# 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 $X_1, \ldots, X_n$ 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 $X_1, \ldots, X_n$ de loi exponentielle $\mathcal{E}(1)$. Tester avec avec $N = 10^3$, $n = 50$ et $\epsilon = 1$.

(2) On utilise maintenant la méthode de réduction de variance par inférence préférentielle.
  (a) Quelle est la valeur maximale $\theta_{\max}$ telle que pour $\theta < \theta_{\max}$, $M_\theta = \mathbb{E}\left[e^{\theta X_1}\right] < \infty$ ?
  (b) Pour $\theta < \theta_{\max}$, on considère la densité de probabilité $f_\theta$ définie par :
  $$\forall x \in \mathbb{R},\ f_\theta(x) = \frac{1}{M_\theta} e^{\theta x} f(x)\,.$$
  De quelle loi usuelle, $f_\theta$ est-elle la densité ?
  (c) Soit $X_\theta$ une v.a. de densité $f_\theta$. Pour quelle valeur de $\theta^*$ a-t-on $\mathbb{E}[X_{\theta^*}] = 1 + \epsilon$ ?
  (d) Soit $X_{1,\theta^*}, \ldots, X_{n,\theta^*}$ $n$ vaiid de densité $f_{\theta^*}$. 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 $f_{\theta^*}$. Tester avec $N = 10^3$, $n = 50$ et $\epsilon = 1$. Comparer avec la méthode proposée à la première question.

**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 $X_1, \ldots, X_n$ 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 $X_1, \ldots, X_n$ de loi uniforme $\mathcal{U}([0, M])$. Tester avec $N = 10^3$, $n = 50$, $\epsilon = 1/2$ et $M = 1$.

(2) On utilise maintenant la méthode de réduction de variance par inférence préférentielle.
  (a) Quelle est la valeur maximale $\theta_{\max}$ telle que pour $\theta < \theta_{\max}$, $M_\theta = \mathbb{E}\left[e^{\theta X_1}\right] < \infty$ ?

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

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

Donner une forme explicite de $f_\theta$.

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

(d) Soit $X_{1,\theta^*}, \ldots, X_{n,\theta^*}$ $n$ vaiid de densité $f_{\theta^*}$. 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 $f_{\theta^*}$. 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 $\theta^*$.*

```
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.

(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 $\alpha > 1$, $\mathbb{E}[X] = \alpha x_0/(\alpha - 1)$.

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.