FEUILLE DE TD/TP 4
SIMULATION D'ÉVÉNEMENTS RARES

Exercice 1. Des requêtes informatiques sont traitées successivement et on s'intéresse à la probabilité que le temps total de traitement soit grand. On modélise les requêtes par une suite de vaiid $X_1, \dots, X_n$ de loi exponentielle $\mathcal{E}(1)$. On note f la densité de la loi $\mathcal{E}(1)$, on pose $S_n = \sum_{i=1}^n X_i$ et on cherche donc à estimer la probabilité

$$p_n = \mathbb{P}(S_n \geq n(1+\epsilon))$$

pour $\epsilon > 0$ fixé.

(1) Écrire une fonction `MC_1` renvoyant l'approximation Monte-Carlo avec N tirages, l'écart-type empirique associé, ainsi que la largeur et les bornes de l'intervalle de confiance asymptotique de niveau 0,95 en utilisant N suites de n vaiid $X_1, \dots, X_n$ de loi exponentielle $\mathcal{E}(1)$. Tester avec avec $N = 10^3$, $n = 50$ et $\epsilon = 1$.

```
def MC_1(sample_size,n,epsilon):
    result = np.zeros(sample_size)
    for k in range(sample_size):
        result[k] = (np.sum(-np.log(np.random.rand(n))) >= n*(1+epsilon)).astype(int)
    mean = np.mean(result)
    std = np.std(result)
    icsize = 1.96*std/np.sqrt(sample_size)
    return np.array([mean,std,mean-icsize,mean+icsize])
```

(2) On utilise maintenant la méthode de réduction de variance par inférence préférentielle.

(a) Quelle est la valeur maximale $\theta_{\max}$ telle que pour $\theta < \theta_{\max}$, $M_\theta = \mathbb{E}\left[e^{\theta X_1}\right] < \infty$?

Pour tout $\theta \in \mathbb{R}$, on a

$$\mathbb{E}\left[e^{\theta X_1}\right] = \int_0^\infty e^{\theta x} e^{-x} dx = \int_0^\infty e^{(\theta-1)x} dx.$$

Cette intégrale diverge vers $+\infty$ si $\theta = 1$, et si $\theta \neq 1$, elle vaut

$$\int_0^\infty e^{(\theta-1)x} dx = \lim_{x\to\infty} \frac{e^{(\theta-1)x} - 1}{\theta - 1} = \begin{cases} +\infty & \text{si } \theta > 1 \\ \frac{1}{1-\theta} & \text{si } \theta < 1. \end{cases}$$

On en déduit que $\theta_{\max} = 1$.

(b) Pour $\theta < \theta_{\max}$, on considère la densité de probabilité f_θ définie par :

$$\forall x \in \mathbb{R},\ f_\theta(x) = \frac{1}{M_\theta} e^{\theta x} f(x)\,.$$

De quelle loi usuelle, f_θ est-elle la densité ?

D'après les calculs de la question (a), pour tout $\theta < 1$ on a $M_\theta = \frac{1}{1-\theta}$. On en déduit que pour tout $x \in \mathbb{R}$,

$$f_\theta(x) = (1-\theta)e^{\theta x} f(x) = (1-\theta)e^{-(1-\theta)x} \mathbb{1}_{\{x>0\}},$$

donc f_θ est la densité de la loi exponentielle de paramètre $1-\theta$.

(c) Soit X_θ une v.a. de densité f_θ. Pour quelle valeur de θ^* a-t-on $\mathbb{E}[X_{\theta^*}] = 1+\epsilon$?

Pour $\theta^* < 1$, $\mathbb{E}[X_{\theta^*}] = \frac{1}{1-\theta}$ donc $\mathbb{E}[X_{\theta^*}] = 1+\epsilon \Leftrightarrow 1 = (1+\epsilon)(1-\theta^*) \Leftrightarrow \epsilon = \theta^*(1+\epsilon) \Leftrightarrow \theta^* = \frac{\epsilon}{1+\epsilon}$.

(d) Soit $X_{1,\theta^*}, \dots, X_{n,\theta^*}$ n vaiid de densité f_{θ^*}. Exprimer p_n en fonction de $\sum_{i=1}^n X_{i,\theta^*}$.

Pour $\theta \in (0,1)$, on a par théorème de transfert

$$\begin{aligned} p_n = \mathbb{P}(S_n \geq n(1+\epsilon)) &= \int_{\mathbb{R}^n} \mathbb{1}_{\{\sum_{i=1}^n x_i \geq n(1+\epsilon)\}} \prod_{i=1}^n e^{-x_i} dx_1 \dots dx_n \\ &= \int_{\mathbb{R}^n} \mathbb{1}_{\{\sum_{i=1}^n x_i \geq n(1+\epsilon)\}} \prod_{i=1}^n M_{\theta^*} e^{-\theta^* x_i} \prod_{i=1}^n f_{\theta^*}(x_i) dx_1 \dots dx_n \end{aligned}$$

$$= \int_{\mathbb{R}^n} \mathbb{1}_{\{\sum_{i=1}^n x_i \geq n(1+\epsilon)\}} M_{\theta^*}^n e^{-\theta^* \sum_{i=1}^n x_i} \prod_{i=1}^n f_{\theta^*}(x_i) dx_1 \dots dx_n$$
$$= \mathbb{E}\left[Z_n^{\theta^*}\right].$$

où $Z_n^{\theta^*} = \mathbb{1}_{\{\sum_{i=1}^n X_{i,\theta^*} \geq n(1+\epsilon)\}} M_{\theta^*}^n e^{-\theta^* \sum_{i=1}^n X_{i,\theta^*}}$

(e) Écrire une fonction `MC_2` renvoyant l'approximation Monte-Carlo avec N tirages, l'écart-type empirique associé, ainsi que la largeur et les bornes de l'intervalle de confiance asymptotique de niveau 0,95 en utilisant des vaiid de densité f_{θ^*}. Tester avec $N = 10^3$, $n = 50$ et $\epsilon = 1$. Comparer avec la méthode proposée à la première question.

```
def MC_2(sample_size,n,epsilon):
    result = np.zeros(sample_size)
    theta = epsilon/(1+epsilon)
    for k in range(sample_size):
        sum_xi_theta = np.sum(-(1/(1-theta))*np.log(np.random.rand(n)))
        result[k] = (sum_xi_theta >= n*(1+epsilon)).astype(int) * ((1/(1-theta))**n) * np.exp(-theta * sum_xi_theta)
    mean = np.mean(result)
    std = np.std(result)
    icsize = 1.96*std/np.sqrt(sample_size)
    return np.array([mean,std,mean-icsize,mean+icsize])
```

Exercice 2. Comme dans l'exercice précédent, on s'intéresse à la probabilité que le temps total de traitement de n requêtes informatiques soit grand. Mais on modélise maintenant les requêtes par une suite de vaiid $X_1, \dots, X_n$ de loi uniforme $\mathcal{U}([0, M])$. On pose $S_n = \sum_{i=1}^n X_i$ et on cherche à estimer la probabilité

$$p_n = \mathbb{P}(S_n \geq nM(1 + \epsilon)/2)$$

pour $\epsilon > 0$ fixé.

(1) Écrire une fonction `MC_1` renvoyant l'approximation Monte-Carlo avec N tirages, l'écart-type empirique associé, ainsi que la largeur et les bornes de l'intervalle de confiance asymptotique de niveau 0,95 en utilisant N suites de n vaiid $X_1, \dots, X_n$ de loi uniforme $\mathcal{U}([0, M])$. Tester avec $N = 10^3$, $n = 50$, $\epsilon = 1/2$ et $M = 1$.

```
def MC_1(sample_size,M,n,epsilon):
    result = np.zeros(sample_size)
    for k in range(sample_size):
        result[k] = (np.sum((np.random.rand(0,M,n))) >= n*M*(1+epsilon)/2).astype(int)
    mean = np.mean(result)
    std = np.std(result)
    icsize = 1.96*std/np.sqrt(sample_size)
    return np.array([mean,std,mean-icsize,mean+icsize])
```

(2) On utilise maintenant la méthode de réduction de variance par inférence préférentielle.

(a) Quelle est la valeur maximale $\theta_{\max}$ telle que pour $\theta < \theta_{\max}$, $M_\theta = \mathbb{E}\left[e^{\theta X_1}\right] < \infty$?

La densité de la loi uniforme étant à support compact, on a $M_\theta < \infty$ pour tout $\theta \in \mathbb{R}$ donc $\theta_{\max} = \infty$.

(b) Pour $\theta < \theta_{\max}$, on considère la densité de probabilité f_θ définie par :

$$\forall x \in \mathbb{R},\ f_\theta(x) = \frac{1}{M_\theta} e^{\theta x} f(x).$$

Donner une forme explicite de f_θ.

On commence par calculer M_θ :

$$M_\theta = \frac{1}{M} \int_0^M e^{\theta x} dx = \frac{e^{\theta M} - 1}{M\theta}.$$

Il s'en suit que pour tout $x \in \mathbb{R}$,

$$f_\theta(x) = \frac{M\theta}{e^{\theta M} - 1} e^{\theta x} \frac{1}{M} \mathbb{1}_{\{x \in [0,M]\}} = \frac{\theta e^{\theta x}}{e^{\theta M} - 1} \mathbb{1}_{\{x \in [0,M]\}}.$$

(c) Soit X_θ une v.a. de densité f_θ. Montrer que pour $\epsilon < 1$, il existe θ^* tel que $\mathbb{E}[X_{\theta^*}] = M(1+\epsilon)/2$.

Soit $\theta > 0$ et $\epsilon \in (0,1)$. On intègre par parties :

$$\begin{aligned}\mathbb{E}[X_\theta] = \frac{\theta}{e^{\theta M}-1}\int_0^M xe^{\theta x}dx &= \frac{\theta}{e^{\theta M}-1}\left(\left[x\frac{e^{\theta x}}{\theta}\right]_0^M - \int_0^M \frac{e^{\theta x}}{\theta}dx\right)\\ &= \frac{\theta}{e^{\theta M}-1}\left(\frac{Me^{\theta M}}{\theta} - \frac{e^{\theta M}-1}{\theta^2}\right)\\ &= \frac{Me^{\theta M}}{e^{\theta M}-1} - \frac{1}{\theta}.\end{aligned}$$

Ainsi, on a $\mathbb{E}[X_\theta] = M(1+\epsilon)/2 \Leftrightarrow h(\theta) = \frac{1+\epsilon}{2}$, où

$$h(\theta) = \frac{e^{\theta M}}{e^{\theta M}-1} - \frac{1}{\theta M}.$$

La fonction h est continue sur $(0,\infty)$, il est clair que $\lim_{\theta\to\infty} h(\theta) = 1$, et on trouve avec un développement limité à l'ordre 2 de l'exponentielle en zéro que $\lim_{\theta\to 0^+} h(\theta) = \frac{1}{2}$. Comme $\epsilon \in (0,1)$, on conclut à l'aide du théorème des valeurs intermédiaires qu'il existe bien $\theta^* \in (0,\infty)$ tel que $h(\theta^*) = \frac{1+\epsilon}{2}$.

(d) Soit $X_{1,\theta^*}, \dots, X_{n,\theta^*}$ n vaiid de densité f_{θ^*}. Exprimer p_n en fonction de $\sum_{i=1}^n X_{i,\theta^*}$.

(e) Écrire une fonction `MC_2` renvoyant l'approximation Monte-Carlo avec n tirages, l'écart-type empirique associé, ainsi que la largeur et les bornes de l'intervalle de confiance asymptotique de niveau 0,95 en utilisant des vaiid de densité f_{θ^*}. Tester avec $N = 10^3$, $n = 50$, $\epsilon = 1/4$ et $M = 4$. Comparer avec la méthode proposée à la première question.

On pourra utiliser la routine suivante pour obtenir une valeur approchée de θ^*.

```
from scipy.optimize import root_scalar
import numpy as np
(...)
 def theta_solve(theta):
    return M*np.exp(theta*M)/(np.exp(theta*M)-1)-1/theta-M*(1+epsilon)/2

 theta_star=root_scalar(theta_solve, bracket=[0, 2/(M*(1-epsilon))]).root
```

Exercice 3. Une compagnie d'assurances souhaite estimer la probabilité que les remboursements qu'elle ait à effectuer dépasse une certaine somme de réserve x. Pour cela, elle modélise les différents remboursements par une suite de variables de loi de Pareto $\mathcal{P}(x_0, \alpha)$, $x_0 > 0$ et $\alpha > 0$, ie de fonction de répartition

$$\forall x \geq \mathbb{R},\ F(x) = \left(1 - \left(\frac{x_0}{x}\right)^\alpha\right)\mathbb{1}_{\{x\geq x_0\}}.$$

(1) Montrer que les lois de Pareto sont des lois à queue lourde.

Pour montrer que les lois de Pareto sont des lois à queue lourde, on veut montrer que si $X \sim \mathcal{P}(x_0,\alpha)$, alors pour tout $\theta > 0$ on a $\mathbb{E}\left[e^{\theta X}\right] = \infty$. Remarquons que comme F est dérivable presque partout, X admet une densité qui est donnée par

$$f(x) = \alpha\frac{x_0^\alpha}{x^{\alpha+1}}\mathbb{1}_{\{x\geq x_0\}}.$$

On en déduit que pour tout $\theta > 0$,

$$\mathbb{E}\left[e^{\theta X}\right] = \alpha x_0^\alpha \int_{x_0}^\infty \frac{e^{\theta x}}{x^{\alpha+1}}dx,$$

et il est clair que cette intégrale diverge en $+\infty$ par croissance comparée de l'exponentielle.

(2) Soit X une va de loi de Pareto $\mathcal{P}(x_0,\alpha)$. Montrer que $X \in \mathbb{L}^p$ pour tout $p < \alpha$ mais que $X \notin \mathbb{L}^\alpha$. Montrer que si $\alpha > 1$, $\mathbb{E}[X] = \alpha x_0/(\alpha-1)$.

Soit $p \in \mathbb{R}$. On calcule

$$\mathbb{E}[|X|^p] = \alpha x_0^\alpha \int_{x_0}^\infty \frac{x^p}{x^{\alpha+1}}dx = \alpha x_0^\alpha \int_{x_0}^\infty \frac{1}{x^{\alpha+1-p}}dx.$$

L'intégrale $\int_{x_0}^{\infty} \frac{1}{x^{\alpha+1-p}} dx$ converge si et seulement si $\alpha + 1 - p > 1 \Leftrightarrow p < \alpha$, ce qui montre que $X \in \mathbb{L}^p \Leftrightarrow p < \alpha$. Si $\alpha > 1$, on a $\lim_{x\to\infty} x^{1-\alpha} = 0$, d'où :

$$\mathbb{E}[X] = \alpha x_0^\alpha \int_{x_0}^{\infty} \frac{1}{x^\alpha} dx = \alpha x_0^\alpha \left[\frac{x^{1-\alpha}}{1-\alpha}\right]_{x_0}^{\infty} = \alpha x_0^\alpha \frac{x_0^{1-\alpha}}{\alpha - 1} = \frac{\alpha x_0}{\alpha - 1}.$$

On consiède maintenant n variables aléatoires iid $X_1, \ldots, X_n$ de loi $\mathcal{P}(x_0, \alpha)$, $x_0 > 0$ et $\alpha > 0$ et on s'intéresse à leur somme $S_n = \sum_{i=1}^n X_i$, plus précisément à la probabilité qu'elle dépasse une certaine valeur $x > 0$, supposée grande :

$$p_x = \mathbb{P}(S_n \geq x)$$

(3) Écrire une fonction `MC_1` renvoyant l'approximation Monte-Carlo avec N tirages, l'écart-type empirique associé, ainsi que la largeur et les bornes de l'intervalle de confiance asymptotique de niveau 0,95 en utilisant N suites de n vaiid $X_1, \ldots, X_n$ de loi $\mathcal{P}(x_0, \alpha)$. Tester avec $N = 10^3$, $n = 50$, $x_0 = 1$, $\alpha = 1.5$ et $x = 300$.

(4) On utilise maintenant la réduction de variance par conditionnement présentée en cours :

$$p_x = \mathbb{E}\left[1 - F(\max(x - S_{n-1}, \max_{1\leq i\leq n-1} X_i)\right].$$

Écrire une fonction `MC_2` renvoyant l'approximation Monte-Carlo avec N tirages, l'écart-type empirique associé, ainsi que la largeur et les bornes de l'intervalle de confiance asymptotique de niveau 0,95 en utilisant cette seconde méthode. Tester avec $N = 10^3$, $n = 50$, $x_0 = 1$, $\alpha = 1.5$ et $x = 300$. Comparer avec la méthode précédente.

Exercice 4. Une compagnie d'assurances souhaite estimer la probabilité que les remboursements qu'elle ait à effectuer dépasse une certaine somme de réserve x. Pour cela, elle modélise les différents remboursements par une suite de variables de loi log-logistique $LL(\alpha, \beta)$, $\alpha, \beta > 0$, de fonction de répartition

$$\forall x \in \mathbb{R},\ F(x) = \frac{x^\beta}{\alpha^\beta + x^\beta} \mathbb{1}_{x\geq 0}.$$

(1) Montrer que les lois log-logistiques sont des lois à queue lourde.

(2) Soit X une va de loi log-logistique $LL(\alpha, \beta)$, $\alpha, \beta > 0$. Montrer que $X \in \mathbb{L}^p$ pour tout $p < \beta$ mais que $X \notin \mathbb{L}^\beta$. Montrer que si $\beta > 1$, $\mathbb{E}[X] = \frac{\alpha\pi}{\beta \sin\frac{\pi}{\beta}}$.

On considère maintenant n variables aléatoires iid $X_1, \ldots, X_n$ de loi $LL(\alpha, \beta)$, $\alpha, \beta > 0$ et on s'intéresse à leur somme $S_n = \sum_{i=1}^n X_i$, plus précisément à la probabilité qu'elle dépasse une certaine valeur $x > 0$, supposée grande :

$$p_x = \mathbb{P}(S_n \geq x)$$

(3) Écrire une fonction `MC_1` renvoyant l'approximation Monte-Carlo avec N tirages, l'écart-type empirique associé, ainsi que la largeur et les bornes de l'intervalle de confiance asymptotique de niveau 0,95 en utilisant N suites de n vaiid $X_1, \ldots, X_n$ de loi $\mathcal{P}(x_0, \alpha)$. Tester avec $N = 10^3$, $n = 50$, $\alpha = 1$, $\beta = 2$ et $x = 200$.

(4) On utilise maintenant la réduction de variance par conditionnement présentée en cours :

$$p_x = \mathbb{E}\left[1 - F(\max(x - S_{n-1}, \max_{1\leq i\leq n-1} X_i)\right].$$

Écrire une fonction `MC_2` renvoyant l'approximation Monte-Carlo avec N tirages, l'écart-type empirique associé, ainsi que la largeur et les bornes de l'intervalle de confiance asymptotique de niveau 0,95 en utilisant cette seconde méthode. Tester avec $N = 10^3$, $n = 50$, $n = 50$, $\alpha = 1$, $\beta = 2$ et $x = 200$. Comparer avec la méthode précédente.