

Thème : Probabilité, statistique et informatique

Titre : Éruptions volcaniques

Auteur : BRO FRÉDÉRIC

Objectifs :

- ▷ Programmer en PYTHON et utiliser notamment le module PANDAS pour étudier une étude statistique.
- ▷ Cartographier le jeu de données sur une carte « dynamique » via OPENSTREETMAP.
- ▷ Modéliser avec une loi exponentielle, binomiale.

Durant les 30 derniers jours, on peut obtenir le relevé des différents séismes dans le monde.

Pour chacun d'eux est renseigné notamment :

- sa latitude et longitude (exprimées en degré)
- le temps (exprimé en UTC : COORDINATED UNIVERSAL TIME)

Exemple : 2018-02-21T17:59:34 .564Z



21 février 2018 à 17h59m34s

- sa magnitude (exprimée selon l'échelle de RICHTER)
- sa profondeur (exprimée en km)

Les données sont disponibles directement depuis le lien :

http://earthquake.usgs.gov/earthquakes/feed/v1.0/summary/all_month.csv

Objectifs :

- Analyser ce jeu de données.
 - Modéliser le temps d'attente entre deux séismes de magnitude supérieure à 5.
 - Calculer la proportion de séismes qui se sont produits ces 30 derniers jours, dans les pays de l'Asie du sud-est (INDONÉSIE, SINGAPORE, *etc*).
- Cette zone appelée aussi insulinde correspond à une latitude allant de -13° à 15° et une longitude allant de 90° à 170° . Ce lieu est le carrefour de plusieurs plaques géologiques et est un lieu réputé sensible.
- Calculer la probabilité d'avoir au moins un séisme sur une période.

Partie A : Extraction des données

1. Saisir dans le SHELLS les instructions suivantes :

```
In [1]: import pandas as pa
```

```
In [2]: T=pa.read_csv("http://earthquake.usgs.gov/earthquakes/
feed/v1.0/summary/all_month.csv", sep=',')
```

2. Afficher l'en-tête du tableau T.

3. Combien de séismes se sont produits au cours de ces 30 derniers jours ?

Partie B : Étude du temps d'attente entre 2 séismes

1. **Conversion d'un temps en format UTC vers datetime**

L'objectif est d'évaluer le nombre de secondes qui séparent deux séismes consécutifs.

Pour soustraire deux temps, on va utiliser la fonction **datetime** du module DATETIME.

La fonction **datetime** prend 6 paramètres (entiers) : l'année - le mois - le jour - l'heure - les minutes - les secondes.

Exemple : le 14 mars 2018 à 15h30 est associé à

```
In [5]: from datetime import datetime
         datetime(2018,3,14,15,30,0)
```

```
Out[5]: datetime.datetime(2018, 3, 14, 15, 30)
```

- a. On définit la fonction **conversion** qui prend pour argument s (une chaîne de caractère correspondant au temps écrit au format UTC) et qui renvoie la liste des entiers (années - mois - jour - l'heure - les minutes - les secondes).

```
In [6]: def conversion(s):
        k=0
        Mot=' '
        while s[k]!='.':
            if s[k]!='-' and s[k]!='T' ..... s[k]!=':':
                Mot=Mot+.....
            else:
                Mot=Mot+.....
            k=.....
        Liste_string=Mot.split('-')
        return [int(a) for a in Liste_string]
```

Principe :

- On crée le mot vide Mot.
- On parcourt chaque lettre du mot s tant que celle-ci est différente de '.'
 - o si la lettre lue est différente de '-' ou 'T' ou ':' alors on l'ajoute au mot Mot
 - o sinon on ajoute '-' au mot Mot.

Compléter les pointillés.

- b. Compléter les instructions en pointillés pour obtenir le temps où s'est produit le dernier séisme exprimé avec **datetime** :

```
In [7]: conversion(T.loc[0,'time'])
```

```
Out[7]: .....
```

```
In [8]: t=conversion(T.loc[...,'time'])
         datetime(t[0],t[1],t[2],,,,,,.....)
```

```
Out[7]: .....
```

- c. On définit la fonction **horaire** qui prendra pour argument UTC (une chaîne de caractère correspondant au temps écrit au format UTC situé dans la colonne **timing** du tableau). Elle renvoie ce temps au format datetime.

```
In [9]: def horaire(UTC):
        t=conversion(UTC)
        return datetime(t[0],,,,,,.....)
```

Compléter les pointillés.

2. Création du tableau T1 ne concernant que les séismes d'amplitude >5

- a. i. Compléter la condition à écrire pour créer le tableau **T1** formé des lignes de **T** concernant seulement les séismes de magnitude >5 :

In [10]: T1=T[.....]

- ii. Afficher l'en-tête de **T1**.

- b. En utilisant la fonction **describe** au tableau **T1**, compléter le tableau ci-dessous :

Séismes de magnitude >5	\bar{x}	Q_1	m_e	Q_3
Magnitude				
Profondeur				

3. Création de la liste des temps entre deux séismes consécutifs

- a. On définit la fonction **attentes** qui prendra pour argument Tab (un tableau de données contenant la colonne **time** formées des temps en format UTC). Cette fonction renvoie la liste des temps d'attente , **exprimés en seconde**, entre deux séismes consécutifs.

```
In [13]: def attentes(Tab):
    L=[]
    Liste_Temps=list(Tab['time'])
    for k in range(1,len(Liste_Temps)):
        d1=horaire(Liste_Temps[k-1])
        d2=horaire(.....)
        d=.....
        L.append(d.seconds)
    return L
```

Compléter les pointillés.

- b. Créer la liste L formée des temps d'attente entre deux séismes consécutifs contenus dans **T1**.

4. Modélisation

On note *S* la variable aléatoire qui, à chaque séisme pris au hasard, associe le temps d'attente (*exprimé en seconde*) du prochain séisme.

Les valeurs de la liste L forment un échantillon de plusieurs réalisations de *S*.

- a. **Proposition « vraisemblable » d'une loi de probabilité pour S :**

- i. Représenter l'histogramme normalisé des fréquences associé à L.
Indication : utiliser la commande pl.hist du module PYLAB pour représenter cet histogramme.

- ii. Calculer avec PYTHON, le temps moyen d'attente entre deux éruptions.

Indications :

- utiliser la commande np.mean du module NUMPY pour calculer ce temps moyen d'attente.
- On stockera ce temps moyen dans une variable nommée m.

- iii. Quelle célèbre loi, permet d'affirmer que $E[S] \approx m$?

- iv. En observant l'histogramme obtenu, ajuster à cet histogramme une loi à densité de référence pour *S*.

Avec ce qui précède, proposer son ou ses éventuels paramètres.

b. Les statisticiens affirment que le modèle choisi est acceptable.

Durant 30 jours étalés sur Janvier-Février 2018, en moyenne 20 728 secondes ont séparé 2 séismes consécutifs (*de magnitude >5*) dans le monde.

- i. Convertir ce temps moyen en heures.
- ii. Calculer la probabilité qu'il ne se produise aucun nouveau séisme (*de magnitude >5*) pendant au moins un jour.

Partie C : Carte des séismes

Pour représenter les séismes (de magnitude >5) sur une carte « dynamique », recopier les instructions suivantes :

In [25]: `import folium`

In [26]: `carte = folium.Map(location=[0,0],tiles='Mapbox Control Room',
zoom_start=2)`

In [27]: `for lat,long,mag in zip(T1['latitude'],T1['longitude'],T1['mag']):
 folium.CircleMarker(location=[lat,long],
 fill=True,
 fill_color='red',
 fill_opacity=0.5,
 color='yellow',
 popup=str(mag),
 radius=np.sqrt(mag)).add_to(carte)`

Partie D : Activité sismique en Asie du Sud - Est

Les pays de l'Asie du sud-est (INDONÉSIE, SINGAPORE, *etc*) appartiennent à la zone (appelée **INSULINDE**) ayant une latitude comprise entre -13° à 15° et une longitude comprise entre 90° et 170° .

1. Avec PANDAS, créer le tableau **T2** formé des lignes du tableau **T1** qui ne concernent que les séismes qui se sont produits en **INSULINDE**.
2. Calculer la proportion des séismes (de magnitude >5) qui se sont produits dans cette zone.
3. Que représente l'aire de cette zone par rapport à celle de la TERRE ?

Partie E : Nombre de séismes sur une période

Durant la même période des 30 jours étalés sur Janvier-Février 2018, il y a eu 108 séismes (*de magnitude >5*) dans le monde.

Un séisme se produit à une date exprimée au format UTC.

1. Calculer la proportion notée p de séismes par seconde durant cette période de 30 jours.
2. On note X le nombre de séismes qui se sont produits durant 1 jour.

On suppose que pour deux séismes quelconques :

- ils sont indépendants entre eux
- ils ne peuvent débiter en même temps.

- a. Déterminer la loi de probabilité de X .
- b. Calculer la probabilité qu'il se produise au moins un séisme en l'espace d'un jour complet.
3. A partir de quelle heure, avons-nous plus d'une chance sur deux d'observer au moins un séisme (*de magnitude >5*) dans le monde ?