

ENSAE TD noté, mardi 24 novembre 2020

Le programme devra être envoyé par mail au chargé de TD et au professeur. Toutes les questions valent 2 points excepté deux questions.

1

On dispose de données issues des hôpitaux relative à l'épidémie. On s'intéresse au temps passé dans l'hôpital. Pour chaque personne, on connaît :

- Sa date d'entrée (jour de l'année)
- Sa date de sortie si la personne est guérie,
- L'issue : décès (0) ou guérison (1)

On veut calculer une fonction $S(t)$ qui détermine la probabilité de survie des malades encore dans l'hôpital t jours après leur arrivée. Cette courbe s'appelle l'estimateur de *Kaplan-Meier*.

1) Récupérer les données disponibles à l'adresse suivante sous la forme d'un DataFrame avec l'instruction :

```
pandas.read_csv("http://www.xavierdupre.fr/enseignement/complements/examen2021.zip")
```

2) Calculer la durée de séjour à l'hôpital pour chaque personne, quelle que soit l'issue.

3) On note $n(t)$ nombre de personnes entrées à l'hôpital et qui y sont encore après t jours, on note $d(t)$ le nombre de personnes qui sont décédées au jour t à partir de leur entrée à l'hôpital. L'estimateur de *Kaplan-Meier* est défini par $S(t) = \frac{n(t)-d(t)}{n(t)}$. Ecrire une fonction qui retourne le résultat de cet estimateur pour un t donné.

4) Calculer tous les $S(t)$ pour t allant de 1 à 100.

5) Tracer cette fonction avec matplotlib. Vous pourrez vous inspirer du code fourni dans l'énoncé (1 point)

```
import matplotlib.pyplot as plt
fig, ax = plt.subplots(1, 1, figsize=(4, 4))
T = [0, 1, 2, 3]
St = [1, 0.8, 0.6, 0.2]
ax.plot(T, St)
```

6) Les données des hôpitaux en accès libre recensent chaque jour le nombre de personnes entrées à l'hôpital, le nombre de personnes décédées et guéries à l'hôpital. Est-il possible de calculer l'estimateur précédent sur ses données? Expliquez pourquoi. (On ne demande pas de le faire si c'est possible). (1 point)

2

On s'intéresse à la propagation d'un virus dans une classe. On suppose que chacun reste assis à la même place sans bouger. Il faut une heure pour qu'une personne contamine son voisin. Une personne a quatre voisins :

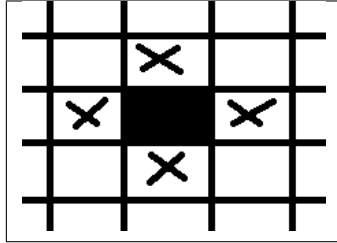


FIGURE 1 : *Voisinage pixels*

1) On considère une matrice carrée $M = \text{numpy.zeros}(N, N)$ avec un élève par place. L'élève ij est contaminé ($M[i, j]=1$) ($i, j = 4, 5$ par exemple). On suppose connu l'état de la contamination à l'instant t (heures), calculer la matrice des personnes contaminées à l'heure $t + 1$ d'après.

Il est possible de répondre à cette en utilisant la fonction `numpy.maximum` et aucune boucle.

2) même chose après T heures.

3) On vaccine une proportion p de la classe, on crée une matrice de la même dimension qui contient 1 si la personne est vaccinée, 0 sinon.

```
vaccine = (numpy.random.rand(N, N) <= p).astype(numpy.int32).
```

Une personne vaccinée stoppe la propagation. Calculer la propagation l'heure $t + 1$.

4) même chose après T heures.

5) Faites varier p , recommencer plusieurs fois. Quel p faut-il pour éviter la propagation ?