Devoir Maison n°9 - Une urne (Écricome 2019)

Une urne contient initialement une boule blanche et une boule noire. On effectue une succession de tirages d'une boule dans cette urne en procédant ainsi :

- Si une boule blanche est tirée, elle est remise dans l'urne et on ajoute une boule noire dans l'urne.
- Si une boule noire est tirée, elle est remise dans l'urne et on ajoute une boule blanche dans l'urne.

Pour tout $k \in \mathbb{N}$, on définit la variable aléatoire X_k représentant le nombre de boules blanches présentes dans l'urne après le k-ième tirage, c'est-à-dire juste avant d'effectuer le $(k+1)$-ième tirage. On a en particulier $X_0 = 1$.

Partie A - Analyse

1. Déterminer la loi de X_1. Donner son espérance et sa variance.
Initialement, il y a une boule blanche et une boule noire. Il y a deux possibilités : soit on tire la boule blanche (avec probabilité $\frac{1}{2}$) et dans ce cas il reste 1 boule ensuite. Soit on tire la boule noire (avec probabilité $\frac{1}{2}$) et on ajoute une boule blanche. Ainsi, $X_1(\Omega) = \{1; 2\}$ et :
$$\mathbb{P}(X_1 = 1) = \mathbb{P}(X_1 = 2) = \frac{1}{2}$$
Ainsi, $E(X_1) = \frac{1}{2}(1+2) = \frac{3}{2}$
Par ailleurs, $X_1 - 1$ suit une loi de Bernoulli de paramètre $\frac{1}{2}$ et donc $V(X_1 - 1) = V(X_1) = \frac{1}{2} \times (1 - \frac{1}{2}) = \frac{1}{4}$

2. Justifier soigneusement que la loi de X_2 est donnée par :
$$\mathbb{P}(X_2 = 1) = \frac{1}{6}, \quad \mathbb{P}(X_2 = 2) = \frac{2}{3}, \quad \mathbb{P}(X_2 = 3) = \frac{1}{6}$$
Notons pour tout $k \in \mathbb{N}$, B_k l'événement «on tire une boule blanche au k-ième tirage» et N_k son complémentaire.
$$\mathbb{P}(X_2 = 1) = \mathbb{P}(B_1 \cap B_2) = \mathbb{P}(B_1)\mathbb{P}_{B_1}(B_2) = \frac{1}{2} \times \frac{1}{3} = \frac{1}{6}$$
$$\mathbb{P}(X_2 = 3) = \mathbb{P}(N_1 \cap N_2) = \mathbb{P}(N_1)\mathbb{P}_{N_1}(N_2) = \frac{1}{2} \times \frac{1}{3} = \frac{1}{6}$$
Puisque la somme des probabilités des événements $X_2 = 1, X_2 = 2, X_2 = 3$ est 1 (système complet d'événements) :
$$\mathbb{P}(X_2 = 2) = 1 - \mathbb{P}(X_2 = 1) - \mathbb{P}(X_2 = 3) = 1 - \frac{1}{6} - \frac{1}{6} = \frac{2}{3}$$

3. Préciser, pour tout $k \in \mathbb{N}$, l'univers-image $X_k(\Omega)$ de la variable X_k.
Étant donné qu'on part d'une boule blanche et qu'on en ajoute une ou zéro à chaque étape, au bout de k étapes il y a au plus $k+1$ boules blanches dans l'urne. Ainsi : $X_k(\Omega) = [|1; k+1|]$

4. (a) Soit $k \in \mathbb{N}$, $i \in \mathbb{N}^*$ et $j \in X_k(\Omega)$. Déterminer la valeur de $\mathbb{P}_{(X_k = j)}(X_{k+1} = i)$.
On distinguera différents cas selon les valeurs relatives de i et j
D'après l'énoncé, à chaque étape le nombre de boules peut rester identique ou augmenter d'un. Ainsi, si $i \notin \{j; j+1\}$, $\mathbb{P}_{(X_k=j)}(X_{k+1} = i) = 0$
Étudions maintenant les cas $i = j$ et $i = j+1$:
   - $\mathbb{P}_{(X_k=j)}(X_{k+1} = j) = \mathbb{P}_{(X_k=j)}(B_{k+1}) = \frac{j}{k+2}$ puisqu'après k tirages il y a $k+2$ boules dans l'urne et que par hypothèse j d'entre elles sont blanches
   - $\mathbb{P}_{(X_k=j)}(X_{k+1} = j+1) = \frac{k+2-j}{k+2}$ par le même argument.

   (b) En déduire :
$$\forall k \in \mathbb{N}, \quad \forall i \in \mathbb{N}^*, \quad \mathbb{P}(X_{k+1} = i) = \frac{i}{k+2}\mathbb{P}(X_k = i) + \frac{3+k-i}{k+2}\mathbb{P}(X_k = i-1) \qquad (\star)$$
Soit $k \in \mathbb{N}$, soit $i \in \mathbb{N}^*$. D'après la loi des probabilités totales dans le système complet d'événements associé à X_k,
$$\mathbb{P}(X_{k+1} = i) = \sum_{j=1}^{k+1} \mathbb{P}(X_k = j)\mathbb{P}_{(X_k=j)}(X_{k+1} = i)$$
Or, d'après la question précédente, $\mathbb{P}_{(X_k=j)}(X_{k+1} = i) = 0$ sauf si $j = i$ ou $j = i-1$.On en déduit :
$$\begin{aligned}\mathbb{P}(X_{k+1} = i) &= \mathbb{P}(X_k = i)\mathbb{P}_{(X_k=i)}(X_{k+1} = i) + \mathbb{P}(X_k = i-1)\mathbb{P}_{(X_k=i-1)}(X_{k+1} = i) \\ &= \frac{i}{k+2}\mathbb{P}(X_k = i) + \frac{k+2-(i-1)}{k+2}\mathbb{P}(X_k = i-1) \\ &= \frac{i}{k+2}\mathbb{P}(X_k = i) + \frac{3+k-i}{k+2}\mathbb{P}(X_k = i-1)\end{aligned}$$

5. À l'aide de la formule $(\star)$, déterminer la loi de X_3.
Prenons $k=2$. $X_3(\Omega)=[|1;2;3;4|]$. D'après la formule $(\star)$:
   - $\mathbb{P}(X_3=1)=\frac{1}{4}\mathbb{P}(X_2=1)+\frac{4}{4}\mathbb{P}(X_2=0)=\frac{1}{4}\times\frac{1}{6}+0\boxed{=\frac{1}{24}}$
   - $\mathbb{P}(X_3=2)=\frac{2}{4}\mathbb{P}(X_2=2)+\frac{3}{4}\mathbb{P}(X_2=1)=\frac{2}{4}\times\frac{2}{3}+\frac{3}{4}\times\frac{1}{6}=\frac{8+3}{24}\boxed{=\frac{11}{24}}$
   - $\mathbb{P}(X_3=3)=\frac{3}{4}\mathbb{P}(X_2=3)+\frac{2}{4}\mathbb{P}(X_2=2)=\frac{3}{4}\times\frac{1}{6}+\frac{2}{4}\times\frac{2}{3}\boxed{=\frac{11}{24}}$
   - $\mathbb{P}(X_3=4)=\frac{4}{4}\times\mathbb{P}(X_2=4)+\frac{1}{4}\times\mathbb{P}(X_3=3)=0+\frac{1}{4}\times\frac{1}{6}\boxed{=\frac{1}{24}}$

6. Montrer que, pour tout $k\in\mathbb{N}, \mathbb{P}\left(X_k=1\right)=\dfrac{1}{(k+1)!}$.
Montrons-le par récurrence.
   - Initialisation : pour $k=0$, $\mathbb{P}(X_0=1)=1=\frac{1}{1!}$
   - Hérédité : soit $k\in\mathbb{N}$ tel que $\mathbb{P}(X_k=1)=\frac{1}{(k+1)!}$. Alors, d'après la relation $(\star)$,

$$\mathbb{P}(X_{k+1}=1)=\frac{1}{k+2}\mathbb{P}(X_k=1)+\frac{k+2}{k+2}\mathbb{P}(X_k=0)=\frac{1}{k+2}\times\frac{1}{(k+1)!}+0=\frac{1}{(k+2)!}$$

   La récurrence est donc établie.

7. Déterminer, pour tout $k\in\mathbb{N}$, la valeur de $\mathbb{P}\left(X_k=k+1\right)$.
On a aussi $\mathbb{P}(X_k=k+1)=\frac{1}{(k+1)!}$. Deux manières de le justifier :
   - avec la même récurrence qu'à la question précédente, puisque de même que $\mathbb{P}(X_k=0)=0$, on a également $\mathbb{P}(X_k=k+2)=0$
   - par un argument plus concret, le protocole étant symétrique (entre les boules noires et blanches : les règles sont les mêmes et la situation initiale aussi). Ainsi, la loi de X_k soit être symétrique : il y a la même probabilité d'avoir une seule boule blanche ou une seule boule noire, c'est-à-dire $k+1$ boules blanches.

8. Pour tout $k\in\mathbb{N}$, on pose $a_k=(k+1)!\times\mathbb{P}\left(X_k=2\right)$ et $b_k=a_k+k+2$.
   (a) Exprimer a_{k+1} en fonction de a_k et de k.
   Soit $k\in\mathbb{N}$. D'après $(\star)$,

$$\begin{aligned}a_{k+1}&=(k+2)!\mathbb{P}(X_{k+1}=2)=(k+2)!\left(\frac{2}{k+2}\mathbb{P}(X_k=2)+\frac{k+1}{k+2}\mathbb{P}(X_k=1)\right)\\&=2(k+1)!\mathbb{P}(X_k=2)+(k+1)!\frac{k+1}{(k+1)!}\\&=2a_k+k+1\end{aligned}$$

   (b) Montrer que la suite $\left(b_k\right)_{k\in\mathbb{N}}$ est géométrique.
   Soit $k\in\mathbb{N}$.

$$\begin{aligned}b_{k+1}&=a_{k+1}+k+1+2\\&=2a_k+k+1+k+3\\&=2(a_k+k+2)\\&=2b_k\end{aligned}$$

   Ainsi, (b_k) est géométrique de raison 2
   (c) En déduire alors que : $\forall k\in\mathbb{N},\quad \mathbb{P}\left(X_k=2\right)=\dfrac{2^{k+1}-k-2}{(k+1)!}$.
   Puisque (b_k) est une suite géométrique de raison 2 et de premier terme $b_0=1!\mathbb{P}(X_0=2)+0+2=2$, alors pour tout $k\in\mathbb{N}$, $b_k=2^{k+1}$
   Ainsi, $\mathbb{P}(X_k=2)=\frac{a_k}{(k+1)!}=\frac{b_k-k-2}{(k+1)!}=\frac{2^{k+1}-k-2}{(k+1)!}$

9. (a) À l'aide de la formule $(\star)$, montrer :

$$\forall k\in\mathbb{N},\quad E\left(X_{k+1}\right)=\frac{k+1}{k+2}E\left(X_k\right)+1$$

Soit $k \in \mathbb{N}$. Par définition, $E(X_{k+1}) = \sum_{i=1}^{k+2} i\mathbb{P}(X_{k+1} = i)$. Utilisons $(\star)$:

$$\begin{aligned}
E(X_{k+1}) &= \sum_{i=1}^{k+2} i \times \left(\frac{i}{k+2}\mathbb{P}(X_k = i) + \frac{3+k-i}{k+2}\mathbb{P}(X_k = i-1) \right) \\
&= \frac{1}{k+2}\left(\sum_{i=1}^{k+2} i^2\mathbb{P}(X_k = i) + \sum_{j=0}^{k+1}(j+1)(2+k-j)\mathbb{P}(X_k = j) \right) \text{ en posant } j = i-1 \text{ dans la deuxième somme} \\
&= \frac{1}{k+2}\sum_{i=1}^{k+1}(i^2 + (i+1)(2+k-i))\mathbb{P}(X_k = i) \text{ car } \mathbb{P}(X_k = 0) = \mathbb{P}(X_k = k+2) = 0 \\
&= \frac{1}{k+2}\sum_{i=1}^{k+1}(i^2 - i^2 + 2i + ki + 2 + k - i)\mathbb{P}(X_k = i) \\
&= \frac{1}{k+2}\sum_{i=1}^{k+1}(k+1)i\mathbb{P}(X_k = i) + (k+2)\mathbb{P}(X_k = i) \\
&= \frac{k+1}{k+2}\sum_{i=1}^{k+1} i\mathbb{P}(X_k = i) + \sum_{i=1}^{k+1}\mathbb{P}(X_k = i) \\
&= \frac{k+1}{k+2}E(X_k) + 1 \text{ car } (X_k = 1, \dots, X_k = k+1) \text{ forme un s.c.e}
\end{aligned}$$

(b) Déduire de ce qui précède que : $\forall k \in \mathbb{N}, \quad E(X_k) = \frac{k+2}{2}$.
Montrons par récurrence que pour tout entier k, $E(X_k) = \frac{k+2}{2}$

— Initialisation : pour $k = 0$, on sait que $X_0 = 1$ (v.a. certaine) donc $E(X_0) = 1 = \frac{0+2}{2}$
— Hérédité : soit $k \in \mathbb{N}$ un entier. Supposons $E(X_k) = \frac{k+2}{2}$. Alors,

$$\begin{aligned}
E(X_{k+1}) &= \frac{k+1}{k+2}E(X_k) + 1 \\
&= \frac{k+1}{k+2} \times \frac{k+2}{2} + 1 \text{ par hypothèse de récurrence} \\
&= \frac{k+1}{2} + 1 \\
&= \frac{(k+1)+2}{2}
\end{aligned}$$

Et la récurrence est établie.

(c) Soit Y_k la variable aléatoire représentant le nombre de boules noires présentes dans l'urne après le k-ième tirage. Justifier que X_k et Y_k ont même espérance, puis retrouver le résultat précédent.
C'est un raisonnement que l'on a déjà fait précédemment : le nombre initial de boules noires et blanches est identique et les règles de modification sont symétriques, donc X_k et Y_k suivent la même loi, et en particulier ont la même espérance m. Par ailleurs, pour tout k, $X_k + Y_k = k+2$ (nombre de boules total dans l'urne) donc $E(X_k + Y_k) = m + m = k + 2$. Ainsi,

$$m = \frac{k+2}{2}$$

Partie B - Simulations

Rappel : ceci est un devoir maison. Pensez à recopier les codes et les exécuter dans Python (avec Pyzo, par exemple) sur votre ordinateur pour vérifier (après avoir réfléchi) que vos réponses sont cohérentes !

1. (a) Que renvoie la fonction Python suivante pour un entier k non nul ?

```
import random as rd

def mystere(k):
    n=1
    b=1
```

```
        for i in range(k):
            r = rd.randint(1,b+n)
            if r <= b:
                n=n+1
            else:
                b=b+1
        return b
```

Cette fonction simule notre expérience : on réalise k étapes (paramètre de la fonction). À chaque étape, on tire un entier au hasard parmi les nombres de 1 à $b+n$. Il y a b nombres entre 1 et b qui représentent les boules blanches, et les autres représentent des boules noires. Lorsqu'on tire une couleur, on ajoute une boule de l'autre couleur. La fonction **renvoie finalement le nombre de boules blanches après k étapes.** Ainsi, la fonction mystère renvoie **une valeur simulée de** $X_k \in X_k(\Omega) = [|1; k+1|]$

(b) Écrire une fonction Python loi_empirique d'arguments k et N (deux entiers strictement positifs). Cette fonction doit effectuer N simulations de k tirages successifs dans l'urne, et retourner une matrice ligne L de taille $k+1$ dont le i-ème coefficient contient la fréquence d'apparition de l'événement $(X_k = i)$, pour $i \in [\![1, k+1]\!]$.
On pourra utiliser la fonction `mystere` de la question précédente.
La matrice L contient alors une estimation de la loi de la variable aléatoire X_k.

```
    def loi_empirique(k,N):
        L = [0 for i in range(k+1)]
        for i in range(N):
            b = mystere(k)
            L[b-1] = L[b-1]+(1/N) #attention au décalage d'indices
        return L
```

En effet, on réalise N simulations de la même expérience, donc chacune contribue $1/N$ à la fréquence d'apparition de chaque valeur. Attention : $X_k(\Omega) = [|1; k+1|]$ mais la liste Python est indexée de 0 à k

(c) Recopier et compléter la fonction Python loi_theorique suivante, d'argument strictement positif n qui calcule la matrice de taille $(n, n+1)$ dont le coefficient (k, i) est $\mathbb{P}(X_k = i)$, puis retourne le vecteur :

$$(\mathbb{P}(X_n = 1), P(X_n = 2), \ldots, \mathbb{P}(X_n = n+1))$$

```
    import numpy as np

    def loi_theorique(n):
        M=np.zeros((n+1,n+1))
        M[1,1] = 0.5
        M[1,2] = 0.5
        for k in range(2,n+1):
            M[k,1] = M[k-1;1]/(k+1)
            for i in range(2,k+2):
                M[k,i] = i/(k+1)*M[k-1,i]+(2+k-i)/(k+1)*M[k,i-1]
        return M[n-1,:]
```

Note : la commande Python A [i, :] extrait de la matrice A la (i+1)-ème ligne.
$M[k,1] = \mathbb{P}(X_k = 1) = \frac{1}{(k+1)!} = \frac{1}{k+1} \times \frac{1}{k!} = \frac{1}{k+1} \times M[k-1;1]$
Par ailleurs pour $k \in \mathbb{N}^*, i \in \mathbb{N}^*$, $M[k,i] = \mathbb{P}(X_k = i) = \frac{i}{k+1} M[k-1,i] + \frac{2+k-i}{k+1} M[k-1, i-1]$
Puisqu'on remplit la matrice dans l'ordre croissant de k, la ligne précédente est déjà remplie donc $M[k-1,i]$ et $M[k-1, i-1]$ contiennent les bonnes valeurs.

(d) Une étudiante propose comme loi de X_5 le résultat suivant :

k	1	2	3	4	5	6
$P(X_5 = k)$	0.001368	0.079365	0.419434	0.418999	0.079454	0.00138

A-t-elle utilisé la fonction loi_empirique ou la fonction loi_theorique ?
C'est la fonction `loi_empirique`. Les deux contiennent des valeurs similaires mais on voit ici par exemple que $\mathbb{P}(X_5 = 1) \neq \mathbb{P}(X_5 = 6)$ alors que les deux devraient valoir $\frac{1}{6!}$. Même si les deux valeurs sont proches, le fait qu'elles soient différentes montrent qu'elles ne résultent pas du même calcul. De même pour $\mathbb{P}(X_5 = 5) \neq \mathbb{P}(X_5 = 2)$ et $\mathbb{P}(X_5 = 3) \neq \mathbb{P}(X_5 = 4)$