DS5

Correction du devoir surveillé

Exercice 1 (Extrait Problème Ecricome 2016)

1. On a une probabilité de $\frac{x}{x+y}$ de tirer une boule rouge. Donc le programme doit retourner 0 avec une probabilité $\frac{x}{x+y}$. Il faut donc que la condition soit satisfaite avec cette probabilité. On va pour cela prendre comme condition $r \leq \dfrac{x}{x+y}$ qui est satisfaite avec une probabilité de $P(R \leq \frac{x}{x+y}) \underset{\frac{x}{x+y}\in[0,1]}{=} \frac{x}{x+y}$ où $R \hookrightarrow \mathscr{U}([0,1])$.

```python
def tirage(x,y):
    r = rd.random()
    if r <= x/(x+y) :
        res = 0
    else :
        res = 1
    return res
```

2. Si $r = 0$ alors on ajoute une rouge sinon on ajoute une blanche, d'où `x = x+1` ou `y = y+1` ; le nombre de rouges ajoutées est donc la différence entre le nombre de boules rouges x au final et le nombre de boules rouges a au départ, soit `Xn = x-a`. Ainsi on a :

```python
def experience(a,b,n):
    x = a
    y = b
    for k in range(n):
        r = tirage(x,y)
        if r == 0 :
            x = x+1 #on a tiré une boule rouge
        else :
            y = y+1 #on a tiré une boule blanche
    Xn = x-a
    return Xn
```

3. On va stocker (une approximation de) la loi de X_n dans la variable `loi`. Commençons par noter que $X_n(\Omega) = [\![0, n]\!]$, et qu'il nous faut une approximation des probabilités $P(X_n = k)$ pour tout $k = 0, \ldots, n$. L'idée est habituelle : on répète l'expérience un grand nombre de fois (ici `m` fois), et on observe la fréquence de chaque issue, qui donnera une approximation de la probabilité théorique. On la stocke alors dans la k-ème composante `loi[k]` du vecteur `loi`. On renvoie alors ce vecteur.

   Voici comment on va procéder plus concrètement :

   - On initialise la variable `loi` en lui affectant un vecteur avec que des 0 à $n+1$ composantes (numérotées de 0 à n), à l'aide de la commande `loi = np.zeros(n+1)`.
   - On répète l'expérience `m` fois à l'aide d'une boucle `for`. Pour chaque résultat `r = experience(a,b,n)` de l'expérience, on ajoute 1 à la r-ème composante du vecteur `loi`.
     Ainsi, à l'issue de la boucle `for`, le vecteur `loi` contient en k-ème composante, le nombre d'expériences (parmi les `m` effectuées) renvoyant l'issue `k`, c'est-à-dire le nombre d'expériences pour lesquelles k boules rouges ont été ajoutées à l'urne.

- Il reste alors à diviser par `m` le vecteur `loi` pour obtenir la fréquence de chaque issue, et donc une approximation de $P(X_n = k)$.

Voici une possibilité de programme :

```python
def simulation(a,b,n,m):
    loi = np.zeros(n+1) # initialisation
    for k in range(m):
        r = experience(a,b,n) # on effectue l'expérience
        loi[r] = loi[r]+1 # +1 à l'effectif de l'issue r
    loi = loi/m # pour obtenir les fréquences
    return loi
```

4. La distribution des fréquences semble équiprobable, donc on peut conjecturer que $X_n \hookrightarrow \mathscr{U}(\llbracket 0, n \rrbracket)$.

Remarque. Ce résultat était ensuite démontré dans la suite du sujet d'Ecricome.

Exercice 2

1. F est continue et strictement croissante car la fonction arctangente l'est, et on a $\lim\limits_{-\infty} F = 0$ et $\lim\limits_{+\infty} F = 1$. Par le théorème de la bijection, F réalise une bijection de $\mathbb{R}$ dans $]0,1[$.

   Déterminons $F^{-1}\ :]0,1[\to \mathbb{R}$. Soit pour cela $x \in \mathbb{R}$ et $y \in]0,1[$. On résout :

$$\begin{aligned} y = F(x) \Leftrightarrow y = \frac{1}{\pi}\left(\arctan(x) + \frac{\pi}{2}\right) &\Leftrightarrow \pi y - \frac{\pi}{2} = \arctan(x) \\ &\Leftrightarrow x = \tan\left(\pi y - \frac{\pi}{2}\right) \end{aligned}$$

   Ainsi on a $F^{-1} : y \in]0,1[\to \tan\left(\pi y - \frac{\pi}{2}\right)$.

2. (a) En utilisant la méthode d'inversion, on obtient :

```python
def cauchy():
    u = rd.random()
    c = np.tan(np.pi*u-np.pi/4)
    return c
```

   (b) Toujours sur le même principe :

```python
def Cauchy(N):
    C = np.zeros(N)
    for k in range(N):
        C[k] = cauchy()
    return C
```

3. La fonction `moyenne(n)` renvoie la moyenne de `n` réalisations d'une variable suivant une loi de Cauchy, à l'aide de la commande `np.mean(Cauchy(n))`. On observe qu'en exécutant plusieurs fois la commande `moyenne(1000)`, on obtient des valeurs sont très dispersées, alors que `n = 1000` est grand, et qu'on pourrait s'attendre à ce qu'elles soient voisine d'une valeur qui serait l'espérance d'une loi de Cauchy. Cela suggère qu'une variable suivant une loi de Cauchy n'admet pas d'espérance.

Vérifions que c'est effectivement le cas. Une variable X suivant une loi de Cauchy admet une espérance si et seulement si l'intégrale $\int_{-\infty}^{+\infty} |t|f(t)\,\mathrm{d}t$ converge, soit par parité si et seulement si $I = \int_0^{+\infty} \frac{t}{\pi(1+t^2)}\,\mathrm{d}t$ converge. La fonction $t \in \mathbb{R}_+ \mapsto \frac{t}{\pi(1+t^2)}$ étant continue sur $[0,+\infty[$, l'intégrale I est généralisée en $+\infty$. Et on a :

- $\frac{t}{\pi(1+t^2)} \underset{t\to+\infty}{\sim} \frac{1}{\pi t}$.
- $\frac{1}{t} \geq 0$ pour tout $t > 0$.
- $\int_1^{+\infty} \frac{\mathrm{d}t}{t}$ diverge (intégrale de Riemann en $+\infty$ d'exposant $1 \leq 1$).

Par théorème de comparaison, l'intégrale I diverge, et $E(X)$ n'existe pas.

Exercice 3

1. Rappelons que la variable `A` contient la matrice $\begin{pmatrix} 1 & 4 & 2 \\ 3 & 2 & 8 \end{pmatrix}$. La commande `np.min(A,0)` renvoie un vecteur de taille le nombre de colonne, contenant le minimum sur chacune des colonnes de A. De même avec les lignes pour la commande `np.min(A,1)`.

2. (a) `X` contient une matrice à 7 lignes et 10000 colonnes dont chaque coefficient est une réalisation d'une variable suivant une loi $\mathscr{E}(1)$.

   D'après la question précédente, `I` est donc un vecteur de taille 10000 tel que pour tout $1 \leq i \leq 10000$, `Y(i)` contient le minimum de la i-ème colonne, c'est-à-dire de 7 réalisations indépendantes de variables qui suivent une loi $\mathscr{E}(1)$. Il s'agit donc d'une réalisation de I_n.

   On dispose donc de 10000 réalisations de I_n, on trace alors l'histogramme des fréquences (bien noté l'ajout de l'argument `density = 'True'`) associé à ces données statistiques, réparties en 100 classes de mêmes amplitudes.

   On trace ensuite sur le même diagramme la densité de la loi $\mathscr{E}(7)$ sur $[0,1]$.

   (b) On remarque que l'aire de chaque rectangle de l'histogramme correspond à l'air sous la densité. Ainsi par comparaison entre l'histogramme des fréquences et la densité de $\mathscr{E}(7)$, on peut conjecturer que I_7 suit une loi $\mathscr{E}(7)$, et plus généralement que I_n suit la loi $\mathscr{E}(n)$ pour tout $n \geq 1$.
