Partiel Mag 372 – Modélisation en probabilités 29-02-2024

1. Le sujet comporte 2 exercices indépendants.
2. Vous ne pouvez pas vous connecter à votre session personnelle, vous vous connectez 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. La réponse aux questions théoriques doit être sur les copies papier.
5. Les graphes doivent avoir un titre et être légendés (les graphes doivent être présents, le code qui les génère 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.
Soit une urne contenant au total $n = 20$ boules dont $r_0 = 10$ sont rouges et $b_0 = 10$ sont bleues. On se propose d'étudier le nombres de boules rouges tirées lors de $k = 10$ tirages d'abord lorsqu'on fait des tirages avec remise puis lorsqu'on fait des tirages sans remises.

1. On regarde le cas où après chaque tirage on remet la boule tirée. Quelle est la loi μ de X le nombre de boules rouges tirées au bout de k tirages ? Écrire un programme qui représente sur le même graphique le diagramme en bâton de la loi μ et le diagramme en bâton de la distribution empirique obtenue à partir de $N = 1000$ simulations indépendantes de X (c'est à dire le diagramme en baton de μ_N la probabilité sur $\{0; 1 \dots; 10\}$ telle que $\mu_N(\{k\}) = \frac{1}{N}\sum_i 1_{X_i=k}$ la proportion des réalisations égales à k). Qu'observez vous ?
2. On regarde maintenant le cas où après chaque tirage on ne remet pas la boule tirée. Écrire un programme qui simule la variable aléatoire Y qui est le nombre de boules rouges tirées au bout de k tirages sans remise (on pourra introduire les variables r et b qui représentent le nombre de boules rouges/bleues présentes et qui évolueront en fonction des tirages effectués).
3. Écrire un programme qui représente le diagramme en bâton de la distribution empirique obtenue à partir de $N = 1000$ réalisations indépendantes de Y. Représenter sur une même figure les diagrammes en bâton de loi empirique : celui pour le tirage avec remise et celui pour le tirage sans remise. À votre avis quelle variable à la plus grande variance : X ou Y ?

Exercice 2. On note $\mathfrak{X} = [\![-5, 5]\!] \times [\![-5, 5]\!]$ l'ensemble des points (x, y) du plan à coordonnées entières plus petites que 5 en valeur absolue. Autrement dit, ce sont les sommets d'une grille de côtés 10×10. Deux points (x_1, y_1) et (x_2, y_2) seront dits *voisins* si $x_1 = x_2$ et $|y_2 - y_1| = 1$, ou si $|x_2 - x_1| = 1$ et $y_1 = y_2$.

1. Le *degré* $\deg(x, y)$ d'un point (x, y) dans $\mathfrak{X}$ est défini comme son nombre de voisins. Quels sont les valeurs prises par la fonction degré $\deg : \mathfrak{X} \to \mathbb{N}$? Pour chaque valeur prise n, décrire la partie $\mathfrak{X}_n = \{(x, y) \in \mathfrak{X} \mid \deg(x, y) = n\}$.
2. Soit $P((x, y), (u, v))$ la matrice de transition de la marche aléatoire sur la grille $\mathfrak{X}$:

$$P((x, y), (u, v)) = \begin{cases} \frac{1}{\deg(x,y)} & \text{si } (u, v) \text{ est voisin de } (x, y), \\ 0 & \text{sinon.} \end{cases}$$

Écrire un programme `transition(x, y)` qui prend en argument un site $(x, y) \in \mathfrak{X}$, et qui renvoie un voisin de (x, y) tiré suivant la mesure de probabilité $P((x, y), \bullet)$; le voisin (u, v) est donc choisi uniformément parmi les voisins de (x, y). Pour gérer efficacement les problèmes de bord mis en évidence par la question 1, on pourra remarquer que

$$P((x, y), (u, v)) = P((-x, y), (-u, v)) = P((x, -y), (u, -v)) = P((-x, -y), (-u, -v)) ;$$
$$P((x, y), (u, v)) = P((y, x), (v, u)) ,$$

ce qui permet de se ramener au cas où $x \geq 0, y \geq 0$, et ce qui ramène la disjonction de cas à trois situations. Exécuter 10 fois ce programme pour les sites suivants : $(0, 0)$, $(0, 5)$, $(5, 5)$.
3. On rappelle que la *chaîne de Markov* de matrice P est la suite de variables aléatoires $(X_0, X_1, X_2, \ldots)$ telle que, conditionnellement au début de la trajectoire $(X_0, X_1, \ldots, X_n)$, X_{n+1} est choisi suivant la mesure de probabilité $P(X_n, \bullet)$:

$$\mathbb{P}[X_{n+1} = (x_{n+1}, y_{n+1}) \mid X_0 = (x_0, y_0), \ldots, X_n = (x_n, y_n)] = P((x_n, y_n), (x_{n+1}, y_{n+1})) .$$

En utilisant le programme de la question précédente, écrire un programme `marche(x, y, n)` qui part de $X_0 = (x, y)$, et qui renvoie les n premiers pas $X_1, X_2, \ldots, X_n$ de la chaîne de Markov de matrice de transition P. Exécuter ce programme avec $(x, y) = (0, 0)$ et $n = 20$.
4. La fonction `plt.plot(X, Y)` relie les points dont les abscisses sont dans la liste X, et dont les ordonnées sont dans la liste Y. En utilisant cette fonction, écrire un programme `dessiner_marche(x, y, n)` qui dessine dans le plan la marche aléatoire de la question précédente. Exécuter ce programme avec les mêmes paramètres que précédemment.
5. On note $\partial\mathfrak{X}$ le bord de la grille, c'est-à-dire la partie de $\mathfrak{X}$ formée par les points (x, y) avec $x \in \{-5, 5\}$ ou $y \in \{-5, 5\}$. Si $(X_n)_{n \in \mathbb{N}}$ est la marche aléatoire sur la grille $\mathfrak{X}$, on admet que le *temps d'atteinte du bord*

$$T = \inf(\{n \in \mathbb{N} \mid X_n \in \partial\mathfrak{X}\})$$

est fini avec probabilité 1, quelque soit le point de départ X_0. Écrire des programmes qui permettent de simuler la variable T, et de faire un histogramme des fréquences de cette variable aléatoire (on prendra $X_0 = (0, 0)$). Dessiner un histogramme pour $N = 10000$ essais et les valeurs de T comprises entre 0 et 100.
6. Le théorème ergodique assure qu'avec probabilité 1, les fréquences de visites

$$F_n(x, y) = \frac{1}{n} \sum_{i=1}^{n} 1_{(X_n = (x,y))}$$

convergent avec probabilité 1 vers $\pi(x, y)$, où π est l'unique mesure de probabilité sur $\mathfrak{X}$ telle que $\pi P = \pi$ (on voit π comme un vecteur ligne dont les coordonnées sont indicées par les sites de $\mathfrak{X}$, et P comme une matrice de taille $\mathfrak{X} \times \mathfrak{X}$). Écrire un programme `frequence(x, y, n)` qui calcule $F_n(x, y)$ (avec $X_0 = (0, 0)$). En fonction de (x, y), conjecturer la valeur de $\pi(x, y)$ parmi les quantités

$$\left\{ \frac{2}{440}, \frac{3}{440}, \frac{4}{440} \right\}.$$

On regardera le comportement de $F_n(x, y)$ en fonction de l'appartenance ou non de (x, y) au bord de la grille, pour n grand.