

Logiciel scientifique - projet

B. DI PIERRO

Année 2021 - 2022

Contexte

Le but de ce projet est de tester vos capacités à traiter d'une base de donnée en python en utilisant les bibliothèque numpy, pandas et matplotlib. Toute remarque pertinente devra explicitement figurer dans les commentaires du programme et cela fera partie de la notation.

La base de donnée à traitée est générée par une fonction se trouvant dans le fichier "data_etu.py", disponible sur moodle. Ce fichier ne doit pas être modifié ! La base de données contient des données Y_i bruitées en fonction d'une variable 'X', dont une seule contient des données d'intérêt (nous ne savons à priori pas laquelle).

Commencer par charger la bibliothèque de génération de la base de données :

```
import data_etu as de
```

puis générez une base de données :

```
data = de.generate_data(N);
```

où N est votre numéro étudiant (sans la lettre "p"). Il est conseillé d'étudier cette base de donnée AVANT de travailler dessus afin d'en comprendre la structure.

1 Travail à réaliser

1. calculez la moyenne de chacune des colonnes " Y_i " de la base de donnée. Une seule d'entre elle aura une moyenne supérieure à 0.05. Elle sera considéré comme la colonne d'intérêt et sera appelée ' Y_k '.
2. Écrivez une fonction qui renverra la moyenne, l'écart type et la position 'X' correspondant au maximum de ' Y_k '. Cette fonction tracera également ' Y_k ' en fonction de 'X'.
3. Écrivez une fonction qui calculera

$$C = \int_{-1}^1 Y_k(X) dx \quad (1)$$

par la méthode de monté Carlo.

4. On souhaite maintenant normaliser Y_k , c'est à dire tel que :

$$\int_{-1}^1 Y_k(X) dx = 1. \quad (2)$$

En vous servant de la question précédente, ajoutez une colone 'Z' à la base de donnée contenant Y_k normalisée.

5. À partir de toutes les questions précédentes, écrivez une fonction qui tentera de reconstruire une forme non bruitée de Y_k .