Examen Mag 372 – Modélisation en probabilités 11-05-2023

1. Le sujet comporte 2 exercices indépendants.
2. Vous ne pouvez pas vous connecter à votre session personnelle ; vous vous connecterez à la session `mathens` avec le mot de passe `mathens`. Tous vos fichiers devront être enregistrés dans le dossier "`reMise_copie`" présent sur le bureau. Il faut impérativement inclure votre nom dans le nom des fichiers.
3. L'idéal est de transmettre l'exportation en pdf de votre feuille Jupyter.
4. Les réponses aux questions théoriques doivent être données sur les copies papier.
5. Les graphes doivent avoir un titre et être légendés ; ils doivent être présents, le code qui les engendre ne suffit pas.
6. Pour toutes les questions qui demandent d'écrire un programme, on demande en plus du programme écrit une exécution du dit programme avec des paramètres convenablement choisis. On rappelle par ailleurs les commandes d'import des bibliothèques adéquates :

```python
import math
import numpy as np
import scipy.stats as scs
import matplotlib.pyplot as plt
import numpy.random as random
```

Exercice 1. Loi $\Gamma(2)$.

Soit X une variable aléatoire positive telle que, pour tous $0 \leq a \leq b$, on ait

$$\mathbb{P}[a \leq X \leq b] = C \int_a^b s\,\mathrm{e}^{-s}\,\mathrm{d}s$$

pour une certaine constante $C \geq 0$.

1. Montrer que la fonction de répartition de X s'écrit :

$$F_X(x) = \mathbb{P}[X \leq x] = \begin{cases} C(1-(x+1)\,\mathrm{e}^{-x}) & \text{si } x > 0, \\ 0 & \text{si } x \leq 0. \end{cases}$$

2. Que vaut $\lim_{x\to+\infty} F_X(x)$? Déterminer la valeur de la constante C.
3. Écrire un programme `fonction_repartition(m)` qui dessine la fonction de répartition de X sur l'intervalle $[0, m]$. Exécuter ce programme avec $m = 10$.
4. La variable aléatoire X peut être simulée par Python avec la commande :

```python
from scipy.stats import gamma
gamma.rvs(2, size=M)
```

où M est le nombre de simulations indépendantes que l'on souhaite faire. Écrire un programme `fonction_repartition_empirique(m, M)` qui réalise M simulations $X_1, \ldots, X_M$ de la variable X, et qui dessine la fonction de répartition empirique de ces variables sur l'intervalle $[0, m]$. On rappelle que c'est la fonction de répartition $F_X^{(M)}$ qui prend ses valeurs dans $\{0, \frac{1}{M}, \frac{2}{M}, \ldots, 1\}$ et qui vaut

$$F_X^{(M)}(t) = \frac{\mathrm{card}\left(\{i \mid 1 \leq i \leq M,\ X_i \leq t\}\right)}{M}.$$

Si $X_1^{\mathrm{ord}} \leq X_2^{\mathrm{ord}} \leq \cdots \leq X_M^{\mathrm{ord}}$ est le réordonnement croissant de $X_1, \ldots, X_M$, alors $F_X^{(M)}$ vaut $\frac{k}{M}$ entre X_k^{ord} et X_{k+1}^{ord}.

5. Superposer sur un même dessin la fonction de répartition F_X de la question 3. et la fonction de répartition empirique $F_X^{(M)}$ de la question 4., pour $m = 10$ et un nombre de simulations $M = 100$, puis pour $M = 10000$. Commenter ces dessins.

Exercice 2. Processus de Galton–Watson, une approche par la simulation.

Soit X une variable aléatoire intégrable à valeurs dans $\mathbb{N} = \{0, 1, 2, 3, \ldots\}$, d'espérance $\mu = \mathbb{E}[X]$. Soit $z_0 \in \mathbb{N}$, et soit $(X_{n,i})_{n \in \mathbb{N}, i \in \mathbb{N}^*}$ une famille de variables aléatoires i.i.d. de même loi que X. On pose alors : $Z_0 = z_0$ et pour tout $n \in \mathbb{N}$,

$$Z_{n+1} = \begin{cases} \sum_{i=1}^{Z_n} X_{n,i} & \text{si } Z_n \neq 0, \\ 0 & \text{si } Z_n = 0. \end{cases}$$

La suite $(Z_n)_{n \in \mathbb{N}}$ s'appelle le processus de Galton–Watson issu de z_0, de loi de reproduction la loi de X. Du point de vue de la modélisation, Z_n rend compte du nombre d'individus présents à la n-ième génération d'une population composée de z_0 individus à la génération 0 ; le i-ième individu de la génération n engendre $X_{n,i}$ individus à la génération $n+1$, qui se trouve donc composée au total de $\sum_{i=1}^{Z_n} X_{n,i}$ individus.

1. Montrer que $\mathbb{E}[Z_{n+1}] = \mu\, \mathbb{E}[Z_n]$. On pourra décomposer selon les valeurs prises par Z_n, en écrivant : $Z_{n+1} = \sum_{k \in \mathbb{N}} Z_{n+1} 1_{\{Z_n = k\}}$.
2. En déduire une expression de $\mathbb{E}[Z_n]$ en fonction de n, z_0 et μ. Discuter en fonction de z_0 et μ le comportement de la suite $(\mathbb{E}[Z_n])_{n \in \mathbb{N}}$ quand $n \to \infty$.

On suppose dans toute la suite que X suit une loi de Poisson de paramètre `theta`, de sorte que $\mu =$ `theta`. On rappelle aussi que la somme de k variables indépendantes de Poisson de paramètre `theta` est une variable aléatoire de Poisson de paramètre k `theta`.

Cas `theta` < 1.

3. Écrire une fonction `GaltonWatson(N,theta,z0)` qui simule une réalisation de $(Z_0, Z_1, \ldots, Z_{N-1})$ avec
   — `theta` le paramètre de la loi de Poisson,
   — $Z_0 =$ `z0` la valeur initiale du processus.

   La fonction `scs.poisson.rvs(theta, size=M)` permet d'engendrer un échantillon de taille `M` de variables aléatoires de Poisson indépendantes de paramètre `theta`. *Les paramètres* `N`, `theta`, `z0` *conserveront dans la suite le même sens que dans cette question.*
4. Écrire une fonction `SamplePathZ(N,M,theta,z0)` qui simule M réalisations indépendantes de $(Z_0, Z_1, \ldots, Z_{N-1})$ et superpose sur une même figure le tracé des `M` trajectoires de l'application $n \in \{0, 1, \ldots, N-1\} \mapsto Z_n$.

5. Évaluer la fonction précédente pour $(\texttt{N}, \texttt{M}, \texttt{theta}, \texttt{z0}) = (50, 10, 0.9, 5)$. Enregistrer la figure et commenter les résultats obtenus. A-t-on convergence presque sûre de la variable aléatoire Z_n lorsque n tend vers l'infini ? Si c'est le cas, quelle est la limite ?

Cas `theta` > 1.

6. Évaluer la fonction `SamplePathZ` précédente pour (`N`, `M`, `theta`, `z0`)$=(50, 10, 1.1, 5)$. Enregistrer la figure et commenter les résultats obtenus.

On pose alors, pour $n \in \mathbb{N}$,

$$W_n = \frac{Z_n}{\mu^n}.$$

7. Écrire une fonction `SamplePathW(N,M,theta,z0)` qui superpose sur un même graphe le tracé de `M` trajectoires indépendantes de l'application $n \mapsto W_n$.
8. Évaluer la fonction précédente pour (`N`, `M`, `theta`, `z0`)$=(100, 10, 1.1, 5)$. Enregistrer la figure et commenter les résultats obtenus. A-t-on convergence presque sûre ? Que dire de la limite si elle existe ?
9. Écrire une fonction `RepFuncW(N,M,theta,z0)` qui trace la fonction de répartition empirique de la variable aléatoire W_N à partir d'un échantillon de taille `M` de cette variable aléatoire.
10. Superposer les *trois* fonctions de répartition empiriques de la variable aléatoire W_N pour $(\texttt{M}, \texttt{theta}, \texttt{z0}) = (500, 1.1, 5)$ et $\texttt{N} \in \{10, 20, 100\}$. Enregistrer la figure. A-t-on convergence en loi ? Discuter des propriétés de la limite si celle-ci existe.