Modèles de mélange de densités

Algorithme EM pour un mélange de deux gaussienne

Véronique Tremblay

Voir la section 8.5.1 du livre Elements of Statistical Learning (Hastie, Tibshirani, and Friedman (2009)).

Algorithme EM

Solutions aux défis de la $\ell\ell$ du modèle de mélanges :

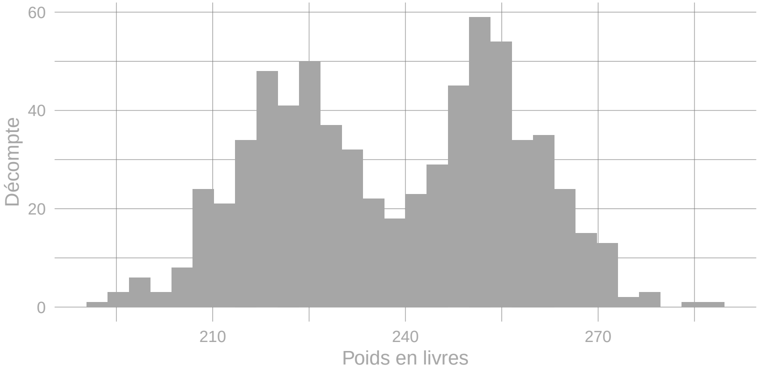
$$\ell(\theta; Y) = \sum_{j=1}^{n} \log \left[\sum_{i=1}^{K} \pi_i f_i(Y_j, \theta_i) \right]$$

Algorithme EM

- Approche générique d'estimation de modèles paramétriques lorsque la vraisemblance est difficile à manipuler (Dempster, Laird, and Rubin (1977))
- Tire son nom des deux étapes:
  - E = expectation
  - M = maximization

Exemple pour 2 gaussiennes

$$\ell(\theta|Y) = \sum_{j=1}^{n} \log\left[\pi_1 \phi_{\theta_1}(Y_j) + \pi_2 \phi_{\theta_2}(Y_j)\right]$$

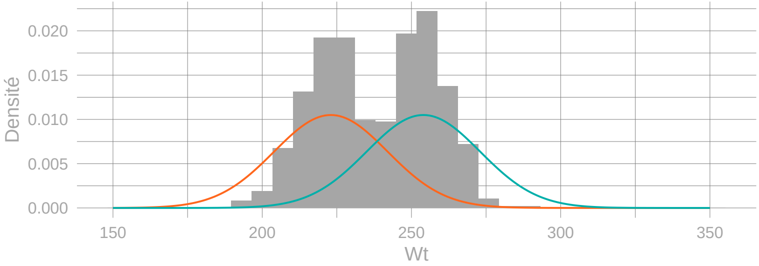


Étape 1: initialisation

$\mu_1^{(0)} = 223$ $\mu_2^{(0)} = 254$

$\sigma_1^2 = \sigma_2^2 = 19$

$\pi^{(0)} = 0.5$

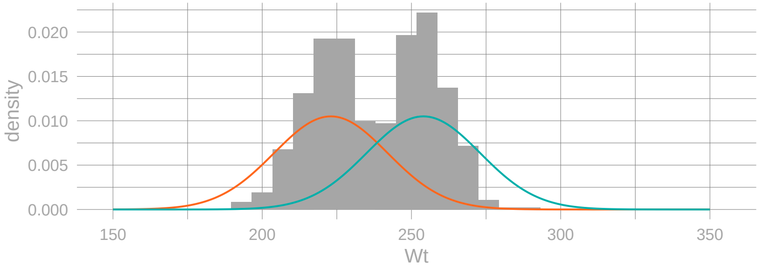


Distributions initiales

Itération 1

Étape 2 : Expectation

$$\hat{\gamma}_i^{(0)} = \frac{\pi_2^{(0)} \phi_2^{(0)}(y_i)}{\pi_1^{(0)} \phi_1^{(0)}(y_i) + \pi_2^{(0)} \phi_2^{(0)}(y_i)}$$



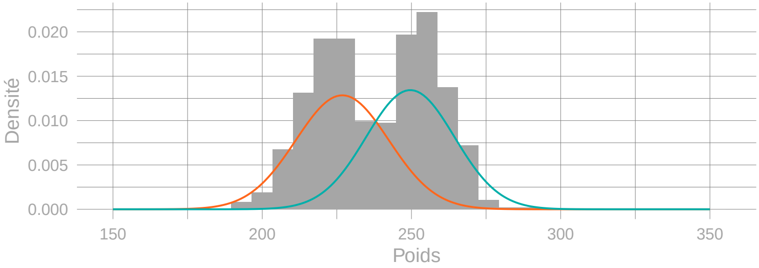
Distributions initiales

Étape 3: Maximisation

$$\hat{\mu}_1^{(1)} = \frac{\sum_{i=1}^{N}(1-\hat{\gamma}_i^{(0)})y_i}{\sum_{i=1}^{N}(1-\hat{\gamma}_i^{(0)})} = 227 \qquad \hat{\mu}_2^{(1)} = \frac{\sum_{i=1}^{N}(\hat{\gamma}_i^{(0)})y_i}{\sum_{i=1}^{N}(\hat{\gamma}_i^{(0)})} = 250$$

$$\hat{\sigma}_1^{2(1)} = \frac{\sum_{i=1}^{N}(1-\hat{\gamma}_i^{(0)}(y_i-\hat{\mu}_1^{(1)}))}{\sum_{i=1}^{N}(1-\hat{\gamma}_i^{(0)})} = 243 \qquad \hat{\sigma}_2^{2(1)} = \frac{\sum_{i=1}^{N}(\hat{\gamma}_i^{(0)}(y_i-\hat{\mu}_2^{(1)}))}{\sum_{i=1}^{N}(\hat{\gamma}_i^{(0)})} = 219$$

$$\hat{\pi}^{(1)} = 0.5$$

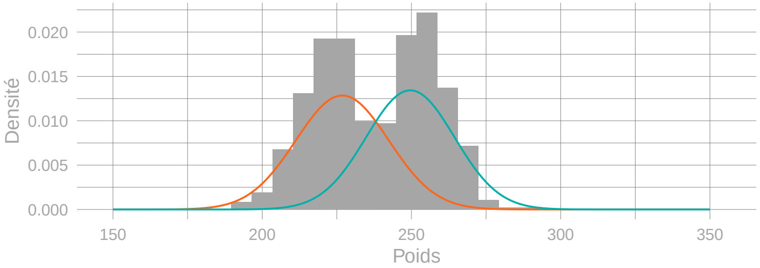


Distributions à la fin de l'itération 1

Itération 2

Étape 2: Expectation

$$\hat{\gamma}_i^{(1)} = \frac{\pi_2^{(1)} \phi_2^{(1)}(y_i)}{\pi_1^{(1)} \phi_1^{(1)}(y_i) + \pi_2^{(1)} \phi_2^{(1)}(y_i)}$$



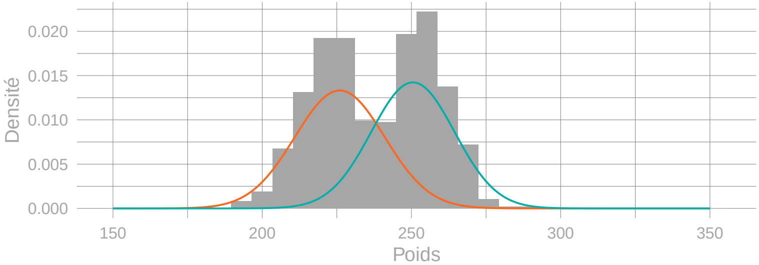
Distributions à la fin de l'itération 1

Étape 3: Maximisation

$\hat{\mu}_1^{(2)} = \frac{\sum_{i=1}^N (1-\hat{\gamma}_i^{(1)}) y_i}{\sum_{i=1}^N (1-\hat{\gamma}_i^{(1)})} = 226$ $\hat{\mu}_2^{(2)} = \frac{\sum_{i=1}^N (\hat{\gamma}_i^{(1)}) y_i}{\sum_{i=1}^N (\hat{\gamma}_i^{(1)})} = 250$

$\hat{\sigma}_1^{2(1)} = \frac{\sum_{i=1}^N (1-\hat{\gamma}_i^{(1)})(y_i-\hat{\mu}_1^{(2)}))}{\sum_{i=1}^N (1-\hat{\gamma}_i^{(1)})} = 226$ $\hat{\sigma}_2^{2(2)} = \frac{\sum_{i=1}^N (\hat{\gamma}_i^{(1)}(y_i-\hat{\mu}_2^{(2)}))}{\sum_{i=1}^N (\hat{\gamma}_i^{(1)})} = 195$

$\hat{\pi}^{(2)} = 0.5$



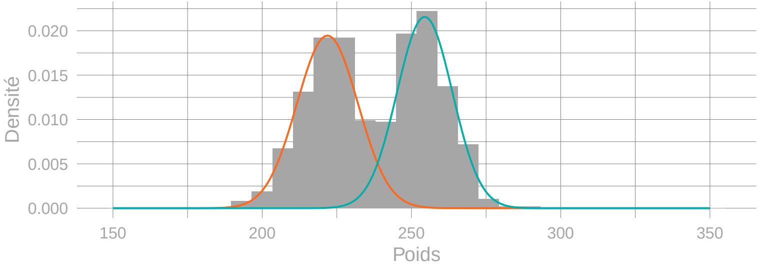
Distributions à la fin de l'itération 2

Itération 12

$$\hat{\mu}_1^{(2)} = \frac{\sum_{i=1}^N (1-\hat{\gamma}_i^{(1)}) y_i}{\sum_{i=1}^N (1-\hat{\gamma}_i^{(1)})} = 222 \qquad \hat{\mu}_2^{(2)} = \frac{\sum_{i=1}^N (\hat{\gamma}_i^{(1)}) y_i}{\sum_{i=1}^N (\hat{\gamma}_i^{(1)})} = 254$$

$$\hat{\sigma}_1^{2(1)} = \frac{\sum_{i=1}^N (1-\hat{\gamma}_i^{(1)})(y_i-\hat{\mu}_1^{(2)}))}{\sum_{i=1}^N (1-\hat{\gamma}_i^{(1)})} = 104 \qquad \hat{\sigma}_2^{2(2)} = \frac{\sum_{i=1}^N (\hat{\gamma}_i^{(1)}(y_i-\hat{\mu}_2^{(2)}))}{\sum_{i=1}^N (\hat{\gamma}_i^{(1)})} = 86$$

$$\hat{\pi}^{(2)} = 0.5$$



Distributions à la fin de l'itération 12

Propriétés

- La vraisemblance augmente nécessairement à chaque itération
- Pour éviter un maximum local, il est préférable de refaire l'estimation avec divers ensembles de valeurs de départ

Résumé

- L'algorithme EM permet d'estimer la distribution des densités d'un modèle de mélange de densités.

Dempster, Arthur P, Nan M Laird, and Donald B Rubin. 1977. "Maximum Likelihood from Incomplete Data via the Em Algorithm." *Journal of the Royal Statistical Society: Series B (Methodological)* 39 (1): 1–22.

Hastie, Trevor, Robert Tibshirani, and Jerome Friedman. 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer Science & Business Media.