# An introduction for mixture modelling for unsupervised clustering Mini-tutorial

#### Nicole M White

Australian Centre for Health Services Innovation (AusHSI) Queensland University of Technology

July 1, 2021





- Pinite mixture models
- 3 Dirichlet Process Mixture models
- 4 Profile regression
- **6** Model fitting and inference
- 6 References

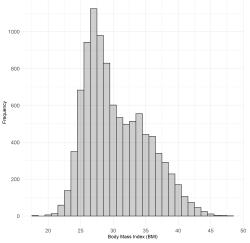
- 2 Finite mixture models
- 3 Dirichlet Process Mixture models
- 4 Profile regression
- **5** Model fitting and inference
- 6 References



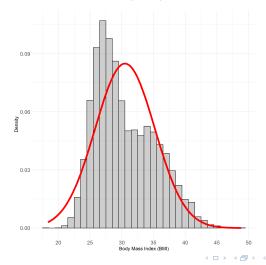
Chapman & Hall/CRC Handbooks of Modern Statistical Methods Handbook of **Mixture Analysis** Edited by Sylvia Frühwirth-Schnatter Gilles Celeux Christian P. Robert CRC Press

## A motivating example

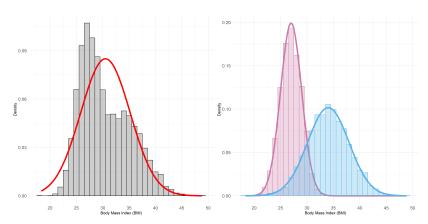
Distribution of body mass index (BMI) for 10,000 participants.



#### Distribution of body mass index (BMI) for 10,000 participants.



#### Distribution of body mass index (BMI) for 10,000 participants



Unsupervised clustering  $\leftrightarrow$  Identifying subgroups

Example clustering methods:

- Hierarchical clustering
- K-means
- Mixture models

Itodo add figure to demonstrate cluster separation here ;e.g. k-means]

#### Examples of clustering using mixture models

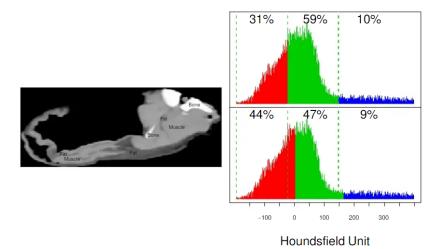


Figure 3: Experimental data, Sheep CT (Alston & Mengersen [ref])

#### Examples of clustering using mixture models

#### Spike sorting

Overview

- Show unsorted datasets from book chapter
- Show mixture solution with spikes in different colours



### Mixture model ingredients

Overview

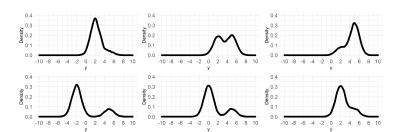
Data are drawn from a convex combination of components For K groups/clusters:

$$p(y) = \eta_1 p(y|\theta_1) + \ldots + \eta_K p(y|\theta_K)$$
$$= \sum_{k=1}^K \eta_k p(y|\theta_k)$$

- $\eta = (\eta_1, \dots, \eta_K)$ : Mixture weights;  $\sum_{k=1}^K \eta_k = 1$
- $p(y|\theta_k)$ :  $k^{th}$  Mixture component; same parametric family

### A simple 2-component mixture model

$$y_i \sim \eta_1 \mathcal{N}\left(\mu_1, 1\right) + \eta_2 \mathcal{N}\left(\mu_2, 1\right)$$



General formulation:

$$p(y_i) = \sum_{k=1}^K \eta_k p(y_i | \boldsymbol{\theta}_k)$$

Latent class analysis (*J* items)

$$p(y_i|\boldsymbol{\theta}_k) = \prod_{j=1}^J p(y_{ij}|\theta_{jk})$$

General formulation:

$$p(y_i) = \sum_{k=1}^K \eta_k p(y_i | \boldsymbol{\theta}_k)$$

Hidden Markov models

[TODO]

0000000000000 0000

## General formulation:

$$p(y_i) = \sum_{k=1}^K \eta_k p(y_i | \boldsymbol{\theta}_k)$$

• Latent class regression:  $\eta_k \to \eta_k(x_i)$ 

$$\eta_{k}(\mathbf{x}_{i}) = \frac{\exp\left(\mathbf{x}_{i}^{T}\beta_{k}\right)}{\sum_{l=1}^{K}\exp\left(\mathbf{x}_{i}^{T}\beta_{l}\right)}$$
$$\beta_{K} = 0$$

#### Mixture model examples

Focus of mini-tutorial: cross-sectional, continuous data

- Finite mixture model
- Dirichlet Process mixture model
- Profile regression

Bayesian approaches to inference: Markov chain Monte Carlo (MCMC)



0000000000000 •000

- Pinite mixture models
- 3 Dirichlet Process Mixture models
- 4 Profile regression
- **5** Model fitting and inference
- 6 References

Assume a fixed number of components KEach data point has a probability of belonging to each

## Estimating a Finite Mixture Model

Aim is to learn  $\eta_{1,...,K}$  and  $\theta_{1,...,K}$ Both are conditional on kIntroduce a latent variable. z

- One per observation: y<sub>i</sub>, z<sub>i</sub>
- Each  $z_i$  is discrete:  $1, \ldots, K$  with  $Pr(z_i = k) = \eta_k$  [check thesis
- $y_i$  belongs to cluster k iff  $z_i = k$



$$Pr(z_i = k|y_i, \cdot) = \frac{p(y_i|\theta_k, z_i = k)Pr(z_i = k)}{\sum_{l=1}^{K} p(y_i|\theta_l, z_i = l)Pr(z_i = l)}$$

- 3 Dirichlet Process Mixture models
  - Stick breaking process
- 4 Profile regression
- **5** Model fitting and inference
- 6 References

- 3 Dirichlet Process Mixture models Stick breaking process
- 4 Profile regression
- **5** Model fitting and inference
- 6 References



- 3 Dirichlet Process Mixture models Stick breaking process Polya Urn
- 4 Profile regression
- **5** Model fitting and inference
- 6 References



- 3 Dirichlet Process Mixture models Stick breaking process Chinese Restaurant Process
- 4 Profile regression
- **5** Model fitting and inference
- 6 References



- 2 Finite mixture models
- 3 Dirichlet Process Mixture models
- 4 Profile regression
- **5** Model fitting and inference
- 6 References

- 3 Dirichlet Process Mixture models
- 4 Profile regression
- **6** Model fitting and inference Choosing K

- 3 Dirichlet Process Mixture models
- 4 Profile regression
- **6** Model fitting and inference R implementation Choosing K



- 3 Dirichlet Process Mixture models
- 4 Profile regression
- **6** Model fitting and inference Inferring likely clusterings Choosing K



- Label switching conumdrum
- Unswitching vs. xxx

- 3 Dirichlet Process Mixture models
- 4 Profile regression
- **6** Model fitting and inference Choosing K

#### Frame Title

- AIC, BIC
- variants of DIC

- 2 Finite mixture models
- 3 Dirichlet Process Mixture models
- 4 Profile regression
- **5** Model fitting and inference
- 6 References

Thanks!

https://www.latexstudio.net/archives/4051.html