

Thème : Probabilité, statistique et informatique

Titre : Hauteurs d'eau de la Seine

Auteur : BRO FRÉDÉRIC

Objectifs :

- ▷ Programmer en PYTHON et utiliser notamment le module PANDAS pour étudier une étude statistique historique.
- ▷ Utiliser les listes.
- ▷ Modéliser le calcul de probabilité d'une crue de la SEINE.
- ▷ Simulations.
- ▷ Calcul de probabilités d'avoir au moins une crue centennale sur n années.

Le ministère de l'écologie et développement durable et de l'énergie met à disposition les mesures de hauteur d'eau de la Seine exprimées en centimètres.

Ces données sont accessibles sur demande, auprès du site <http://www.hydro.eaufrance.fr/>.

Parmi ces données, on a retenu les hauteurs d'eau maximales journalières de la Seine à PARIS pour la période 04/11/1885 - 18/02/2019 et on a obtenu le fichier 'Seine.csv'.

Partie A : Hauteurs d'eau maximales annuelles

1. Ouvrir un notebook et recopier les instructions suivantes permettant de charger les modules nécessaires :

```
In [1]: import pandas as pa
import pylab as pl
T=pa.read_csv('Seine.csv')
T.head()
```

T est une « table de données » (*dataframe*) liée au fichier Seine.csv.

2. Pour obtenir la table des hauteurs d'eau maximales par année recopier l'instruction suivante :

```
In [2]: A=T.groupby(by='annee').max()[['hauteur']]
A.head()
```

3. Pour transformer la colonne hauteur en une liste H, exécuter l'instruction suivante :

```
In [3]: H=list(A['hauteur'])
```

Partie B : Fréquences cumulées croissantes des hauteurs

Objectif : construire la courbe des fréquences cumulées croissantes des hauteurs d'eau.

1. Compléter la fonction nommée **fcc** de paramètres x (*un nombre réel*) et L (*une liste de valeurs*). Elle renvoie la fréquence des valeurs contenues dans L et inférieures à x .

```
In [4]: def fcc(x,L):
c=0
for a in L:
if .....:
c.....
return c/len(L)
```

2. Compléter les pointillés ci-dessous pour représenter le diagramme des fréquences cumulées croissantes des hauteurs d'eau.

```
In [5]: H.sort()
        pl.plot(H, [..... for x in H])
```

Avec Python

- `H.sort()` remplace la liste H par la liste de ses valeurs rangées par ordre croissant.
- `pl.plot(X,Y)` permet de représenter la liste Y en fonction de la liste X.

3. a. Lire graphiquement la valeur médiane de H.
 b. Compléter : pour la période 1885 - 2019, % des années ont vu une hauteur d'eau inférieure à ».
4. a. Lire graphiquement le troisième quartile de H.
 b. Compléter : pour la période 1885 - 2019, % des années ont vu une hauteur d'eau supérieure à ».
5. D'après cette courbe, quelle est la probabilité qu'une année corresponde à une hauteur d'eau supérieure à 600 cm ?

Partie C : Modélisation du calcul de la probabilité d'une crue

Objectif : ajuster au mieux notre courbe des FCC par celle d'une fonction de GUMBEL d'expression suivante :

$$f(x) = \exp\left(-\exp\left(-\frac{(x-a)}{b}\right)\right),$$

avec *a* et *b* deux réels positifs.

1. a. Charger la fonction `exp` du module **math**.
 b. Écrire la fonction nommée GUMBEL de paramètres *x*, *a* et *b* (3 réels positifs avec *b* ≠ 0), qui renvoie *f(x)*.
2. On considère la fonction nommée **courbe** de paramètres *a* et *b*.
 Elle représente :
 — en bleu le diagramme des fréquences cumulées croissantes des hauteurs d'eau
 — en rouge la courbe de la fonction de GUMBEL pour ces deux paramètres.

Compléter les instructions en pointillé.

```
In [8]: def courbe(a,b):
        pl.grid()
        pl.plot(H,.....)
        X=range(100,900)
        Y=[..... for x in X]
        pl.plot(X,Y,color='red')
```

3. Pour modifier les valeurs des paramètres *a* et *b*, on utilise les fonctions interactive et FloatSlider du module **ipywidgets**.
 a. Recopier et exécuter le code suivant :

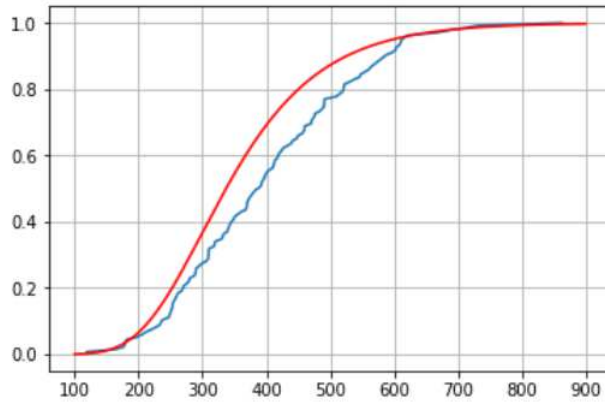
```
In [9]: from ipywidgets import interactive, FloatSlider

        interactive(courbe,
                    a=FloatSlider(value=300,min=300,max=400,step=1),
                    b=FloatSlider(value=100,min=100,max=150,step=1)
                    )
```

Out[9]:

a 300.00

b 100.00



- b. Proposer les valeurs de a et de b qui permettent d'ajuster au mieux notre courbe des FCC par celle de GUMBEL.
4. Pour une année quelconque, si la probabilité de survenue d'une crue est de $\frac{1}{100}$, alors cette crue est dite **crue centennale**.
- On suppose que** a = 338 et b = 111.
- a. Écrire un algorithme qui calcule la plus petite valeur entière de h telle que $f(h) \geq 0,99$.
 - b. Selon le modèle de GUMBEL, donner la hauteur d'eau atteinte lors d'une **crue centennale**.

Partie D : Probabilité d'au moins une crue centennale sur n années

1. Simulations :

- a. Charger la fonction random du module **random**.
- b. Compléter la fonction nommée **crue_centennale** de paramètre n (le nombre d'années consécutives) .

Elle simule, sur n années consécutives, le nombre de crues centennales observées et renvoie 1 si au moins une crue centennale est survenue et 0 sinon.

```
In [15]: def crue_centennale(n):
        C=0
        for i in range(n):
            if .....:
                C=C+1
        if C.....:
            return 1
        else:
            return 0
```

- c. Écrire une fonction nommée **freq** paramètre n (*le nombre d'années consécutives*). Elle simule 1 000 séries de n années consécutives et renvoie la fréquence d'observation d'au moins une crue centennale sur ces n années.
- d. Représenter graphiquement la valeur de cette fréquence pour chaque valeur entière de n comprise entre 2 et 100.

2. Théorie :

On note C_n l'évènement : « il survient au moins une crue centennale sur les n années consécutives » et p_n sa probabilité.

- a. On s'intéresse au cas de $n = 3$ années consécutives. Quelle que soit l'année, on note :

- C est l'issue « il survient une crue centennale »
- \overline{C} est l'issue contraire.

Représenter un arbre à issues associé à l'expérience aléatoire.

- b. Traduire l'évènement contraire de C_3 , noté \overline{C}_3 puis calculer sa probabilité.
- c. En déduire p_3 .
- d. Par un raisonnement analogue, calculer p_n en fonction de n .
- e. Représenter dans un même graphique, les fréquences simulées (*comme à la question 1.c.*) et les probabilités p_n .

Les simulations obtenues sont-elles en accord avec la théorie ?

3. Dernière crue centennale :

- a. Saisir l'instruction suivante :

```
In [20]: A.query('hauteur>=849')
```

En quelle année se produisit la dernière crue centennale ?

- b. Était-ce certain d'observer une crue centennale sur les 100 années qui ont suivie ? Justifier.