Réflexe :
Importer les librairies nécessaires au calcul scientifique, à la simulation de l'aléatoire et à la représentation graphique.

Entrée[ ]:
```
1
```

TP Python 16 : Variables aléatoires discrètes

Ce TP est un cours à part entière :

- la première et la deuxième parties consistent en des rappels de cours, ainsi que des exercices de bases qu'il convient de maîtriser à 100%.
- La troisième partie est nouvelle et porte sur la simulation des lois discrètes usuelles : elle doit faire l'objet d'un **travail intensif** car elle est au coeur du programme de probabilités en ECG.
- La quatrième partie traite des diagrammes en bâtons : il est très fréquent de voir ce type de diagramme dans les sujets de concours... A bon entendeur!
- La cinquième partie comporte des exercices d'entraînement qu'il sera bon de revoir pour la deuxième année. Ils sont à faire en autonomie et ne sont pas prioritaires.

Partie I - Rappels : simulation d'une variable aléatoire avec `rd.random()`

Commençons par quelques rappels.

Soit $(\Omega, \mathcal{P}(\Omega))$ un espace probabilisable. On appelle variable aléatoire réelle (ou v.a.r) sur $(\Omega, \mathcal{P}(\Omega))$ toute application

$$X : \Omega \to \mathbb{R}.$$

L'ensemble $X(\Omega)$ s'appelle l'univers image de X. C'est l'ensemble des valeurs prises par X.

Une variable aléatoire est caractérisée par sa loi.

- En pratique, dans le cas d'un univers Ω discret (fini ou infini), la loi d'une variable aléatoire sera la donnée de tous les couples $(x, \mathbb{P}(X = x))$ avec $x \in X(\Omega)$.
- La loi sera la principale information dont on disposera sur une variable aléatoire, l'univers Ω étant le plus souvent implicite et non précisé.

Avant de simuler une variable aléatoire, il est important d'avoir **compris comment simuler un évènement de probabilité p.**

La commande `rd.random()` renvoie un nombre réel aléatoire entre 0 et 1 suivant la loi

Typesetting math: 100%

uniforme continue $U([0;1])$.

Plus précisément, si $p \in [0,1]$, la probabilité que le nombre renvoyé soit inférieure ou égale à p (ou dans un intervalle de longueur p) vaut exactement p.

On identifie et décide de représenter donc un évènement de probabilité p par le fait que le nombre aléatoire renvoyé soit inférieur ou égal à p (ou dans un intervalle de longueur p).

Exercice 1

Le score d'un joueur joueur lors d'un lancer de fléchette est modélisé par une variable aléatoire X telle que $X(\Omega) = \{0, 2, 5, 10\}$ et

$$\mathbb{P}(X = 0) = \frac{1}{5}, \quad \mathbb{P}(X = 2) = \frac{1}{2}, \quad \mathbb{P}(X = 5) = \frac{1}{5}, \quad \text{et} \quad \mathbb{P}(X = 10) = \frac{1}{10}.$$

1. Calculer l'espérance et la variance de X.
2. Découper l'intervalle $[0;1]$ en quatre intervalles où pourrait tomber un nombre aléatoire généré avec `rd.random()` dans le but de simuler X. Écrire alors une fonction `simul_X()` qui renvoie une simulation de la variable £X£.

Entrée[ ]:
```
1
```

3. Que fait le programme suivant?

Entrée[ ]:
```
def sample_X(n):

    S = []

    for k in range(n):
        S.append(simul_X())

    return S

print(np.mean(sample_X(1000)))
```

Un n−échantillon d'une variable aléatoire X est un n−uplet dont chaque composante est une variable aléatoire X_i de même loi que X et qui sont toutes mutuellement indépendantes.

La moyenne empirique des réalisations d'un n−échantillon (que l'on obtient grâce à la commande `np.mean( )`) fournit alors une estimation (valeur approchée) de l'espérance de X (quand celle-ci existe).

Exercice 2

On désigne par n un entier naturel supérieur ou égal à 2. On lance n fois une pièce équilibrée (c'est-à-dire donnant Pile avec la probabilité $1/2$, les lancers étant supposés indépendants.

On note Z la variable aléatoire qui vaut 0 si l'on n'obtient aucun Pile pendant ces n lancers et qui, dans le cas contraire, prend pour valeur le rang du premier pile.

Typesetting math: 100%

1. Compléter la fonction suivante permettant de simuler la variable aléatoire Z.

Entrée[ ]:
```
def simul_Z(n) :

    for i in range(1,n+1):
        if ........ :
            return ...

    return ...
```

2. Réécrire cette fonction à l'aide d'une boucle `while` ;

Entrée[ ]:
```

```

Partie II - Lois usuelles discrètes finies

II.1) Lois uniforme U(⟦m,n⟧)

Soit n, m deux entiers naturels, $m < n$. On dit que X suit une loi uniforme sur $[\![\mathrm{m},\mathrm{n}]\!]$, si $X(\Omega) = [\![\mathrm{m},\mathrm{n}]\!]$, et que

$$\forall k \in [\![\mathrm{m},\mathrm{n}]\!], \quad \mathbb{P}(X = k) = \frac{1}{n - m + 1}.$$

De plus, $\mathbb{E}[X] = \dfrac{n + m}{2}$ et $\mathbb{V}[X] = \dfrac{(n - m + 1)^2 - 1}{12}$.

La commande `rd.randint(m,n+1)` permet de simuler une telle variable.

Exercice 3
On dispose de $n \geq 1$ urnes, numérotées de 1 à n. Pour chaque $k \in [\![1,\mathrm{n}]\!]$, l'urne k est composée de boules numérotées de 1 à k. On choisit une urne au hasard puis on tire une boule dans cette urne et on note X_n la variable aléatoire qui prend la valeur du numéro de la boule piochée.

Écrire une fonction `simul_X(n)` qui simule X_n.

Entrée[ ]:
```

```

II.2) Loi de Bernoulli $\mathcal{B}(p)$

On dit que X suit une loi de Bernoulli de paramètre $p \in [0, 1]$, ce qu'on note $X \hookrightarrow \mathcal{B}(p)$, si $X(\Omega) = \{0, 1\}$, et que

$$\mathbb{P}(X = 1) = p \qquad \text{et} \qquad \mathbb{P}(X = 0) = 1 - p.$$

De plus, $\mathbb{E}[X] = p$ et $\mathbb{V}[X] = p(1 - p)$.

Typesetting math: 100%

Exercice 4
Écrire, à l'aide de la commande `rd.random()` une fonction `Bernoulli(p)` permettant de simuler une variable $X \hookrightarrow \mathcal{B}(p)$.

Entrée[ ]:
```
1
```

II.3) Loi binomiale $\mathcal{B}(n, p)$

On dit que X suit une loi binomiale de paramètres $(n, p) \in \mathbb{N} \times [0, 1]$, ce qu'on note $X \hookrightarrow \mathcal{B}(n, p)$, si $X(\Omega) = [\![0,n]\!]$, et que

$$\forall k \in [\![0,n]\!], \quad \mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

De plus, $\mathbb{E}[X] = np$ et $\mathbb{V}[X] = np(1-p)$.

La commande `rd.binomial(n,p)` permet de simuler une telle variable.

Remarque : On peut alors très rapidement simuler une variable aléatoire suivant une loi de Bernoulli de paramètre p à l'aide de la commande `rd.binomial(1,p)` .

Exercice 5
Écrire, à l'aide de la commande `rd.random()` et d'une boucle `for` , une fonction `Binomiale(n,p)` permettant de simuler une variable $X \hookrightarrow \mathcal{B}(n, p)$.

Entrée[ ]:
```
1
```

Partie III - Lois usuelles discrètes à support infini

III.1) Loi géométrique $\mathcal{G}(p)$

On dit que X suit une loi géométrique de paramètres $p \in]0, 1[$, ce qu'on note $X \hookrightarrow \mathcal{G}(p)$, si $X(\Omega) = \mathbb{N}^*$, et que

$$\forall k \in \mathbb{N}^*, \quad \mathbb{P}(X = k) = (1-p)^{k-1} p.$$

De plus, $\mathbb{E}[X] = \dfrac{1}{p}$ et $\mathbb{V}[X] = \dfrac{1-p}{p^2}$.

Une loi géométrique correspond au **temps d'attente du premier succès** (de probabilité p) lors de répétitions (infinies) **indépendantes** d'une même épreuve de Bernoulli.

La commande `rd.geometric(p)` permet de simuler une telle variable.

Typesetting math: 100%

Exercice 6
Écrire, à l'aide de la commande `rd.random()` et d'une boucle `while`, une fonction `Geometrique(p)` permettant de simuler une variable $X \hookrightarrow \mathcal{G}(p)$.

Entrée[ ]:
```
1
```

III.2) Loi de Poisson $\mathcal{P}(\lambda)$

On dit que X suit une loi de Poisson de paramètres $\lambda > 0$, ce qu'on note $X \hookrightarrow \mathcal{P}(\lambda)$, si $X(\Omega) = \mathbb{N}$, et que

$$\forall k \in \mathbb{N}, \quad \mathbb{P}(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}.$$

De plus, $\mathbb{E}[X] = \lambda$ et $\mathbb{V}[X] = \lambda^2$.
La commande `rd.poisson(m)` permet de simuler une variable $X \hookrightarrow \mathcal{P}(m)$, $m > 0$.

Exercice 7
Chaque jour, le nombre de personnes subissant un test de dépistage dans une certaine pharmacie est modélisé par une variable X suivant une loi de Poisson de paramètre $m = 5$. Chaque test a une certaine probabilité $p = \frac{1}{10}$ d'être positif et les tests sont indépendants les uns des autres. On note N le nombre de tests positifs un jour donné.

1. Écrire une fonction permettant simuler la variable N.
2. Créer un échantillon de taille 1000 de cette variable aléatoire et le comparer, avec un histogramme, à un échantillon de même taille d'une loi de Poisson de paramètre $m \times p = 0.5$. Conjecturer.

   On pourra utiliser la fonction `hist()` du module `pyplot` de la bibliothèque `mathplotib` à l'aide de la commande `plt.hist(....)`.

3. (*Au brouillon*) En commençant par expliciter $\mathbb{P}_{X=n}(N = k)$ (en faisant attention sur le couplage des indices n et k), démontrer le résultat précédemment conjecturé.

Entrée[ ]:
```
1
```

Partie IV - Diagrammes à bâtons

La commande `bar(x, y)` du package `matplotlib.pyplot` permet de représenter un **diagramme à bâtons**.

Plus précisément, en notant les arguments $x = [x_1, \ldots, x_n]$ et $y = [y_1, \ldots, y_n]$ (deux listes de mêmes longueurs) elle permet de positionner un bâton de hauteur y_i en face de la valeur x_i.

Typesetting math: 100%

À vous de jouer
Exécuter le programme ci-dessous. Commenter chacune des lignes du programme à l'aide du symbole '#'.

Entrée[ ]:
```
p=1/2

K = [ k for k in range(1,10) ]
p_th = [ p*(1-p)**(k-1) for k in K ]

plt.clf()
plt.bar(K, p_th )
plt.show()
```

→Cliquez-ici pour la réponse ←

Exercice 8
Représenter le diagramme à bâtons dont les hauteurs des bâtons sont les valeurs théoriques de la loi uniforme $U(\text{⟦1,n⟧})$, $n \in \mathbb{N}^*$.

Entrée[ ]:
```
1
```

Attention
On ne fera pas de confusion avec la commande `hist(U)` (du même package) qui représente l'histogramme des valeurs d'une liste `U` .

- Il est souvent très efficace d'utiliser de tels diagrammes pour **conjecturer sur la loi suivie**.

  Par exemple, si tous les bâtons ont à peu près la même hauteur, il est raisonnable de penser que la loi est uniforme!

- Nous verrons en fin d'année, et vous le reverrez l'année prochaine, les diagrammes à bâtons sont aussi pratiques pour **visualiser(ou conjecturer graphiquement) une convergence en loi**.

  Par exemple, les commandes du programme du précédent "À vous de jouer" permettent d'afficher les bâtons dont les hauteurs sont les valeurs théoriques de la loi géométrique de paramètre $p = \dfrac{1}{2}$ (pour les 10 premiers entiers).

L'idée est alors de **créer un n−échantillon** (avec n grand) et de créer un diagramme à bâtons dont la liste x en **abscisses** est celles des **valeurs obtenues par simulation** et la liste y en **ordonnées** celle des **fréquences de chaque valeur dans l'échantillon**.

Exercice 9

Typesetting math: 100%

Une urne contient des boules numérotées de 1 à n, indiscernables au toucher. On effectue alors n tirages sans remise dans l'urne et on note, pour tout $k \in \text{⟦1,n⟧}$, X_k la variable aléatoire qui prend la valeur de la boule obtenue au k−ième tirage.

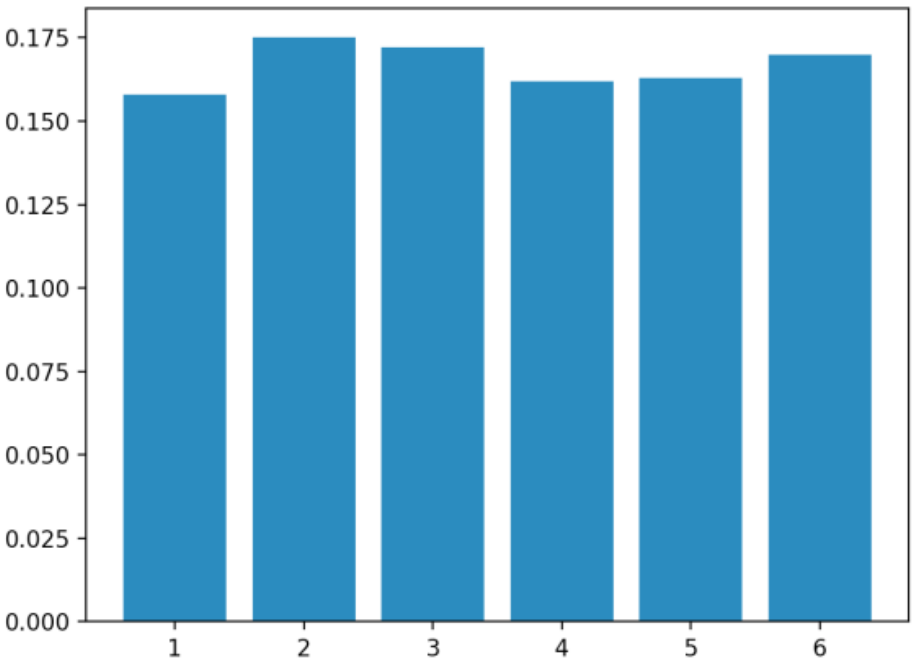
1. Écrire une fonction Python `def simul_X` qui prend en argument deux entiers naturels non nuls k et n, et qui renvoie une simulation de X_k.

Entrée[ ]:
```
1
```

2. On ajoute les commandes suivantes. Expliquer ce qu'elles font. On insistera que la commande de la ligne 7.

Entrée[ ]:
```
1  n=6
2  k=4
3  echantillon = [simul_X(k,n) for j in range(1000) ]
4  freq = [0]*6
5
6  for j in range(1000):
7      freq[echantillon[j]-1] += 1
8
9  freq = [freq[j]/1000 for j in range(n) ]
10
11 plt.clf()
12 plt.bar([j for j in range(1,n+1)],freq)
13 plt.show()
```

3. Leur exécution permet l'affichage suivant:



Émettre une conjecture.

Exercice 10

Typesetting math: 100%

On lance indéfiniment une pièce donnant Pile avec la probabilité p et Face avec la probabilité $q =1-p$. On suppose que $p \in]0,1[$ et on admet que les lancers sont mutuellement indépendants.

Pour tout entier naturel $k \geq 2$, on dit que le k-ième lancer est un **changement** s'il amène un résultat différent de celui du $(k-1)$-ième lancer.

On note P_k (resp. F_k) l'événement : "on obtient Pile (resp. Face) au k_ième lancer".

Pour tout entier naturel $n \geq 2$, on note X_n la variable aléatoire égale au nombre de changements survenus durant les n premiers lancers.

1. Écrire une fonction d'en-tête `simul_X` prenant en argument un entier $n \geq 2$ et un réel $p \in]0,1[$ et qui renvoie une simulation de X_n.
2. Quelles commandes peut-on ajouter pour représenter le diagramme à bâtons des fréquences des valeurs prises par un 1000−échantillon de X_n ?
3. Représenter le diagramme à bâtons susmentionné dans le cas où $p = \dfrac{1}{2}$, avec $n = 3$ puis avec $n = 4$, et enfin avec $n = 5$. Représenter en parallèle des batôns correspondant aux hauteurs théoriques de la loi $\mathcal{B}(n-1, 0.5)$. Émettre une conjecture.

Entrée[ ]:
```
1
```

Partie V - Exercices d'entraînement en autonomie

Exercice A
On considère la variable aléatoire X telle que $X(\Omega)= \left\{-1,0,1 \right\}$, et

$$\mathbb{P}(X=-1) = \mathbb{P}(X=1) = \dfrac{1}{4}, \quad \mathbb{P}(X=0) = \dfrac{1}{2}.$$

Écrire une fonction Python permettant de simuler X.

Entrée[ ]:
```
1
```

Exercice B
Une urne contient 1 boule bleue, 2 boules blanches et 3 boules rouges. On pioche successivement et avec remise 3 boules dans cette urne. Si on obtient les 3 couleurs, on marque 2 points, si on obtient une seule couleur 1 point et 0 point sinon. On note X la variable aléatoire qui prend la valeur du nombre de points marqués sur une partie.

1. Déterminer la loi de X.
2. Écrire une fonction Python qui simule X.
3. Calculer $\mathbb{E}[X]$ puis $\mathbb{V}[X]$.

Entrée[ ]:
```
1
```

Exercice C
Soit N un entier supérieur ou égal à 3. On considère une urne qui contient $(N-1)$ boules blanches et une seule boule noire.

Typesetting math: 100%

On effectue des tirages sans remise jusqu'à l'obtention de la boule noire et on note X la variable aléatoire qui prend pour valeur le nombre de tirages nécessaires.

1. Écrire une fonction `simul_X` prenant en argument un entier naturel $N\geq 3$, et qui renvoie une simulation de la variable X.

Entrée[ ]:
```
1
```

2. Exécuter le programme ci-dessous. Comparer avec le diagramme théorique de l'exercice 8. Commenter, conjecturer, démontrer.

Entrée[ ]:
```
N = 5 # on fera varier les valeurs de N
freq_val_X = np.zeros(N)

for k in range(10000):
    i = simul_X(N)
    freq_val_X[i-1] += 1


freq_val_X = freq_val_X/10000

plt.clf()
plt.bar(range(1,N+1),freq_val_X)
plt.show()
```

Exercice D

On effectue une succession infinie de lancers indépendants d'une pièce donnant Pile avec probabilité p. On note $q = 1 - p$.
On dit que la première série est de longueur $n\geq1$ si les n premiers lancers ont amené le même côté de la pièce et le $(n+1)$-ième l'autre côté. De même la deuxième série commence au lancer suivant la fin de la première série et se termine au lancer précédant un changement de côté.

Par exemple si les lancers donnent les résultats F F P P P P P F F F P P ... alors la première série est de longueur 2 et la deuxième est de longueur 6.

Soient X_1 et X_2 les variables aléatoires égales aux longueurs de la première et deuxième série.

1. Compléter la fonction suivante, prenant en argument la probabilité p d'obtenir Pile et permettant de simuler la variable aléatoire X_1 .

Typesetting math: 100%

Entrée[ ]:

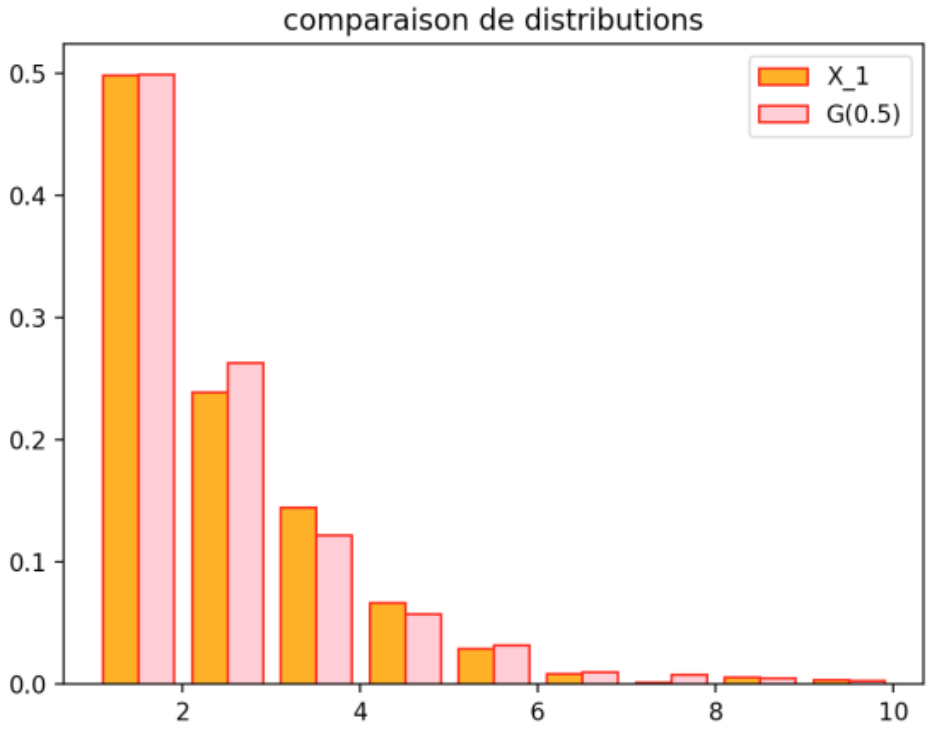
```
def simul_X 1(p) :

    X=1
    t_old = rd.binomial(1,p)
    t_new = rd.binomial(1,p)

    while .... :
        X = ...
        t_old = ...
        t_new = ...

    return X
```

2. Exécuter le code suivant, qui affiche la figure reproduite ci-après. Interpréter la figure obtenue.



Entrée[ ]:

```
p=0.5
N=1000

x1 = [ simul_X1(p) for k in range(N) ]
x2 = [ rd.geometric(p) for k in range(N) ]

plt.clf()
plt.hist([ x1,x2 ], color = ['orange','pink'] , density = True, label = ['
plt.title("comparaison de distributions")
plt.legend()
plt.show()
```

3. Déterminer au brouillon la loi de X_1. Retrouver le résultat conjecturé à la question précédente. Montrer qu'elle admet une espérance que l'on explicitera.
4. Écrire une fonction `simul_X2` prenant en argument un réel $p \in [0,1]$ permettant de simuler la variable X_2 .
5. Déterminer, pour $(k,j) \in \mathbb{N}^{*}\times \mathbb{N}^{*}$, $\mathbb{P}(X_1 = k \cap X_2 = j)$. En déduire, à l'aide de la formule des probabilités totales, la loi de X_2.

Typesetting math: 100%

Entrée[ ]:
```
1
```

Typesetting math: 100%