

# Un modèle pour les nids d'oiseaux

CARVAILLO, CÔME, PRALON

Soutenance de projet de Master 1

3 juin 2022



# Introduction

## Introduction

Je bite dans vos culs, je bite dans vos bouches!!!



# Préliminaires mathématiques

## Préliminaires mathématiques

À remplir tout comme j'ai rempli la chatte de vos mères avec ma semence !!!



# L'algorithme EM

## Pseudo code de l'algorithme EM

---

**Algorithm 1** L'algorithme EM (Dempster et al., 1977).

---

**Entrée(s)** :  $\tilde{\theta}_0 \in \Theta$ , un jeu de données  $X_1 \cdots X_n$ ,  $K \in \mathbb{N}$ ;

1: **pour**  $k$  allant de 1 à  $K$  **faire**

2:   **ETAPE E** : Calculer la probabilité  $\mathbb{P}_{\tilde{\theta}_{k-1}}(Z = j | X = X_i) = \frac{\alpha_j \times \gamma_{\mu_j, j_v}}{\sum_{k=1}^J \alpha_k \times \gamma_{\mu_k, v_k}}$ ,  $\forall i \in \llbracket 1, n \rrbracket$

3:   **ETAPE M** : Calculer  $\tilde{\theta}_k = \underset{\theta = (\alpha_j, \mu_j, v_j)_{j \in \llbracket 1, J \rrbracket}}{\operatorname{argmax}} \mathbb{P}_{\tilde{\theta}_{k-1}}(Z = j | X = X_i)$ ;

4: **fin du pour**

5: **retourner**  $\tilde{\theta}_K$ ;

---



# Les étapes de l'algorithme EM

## L'étape E (Expectation)

Consiste à déterminer  $\mathbb{P}_{\tilde{\theta}}(Z = j | X = X_i)$  à l'aide de la formule suivante :

$$\mathbb{P}_{\tilde{\theta}}(Z = j | X = X_i) = \frac{\alpha_j \times \gamma_{\mu_j, v_j}}{\sum_{k=1}^J \alpha_k \times \gamma_{\mu_k, v_k}}$$



# Les étapes de l'algorithme EM

## L'étape M (Maximization)

Consiste à déterminer les EMV  $(\widehat{\alpha}_j, \widehat{\mu}_j, \widehat{\sigma}_j)$  de la log-vraisemblance conditionnelle via les formules suivantes :

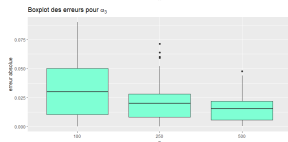
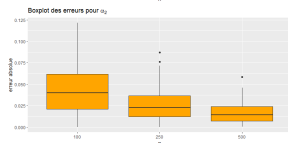
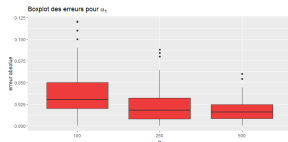
$$\widehat{\alpha}_j = \frac{1}{n} \sum_{i=1}^n \mathbb{P}_{\tilde{\theta}}(Z = j | X = X_i)$$

$$\widehat{\mu}_j = \frac{\sum_{i=1}^n X_i \mathbb{P}_{\tilde{\theta}}(Z = j | X = X_i)}{\sum_{i=1}^n \mathbb{P}_{\tilde{\theta}}(Z = j | X = X_i)}$$

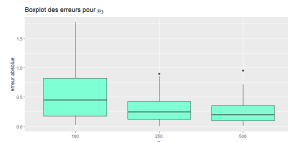
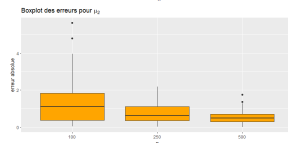
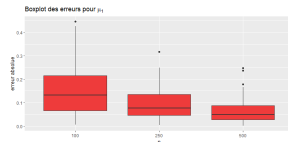
$$\widehat{\nu}_j = \frac{\sum_{i=1}^n (X_i - \widehat{\mu}_j)^2 \mathbb{P}_{\tilde{\theta}}(Z = j | X = X_i)}{\sum_{i=1}^n \mathbb{P}_{\tilde{\theta}}(Z = j | X = X_i)}$$



# Cas des variables à "fortes séparations"

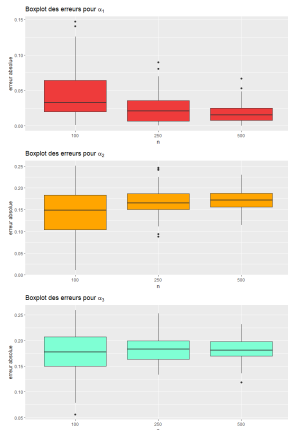


(a) Boxplot des erreurs  
pour  $\alpha_1$ ,  $\alpha_2$  et  $\alpha_3$

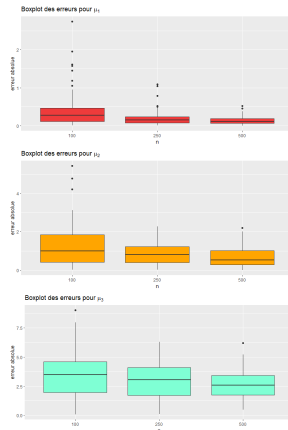


(b) Boxplot des erreurs  
pour  $\mu_1$ ,  $\mu_2$  et  $\mu_3$

# Cas des variables à "faibles séparations"



(c) Boxplot des erreurs  
pour  $\alpha_1$ ,  $\alpha_2$  et  $\alpha_3$



(d) Boxplot des erreurs  
pour  $\mu_1$ ,  $\mu_2$  et  $\mu_3$



# Modélisation des nids d'oiseaux

## Première exploration des données

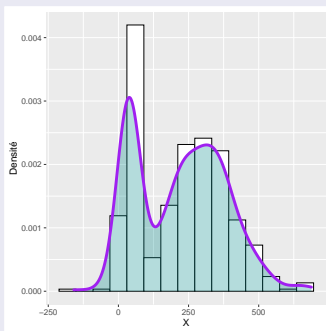


Figure – Densité du mélange

# Heuristique graphique

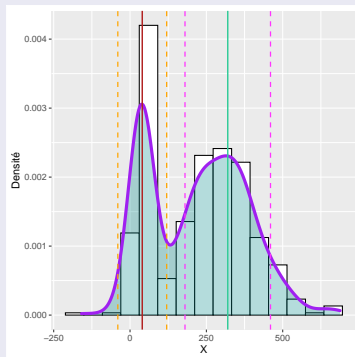


Figure – Densité du mélange

# Conclusion

## Conclusion

J'encule vos grosses marraines bien profond !!!

