

# Apprentissage Statistique - TP 1

M2 EI2D & MATD – Le 22 sept 2025

Instructions : Préparez un rapport incluant le code source et vos résultats, et déposez-le sur Teams. Pas plus de 2 personnes par groupe. N'oubliez pas de mettre les 2 noms sur le rapport, ou de faire 2 rendus. La date limite du rendu TP est le 6 oct 2025

## A Les Données

Dans cet exercice, on va utiliser un jeu de données généré par vous même.

1. Engendrez un tableau d'observations de la forme suivante :

$$\begin{pmatrix} (B, T, G) & (B, T, G) & (B, T, P) \\ (D, T, G) & (D, T, G) & (B, T, P) \\ (D, T, G) & (D, T, G) & (B, T, P) \\ (D, T, G) & (D, T, G) & (D, T, P) \\ (B, T, G) & (B, T, G) & (D, T, P) \\ (B, S, G) & (B, S, G) & (D, S, P) \\ (B, S, G) & (B, S, G) & (B, S, P) \\ (D, S, G) & (D, S, G) & (D, S, P) \end{pmatrix}$$

2. Dans l'exemple ci-dessus vous avez 8 rangés. Ecrivez une fonction Python qui vous permettent d'engendrer d'une manière aléatoire en utilisant une distribution uniforme un tableau de la même forme, mais avec  $n$  rangés et 3 colonnes.
3. On suppose que nous l'ensemble de données ainsi obtenu décrit les trois variables suivantes : Cheveux = B, D, où B=blond, D=dark. Hauteur = T, S, où T=tall, S=short. Pays = G, P, où G=Greenland, P=Poland.

## B Question

Nous voulons répondre à la question suivante : si vous observez un nouvel individu grand ayant cheveux blonds, quel est son pays d'origine le plus probable?

### B.1 Remarques :

1. Il s'agit d'écrire une fonction Python qui engendre une matrice ayant  $N$  lignes, dont chaque élément est constitué d'une triplet du type  $:(X_1, X_2, X_3), X_i \neq X_j, X_i \in \{D, T, G\}$ .

2. L'ensemble des fonctions du TP doivent être écrites par rapport à cette matrice créée par vous-même. Ne donnez pas des réponses par rapport à la matrice exemple de l'énoncé.

## C Maximum a posteriori

1. Ecrivez une fonction Python qui donnera la solution maximum a posteriori (MAP) à la question, en utilisant le cadre naïf de Bayes.
2. Ecrivez une fonction Python qui donnera la solution maximum a posteriori (MAP) à la question, sans utiliser le cadre naïf de Bayes.

### C.1 Remarques :

1. Dans ce cadre vous devons comparer  $\mathbb{P}(G|B, T)$  avec  $\mathbb{P}(P|B, T)$ .
2. Sous l'hypothèse de Bayes naïf, on a

$$\mathbb{P}(G|B, T) = \frac{\mathbb{P}(B, T|G)\mathbb{P}(G)}{\mathbb{P}(B, T)}$$

et un calcul analogue pour  $\mathbb{P}(P|B, T)$ .

3. En conclusion on doit estimer  $\mathbb{P}(B, T|G)$ ,  $\mathbb{P}(G)$ ,  $\mathbb{P}(B, T)$  et  $\mathbb{P}(P)$
4. A titre d'exemple  $\mathbb{P}(T|G) = \frac{5}{8}$  et  $\mathbb{P}(G) = \frac{2}{3}$ .
5. Sans l'hypothèse de Bayes naïf nous devons estimer  $\mathbb{P}(B, T|G)$  et  $\mathbb{P}(B, T|P)$  à partir de l'ensemble d'apprentissage fourni.

## D Maximum de vraisemblance

1. Ecrivez une fonction Python qui donnera la solution de maximum de vraisemblance (MLE) à la question, en utilisant le cadre naïf de Bayes.
2. Ecrivez une fonction Python qui donnera la solution de maximum de vraisemblance (MLE) à la question, sans utiliser le cadre naïf de Bayes.

### D.1 Remarques :

1. La réponse de MLE repose uniquement sur la comparaison des termes de vraisemblance. Par conséquent, nous devons comparer  $\mathbb{P}(B, T|G)$  avec  $\mathbb{P}(B, T|P)$ .
2. A titre d'exemple sous l'hypothèse de Bayes naïf (NB), dans le cadre de la matrice exemple on obtient :

$$\mathbb{P}(B, T|G) = \mathbb{P}(B|G)\mathbb{P}(T|G) = \frac{8}{16} \cdot \frac{10}{16} = \frac{5}{16}$$

3. A titre d'exemple sans l'hypothèse de Bayes naïf (NB), nous devons comparer  $\mathbb{P}(B, T|G)$  avec  $\mathbb{P}(B, T|P)$  et l'on obtient  $\mathbb{P}(B, T|G) = \frac{2}{8}$ .

## E Point de vue continu

Dans cette partie on suppose que la mesure estimée de la couleur des cheveux est continue donnée par une valeur appelée intensité variant entre  $(0, 1)$  représentant l'intensité de la couleur de cheveux allant de  $B$  à  $D$ , où B=blond, D=dark.

1. Engendrez un tel jeu de données via une fonction Python.
2. Expliquer comment vous résoudriez la même question dans ce contexte.
3. Quelles sont les modifications à faire dans les fonctions Python précédentes afin de répondre à la même questions dans ce contexte.