Devoir à la maison n°13

- Le devoir devra être rédigé sur des copies *doubles*.
- Les copies ne devront comporter ni rature, ni renvoi, ni trace d'effaceur.
- Toute copie ne satisfaisant pas à ces exigences devra être intégralement récrite.

Problème 1

1 **1.a** Comme il n'y a que deux variables X_1, X_2, l'ensemble des valeurs prises par ces variables est de cardinal 1 ou 2. Le cardinal 1 est atteint quand $X_1 = X_2 = 1$ par exemple et le cardinal 2 quand $X_1 = 1$ et $X_2 = 2$ ($\ell \geq 2$).

$$\boxed{U_2(\Omega) = \{1, 2\}}$$

1.b $(U_2 = 1) = (X_1 = X_2) = \bigcup_{k=1}^{\ell} (X_1 = k) \cap (X_2 = k)$. La réunion est disjointe et les variables X_1, X_2 sont indépendantes donc

$$\mathbb{P}(U_2 = 1) = \sum_{k=1}^{\ell} \mathbb{P}(X_1 = k)\mathbb{P}(X_2 = k) = \sum_{k=1}^{\ell} \frac{1}{\ell^2} = \frac{1}{\ell}$$

On en déduit $\mathbb{P}(U_2 = 2) = 1 - \mathbb{P}(U_1 = 1)$.

$$\boxed{\mathbb{P}(U_2 = 1) = \frac{1}{\ell} \text{ et } \mathbb{P}(U_2 = 2) = 1 - \frac{1}{\ell}}$$

1.c L'espérance vaut $\mathbb{P}(U_2 = 1) + 2\mathbb{P}(U_2 = 2)$ et donc

$$\boxed{\mathbb{E}(U_2) = 2 - \frac{1}{\ell}}$$

2 **2.a** On écrit (pour le plaisir) une fonction plus générale prenant en argument n et ℓ. On gère une liste `liste` de booléen, la case numéro i étant un booléen indiquant si la valeur i a été prise par l'une des variables X_k (il faut donc $\ell + 1$ cases numérotés de 0 à ℓ). Il s'agit alors de compter combien de cases valent `True`.

```python
def simulU(n,ell):
    liste=[False]*(ell+1)
    for i in range(n):
        liste[random.randint(1,ell)]=True
    s=0
    for x in liste:
        if x:
            s=s+1
    return s
```

2.b La loi faible des grands nombres dit que :
Si (Y_n) une suite de variables aléatoires deux à deux indépendantes, de même loi et admettant des moments d'ordre 2 et si on note $S_n = \sum_{k=1}^{n} Y_k$, $m = \mathbb{E}(Y_1)$, on a

$$\lim_{n \to +\infty} \mathbb{P}\left(\left|\frac{S_n}{n} - m\right|\right) = 0$$

Ainsi, pour calculer une valeur approchée de l'espérance de Y_1, on fait la moyenne sur un grand nombre d'essais des résultats obtenus.

```python
def espU(n,ell):
    s=0
    for i in range(10000):
        s=s+simulU(ell,n)
    return s/(10000)
```

3 Les X_k sont à valeurs dans un ensemble à ℓ éléments et on choisit n de ces valeurs. Ainsi,

$$\boxed{U_n(\Omega) = [\![1, \min(n, \ell)]\!]}$$

4 Remarquons que $\{X_i \in S\} = \bigsqcup_{s \in S} \{X_i = s\}$ de sorte que

$$\mathbb{P}(X_i \in S) = \sum_{s \in S} \mathbb{P}(X_i = s)$$

Puisque X_i suit une loi uniforme sur $[\![1, \ell]\!]$,

$$\boxed{\mathbb{P}(X_i \in S) = \sum_{s \in S} \frac{1}{\ell} = \frac{|S|}{\ell}}$$

5 Les variables X_i étant indépendantes et de même loi, on a

$$\mathbb{P}(X_1 \neq a, \dots, X_{n-1} \neq a) = \prod_{i=1}^{n-1} \mathbb{P}(X_i \neq a) = \mathbb{P}(X_1 \neq a)^{n-1}$$

Avec la question précédente,

$$\boxed{\mathbb{P}(X_1 \neq a, \dots, X_{n-1} \neq a) = \left(\frac{\ell - 1}{\ell}\right)^{n-1}}$$

6 On utilise la formule des probabilité totales avec le système complet d'événements $\{X_n = a\}_{a \in [\![1, \ell]\!]}$:

$$\begin{aligned} \mathbb{P}(X_1 \neq X_n, \dots, X_{n-1} \neq X_n) &= \sum_{a=1}^{\ell} \mathbb{P}(X_1 \neq X_n, \dots, X_{n-1} \neq X_n, X_n = a) \\ &= \sum_{a=1}^{\ell} \mathbb{P}(X_1 \neq a, \dots, X_{n-1} \neq a, X_n = a) \end{aligned}$$

Comme les variables sont indépendantes

$$\mathbb{P}(X_1 \neq X_n, \dots, X_{n-1} \neq X_n) = \sum_{a=1}^{\ell} \mathbb{P}(X_1 \neq a, \dots, X_{n-1} \neq a)\mathbb{P}(X_n = a)$$

Chaque terme dans la somme vaut (question précédente et définition de X_n) $\left(\frac{\ell - 1}{\ell}\right)^{n-1} \frac{1}{\ell}$. En sommant, on obtient donc

$$\boxed{\mathbb{P}(X_1 \neq X_n, \dots, X_{n-1} \neq X_n) = \left(\frac{\ell - 1}{\ell}\right)^{n-1}}$$

7 On utilise la formule des probabilités totales avec le système complet d'événements $(\{X_1, \dots, X_{n-1}\} = S)_{S \in \mathcal{P}_\ell}$ pour obtenir

$$\begin{aligned} \mathbb{P}(X_1 \neq X_n, \dots, X_{n-1} \neq X_n) &= \sum_{S \in \mathcal{P}_\ell} \mathbb{P}(\{X_1, \dots, X_{n-1}\} = S, X_1 \neq X_n, \dots, X_{n-1} \neq X_n) \\ &= \sum_{S \in \mathcal{P}_\ell} \mathbb{P}(\{X_1, \dots, X_{n-1}\} = S, X_n \notin S) \end{aligned}$$

Par lemme des coalitions, les événements $\{X_1, \dots, X_{n-1}\} = S$ et $X_n \notin S$ sont indépendants et donc

$$\mathbb{P}(X_1 \neq X_n, \dots, X_{n-1} \neq X_n) = \sum_{S \in \mathcal{P}_\ell} \mathbb{P}(\{X_1, \dots, X_{n-1}\} = S)\mathbb{P}(X_n \notin S)$$

La question **4** donne alors

$$\boxed{\mathbb{P}(X_1 \neq X_n, \dots, X_{n-1} \neq X_n) = \sum_{S \in \mathcal{P}_\ell} \mathbb{P}(\{X_1, \dots, X_{n-1}\} = S)\left(\frac{\ell - |S|}{\ell}\right)}$$

8 Remarquons que $U_{n-1} = |Y_{n-1}|$ avec $Y_{n-1} = \{X_1, \dots, X_{n-1}\}$. La variable aléatoire Y_{n-1} est à valeurs dans $\mathcal{P}_\ell$ et, d'après la formule de transfert

$$\mathbb{E}(U_{n-1}) = \sum_{S \in \mathcal{P}_\ell} |S| \mathbb{P}(\{X_1, \dots, X_{n-1}\} = S)$$

La question précédente donne

$$\mathbb{P}(X_1 \neq X_n, \dots, X_{n-1} \neq X_n) = \sum_{S \in \mathcal{P}_\ell} \mathbb{P}(\{X_1, \dots, X_{n-1}\} = S) - \frac{1}{\ell} \sum_{S \in \mathcal{P}_\ell} |S| \mathbb{P}(\{X_1, \dots, X_{n-1}\} = S)$$

La première somme du membre de droite vaut 1 car $(\{X_1, \dots, X_{n-1}\} = S)_{S \in \mathcal{P}_\ell}$ est un système complet et ainsi

$$\boxed{\mathbb{E}(U_{n-1}) = \ell(1 - \mathbb{P}(X_1 \neq X_n, \dots, X_{n-1} \neq X_n))}$$

9 Il suffit de combiner les résultats des questions **5** et **8** pour obtenir

$$\boxed{\mathbb{E}(U_n) = \ell\left(1 - \left(\frac{\ell - 1}{\ell}\right)^n\right)}$$

10 A ℓ fixé, si n est très grand, on est presque sûr de trouver toutes les valeurs de $[\![1, \ell]\!]$ et l'espérance devrait être proche de ℓ. C'est bien le cas car $\frac{\ell - 1}{\ell} \in [0, 1[$ de sorte que $\lim_{n \to +\infty} \left(\frac{\ell - 1}{\ell}\right)^n = 0$ puis

$$\boxed{\lim_{n \to +\infty} \mathbb{E}(U_n) = \ell}$$

11 A n fixé et si ℓ est très grand, il est fort probable qu'on ne tombe jamais deux fois sur la même valeur et l'espérance doit être proche de n.
On rappelle que $(1 + u)^\alpha - 1 \underset{u \to 0}{\sim} \alpha u$ donc

$$1 - \left(\frac{\ell - 1}{\ell}\right)^n = 1 - \left(1 - \frac{1}{\ell}\right)^n \underset{n \to +\infty}{\sim} \frac{n}{\ell}$$

puis

$$\boxed{\lim_{\ell \to +\infty} \mathbb{E}(U_n) = n}$$

12 **12.a** Ici, on considère qu'il y a n individus et on note X_k son jour de naissance (un nombre entre 1 et 365). D_n est le nombre des valeurs prises par les X_k. On en dans le cadre de l'exercice avec $\ell = 365$. Ainsi

$$\boxed{\mathbb{E}(D_n) = 365\left(1 - \left(\frac{364}{365}\right)^n\right)}$$

12.b On a alors

$$\boxed{\lim_{n \to +\infty} \mathbb{E}(D_n) = 365}$$