INTERROGATION DE PYTHON N°8

EXERCICE **10 points**

On considère `x` la série statistique obtenue par simulations de `nbsim` variables de loi binomiale de paramètres $n = 100$ et $p = 0.5$ indépendantes. On suppose avoir importé la bibliothèque random de la façon suivante :

```
import random as rd
```

Et on rappelle que la commande `rd.binomial(n,p)` revoie la simulation d'une variable de loi binomiale de paramètres n, p.

1. (a) Compléter la commande permettant de stocker dans la variable `x` une telle série statistique :

```
x=[rd. ...................... ( .... , .... ) for ................................. ]
```

   (b) Compléter la fonction `mod(x)` qui prend en argument un série statistique et renvoie la liste des modalités de cette série :

```
def mod(x):
  L= .....
  for k in ....... :
    if k not in ........ :
      L.append( .... )
  return L
```

   (c) La commande `55 in mod(x)` renvoie `True`.
   On tape ensuite `np.sum(x==55)` et on trouve 4 760. En déduire la fréquence de la modalité associée.

   ....

   ....

   (d) Rappeler l'expression de $\overline{x}$.

   ....

   Soit $(m_i)_{i\in[\![1;p]\!]}$ les modalités de x et $(f_i)_{i\in[\![1;p]\!]}$ les fréquences associées. Montrer que $\overline{x} = \sum_{i=1}^{p} m_i f_i$.

   ....

   ....

   ....

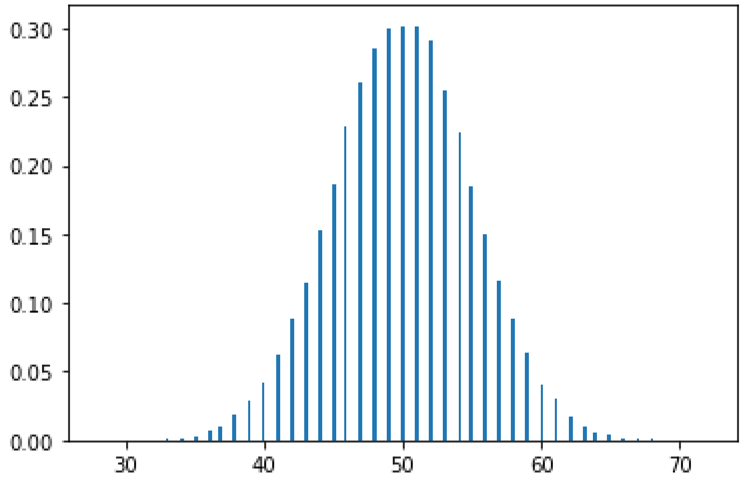
   ....

   (e) On tape la commande `np.mean(x)` et on obtient le résultat `49.98638`. Ce résultat vous parait-il cohérent avec la définition de la série statistique ? Justifier votre réponse.

   ....

   ....

2. On tape la commande suivante : `plt.hist(x,bins='auto',density=True)`
   et on obtient le graphique suivante :



   (a) Rappeler les définitions suivantes :
   Médiane : ....

   ....

   1er quartile : ....

   ....

   3ème quartile : ....

   ....

   On donne :

   `np.mean(x<=50)=0.5398`, `np.mean(x<=46)=0.24433`, `np.mean(x<=53)=0.75817`.
   Que pouvez vous en déduire ?
   ....

   ....

   ....

   (b) On tape la commande `np.std(x)` et on obtient le résultat ≈ **5.03192**. Commentez.
   ....

   ....

INTERROGATION PYTHON N°8 : CORRECTION

EXERCICE **10 points**

On considère x la série statistique obtenue par simulations de 100 000 variables de loi binomiale de paramètres $n = 100$ et $p = 0.5$ indépendantes. On suppose avoir importé la bibliothèque random de la façon suivante :

`import random as rd`

Et on rappelle que la commande `rd.binomial(n,p)` revoie la simulation d'une variable de loi binomiale de paramètres n, p.

1. (a) La commande suivante permet de stocker dans la variable x une telle série statistique :

```
x=[rd.binomial ( 100, 0.5) for_ in range(100000)]
```

   (b) La fonction `mod(x)` prend en argument un série statistique et renvoie la liste des modalités de cette série :

```
def mod(x):
  L= []
  for k in x :
    if k not in L:
      L.append( k )
  return L
```

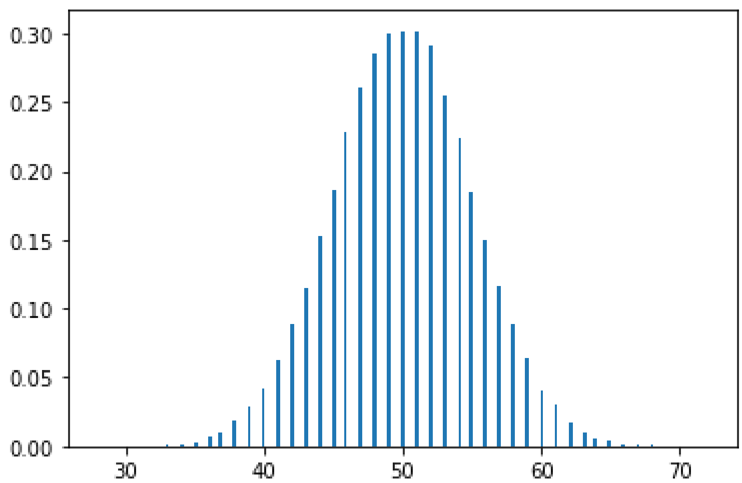
   (c) La commande `55 in mod(x)` renvoie `True`.
   On tape ensuite `np.sum(x==55)` et on trouve 4 760.
   Le fait que `55 in mod(x)` renvoie `True` signifie que $m_i = 55$ est une modalité de x.
   La commande `np.sum(x==55` donne l'effectif associé qui donc de $n_i = 4\,760$. La fréquence associée est donc de $f_i = \dfrac{n_i}{n}$ soit $\boxed{f_i = \dfrac{4\,760}{100\,000} = 0{,}0476}$.

   (d) On note `x=[x_1,....,x_n]`. $\overline{x} = \dfrac{1}{n}\sum_{i=1^n} x_i$.
   Soit $(m_i)_{i\in[\![1;p]\!]}$ les modalités de x et $(f_i)_{i\in[\![1;p]\!]}$ les fréquences associées.
   On a $f_i = \dfrac{n_i}{n}$ où n_i désigne l'effectif de la modalité m_i. Ainsi $m_i f_i = \frac{m_i n_i}{n}$ et $m_i n_i$ revient donc à sommer la modalité m_i autant de fois qu'elle apparait dans la série.
   Il vient donc que $\sum_{i=1}^{p} m_i n_i = \sum_{i=1}^{p} x_i \iff \sum_{i=1}^{p} m_i n_i = n\overline{x}$ et donc que :

$$\frac{1}{n}\sum_{i=1}^{p} m_i n_i = \overline{x} \iff \overline{x} = \sum_{i=1}^{p} \frac{m_i n_i}{n} \iff \boxed{\overline{x} = \sum_{i=1}^{p} m_i f_i.}$$

   (e) On tape la commande `np.mean(x)` et on obtient le résultat `49.98638`.
   Oui le résultat est cohérent puisque la série statistique représente 100 000 simulations d'une variable $X \hookrightarrow \mathscr{B}(100, 0.5)$ qui a pour espérance $n \times p = 100 \times \frac{1}{2} = 50$.

2. On tape la commande suivante : `plt.hist(x,bins='auto',density=True)`
   et on obtient le graphique suivante :



(a) Médiane : la médiane d'une série statistique est un réel m qui sépare la série en deux séries de mêmes effectifs.

1er quartile : le 1er quartile d'une série est le plus grand réel q_1 tel que la série correspondante aux valeurs de x inférieures à q_1 représente 25% de la série x.

3ème quartile : le 3ème quartile d'une série est le plus grand réel q_3 tel que la série correspondante aux valeurs de x inférieures à q_1 représente 75% de la série x.

`np.mean(x<=50)=0.5398`, `np.mean(x<=46)=0.24433`, `np.mean(x<=53)=0.75817`.

$\boxed{m \approx 50}$ $\boxed{q_1 \approx 46}$ $\boxed{q_3 \approx 53.}$

(b) On tape les commandes `np.max(x)-np.min(x)` et `np.std(x)` et on obtient les résultat `44` et `5.03192`.

La commande `np.max(x)-np.min(x)` renvoie l'étendue de la série.

La commande `np.std(x)` renvoie l'écart-type de la série.

Avec une étendue de 44 et un écart-type de ≈ 5 on peut dire que la série n'est pas très dispersée autour de sa moyenne empirique qui vaut ≈ 50 pour une médiane de ≈ 50.

Ce que confirme largement l'histogramme.