une valeur de s, que l'on définira à l'aide de la fonction de répartition F_n de Z_n, pour laquelle $\mathbb{P}\big([Y_{n,s} = Z_n]\big) \geqslant \dfrac{1}{\mathrm{e}}$.

▶ On suppose désormais que les X_k suivent la même loi donc que les p_k sont tous égaux, non nuls. On note p cette valeur commune, F la fonction de répartition et f une densité communes aux X_k.

On admet que si X et Y sont deux variables à densité indépendantes, de fonction de répartition F_X pour X et de densité f_Y pour Y, on a alors $\mathbb{P}\big([X \leqslant Y]\big) = \displaystyle\int_{-\infty}^{+\infty} F_X(t) f_Y(t)dt$.

On rappelle que si A est un événement de probabilité non nulle, $(\Omega, \mathcal{A}, \mathbb{P}_A)$ est un espace probabilisé admettant les mêmes variables aléatoires et ayant les mêmes propriétés que $(\Omega, \mathcal{A}, \mathbb{P})$.

18. *Une estimation* - On suppose dans cette question les X_k suivent la loi uniforme sur $[0;1[$.

18.a. Montrer que $s = 1 - p$.

18.b. Écrire une fonction **Python** `simulCouple(n,p)` qui renvoie une simulation du couple $(Z_n, Y_{n,s})$.

18.c. Écrire un programme **Python** qui réalise et affiche une estimation de r_{10} dans ces conditions lorsque $n = 10$ et $p = 0,15$.

19. *Une expression explicite de r_n* - Soit $k \in [\![1;n]\!]$, on note $I_1, \dots, I_{\binom{n}{k}}$ les parties à k éléments de $[\![1;n]\!]$. Pour tout $j \in \left[\!\!\left[1; \dbinom{n}{k}\right]\!\!\right]$, on définit A_j l'événement $\left(\displaystyle\bigcap_{i \in I_j} [X_i > s]\right) \cap \left(\displaystyle\bigcap_{i \notin I_j} [X_i \leqslant s]\right)$.

19.a. Montrer que pour tout x réel et $i \in I_j$,

$$\mathbb{P}_{A_j}\big([X_i \leqslant x]\big) = \begin{cases} \dfrac{F(x) - F(s)}{p} & \text{si } x > s \\ 0 & \text{sinon} \end{cases}$$

et que les X_i pour $i \in I_j$ sont indépendantes pour la probabilité $\mathbb{P}_{A_j}$.

19.b. En déduire que pour tout $r \in I_j$:

$$\mathbb{P}_{A_j}\left(\left[X_r = \max_{i \in I_j}(X_i)\right]\right) = \frac{1}{p^k} \int_s^{+\infty} \big(F(t) - F(s)\big)^{k-1} f(t)dt = \frac{1}{k}$$

19.c. Montrer que :

$$\mathbb{P}\big([Y_{n,s} = Z_n] \cap [K_{n,s} = k]\big) = \sum_{j=1}^{\binom{n}{k}} \mathbb{P}\big([Y_{n,s} = Z_n] \cap A_j\big) \text{ et que } \mathbb{P}_{A_j}\big([Y_{n,s} = Z_n]\big) = \frac{1}{k}$$

19.d. Montrer de même que $\mathbb{P}\big([Y_{n,s} = Z_n] \cap [K_{n,s} = 0]\big) = \dfrac{1}{n}(1-p)^n$.

19.e. En conclure que $r_n = \dfrac{1}{n}(1-p)^n + \displaystyle\sum_{k=1}^{n} \frac{1}{k} \binom{n}{k} p^k (1-p)^{n-k}$.

20. *Comportement asymptotique* - On suppose que $p = \dfrac{\lambda}{n}$, λ étant un réel strictement positif ne dépendant pas de n.

20.a. En utilisant un résultat de la partie 2, montrer que pour tout $n \in \mathbb{N}^*$,

$$\sum_{k=1}^{n} \frac{1}{k} \left| \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} - \frac{\lambda^k}{k!} \mathrm{e}^{-\lambda} \right| \leqslant \frac{4\lambda^2}{n}$$

20.b. En déduire que $\displaystyle\lim_{n \to +\infty} r_n = \left(\sum_{k=1}^{+\infty} \frac{\lambda^k}{k!k}\right) \mathrm{e}^{-\lambda}$.

21. On suppose que n est assez grand pour pouvoir considérer que r_n vaut $\left(\displaystyle\sum_{k=1}^{+\infty} \frac{\lambda^k}{k!k}\right) \mathrm{e}^{-\lambda}$. Comment choisir s pour que cette probabilité soit maximale ?

★★★★★★★ FIN ★★★★★★★