

# Normal Mixtures for Clustering

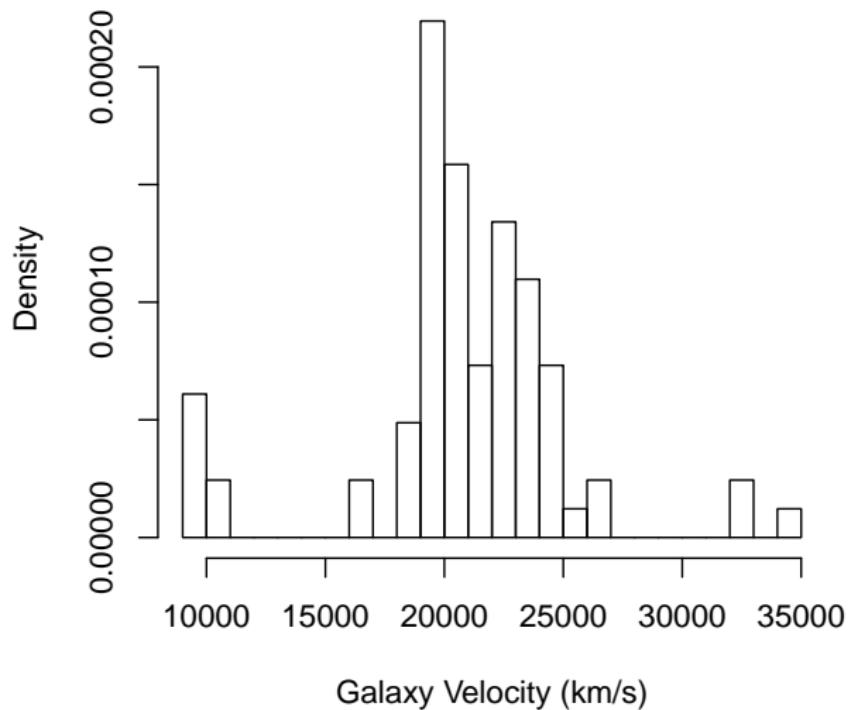
Matthew S. Shotwell, Ph.D.

Department of Biostatistics  
Vanderbilt University School of Medicine  
Nashville, TN, USA

April 17, 2020

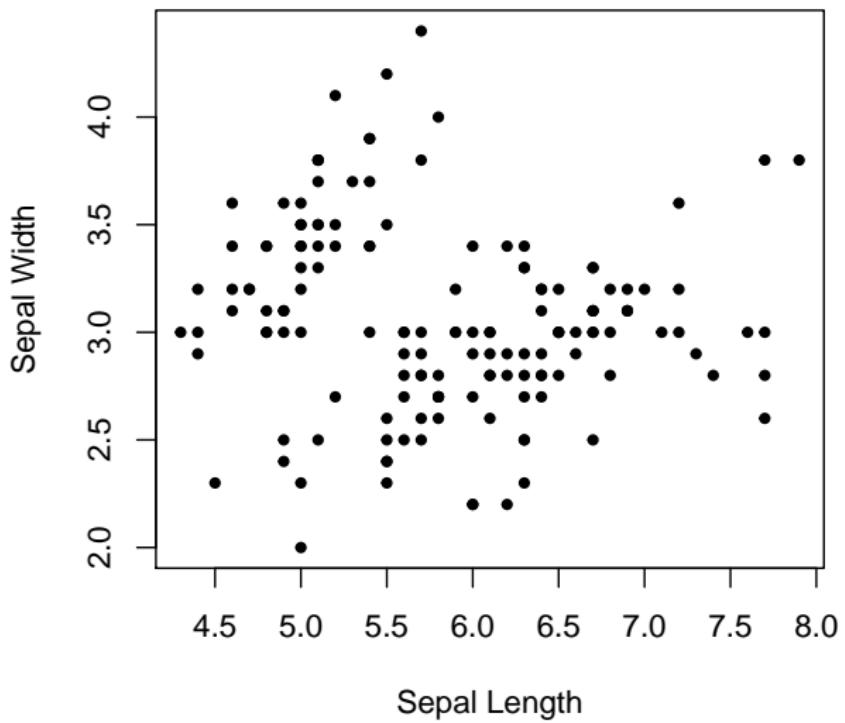
# Galaxies Data

## Histogram of Galaxy Velocities



# Iris Data

Iris (flower) Morphology



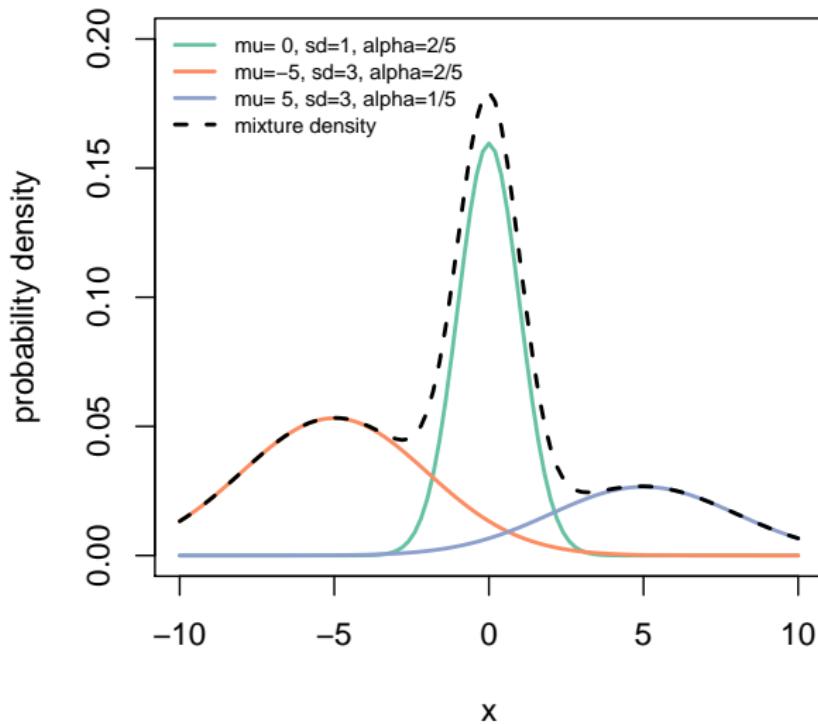
# Mixture model

- ▶  $f(x) = \sum_{m=1}^M \alpha_m f_m(x, \theta_m)$
- ▶  $\alpha_m$  are scalar weights,  $\sum_m \alpha_m = 1$
- ▶  $f_m(x, \theta_m)$  is a probability density function with parameter  $\theta_m$
- ▶ homogeneous:  $f_m$  all of same type, e.g., normal
- ▶ non-homogeneous:  $f_m$  have different types, e.g., normal, skew-normal

# Normal mixture model

- ▶  $f(x) = \sum_{m=1}^M \alpha_m \phi(x, \mu_m, \Sigma_m)$
- ▶  $\mu_m$  - mean
- ▶  $\Sigma_m$  - variance-covariance

# Example normal mixture w/3 components (M=3)



# Maximum likelihood estimation

- ▶ fit to data using maximum likelihood estimation
- ▶ given a sample of data  $x_1, \dots, x_N$
- ▶ estimate  $\theta$  by maximizing log likelihood:

$$l(\theta; x_1, \dots, x_N) = \sum_i \log f(x_i, \theta)$$

where

$$f(x_i, \theta) = \sum_{m=1}^M \alpha_m f_m(x_i, \theta_m)$$

- ▶ log-likelihood difficult because  $f(x_i, \theta)$  is a sum (log of sum?)
- ▶ cannot find close-form MLEs

## Augmented likelihood (data-augmentation)

The MLE problem is made easier by augmenting the log likelihood with a “latent variable”, i.e., an unknown variable that, if known, would make the log likelihood easier to maximize

# Augmented likelihood (data-augmentation)

- ▶ define latent variable  $z_i = [z_{i1}, \dots, z_{iM}]$  where

$$z_{im} = \begin{cases} 1 & \text{if } x_i \text{ generated by } f_m \\ 0 & \text{otherwise} \end{cases}$$

- ▶  $z$ 's define the clusters

# Augmented likelihood (data-augmentation)

- ▶ treat  $z_i$  as a multinomial R.V. with probabilities  $\alpha_m$

$$f(z_i) = \prod_{m=1}^M \alpha_m^{z_{im}}$$

- ▶ rewrite density as a product instead of sum

$$f(x_i|z_i) = \prod_{m=1}^M f_m(x_i, \theta_m)^{z_{im}}$$

- ▶ augmented log likelihood is easier to work with:

$$l(\theta; x_1, \dots, x_N, z_1, \dots, z_N) = \sum_{i=1}^N \sum_{m=1}^M z_{im} \log(\alpha_m f_m(x, \theta_m))$$

## Augmented likelihood (data-augmentation)

$$l(\theta; x_1, \dots, x_N, z_1, \dots, z_N) = \sum_{i=1}^N \sum_{m=1}^M z_{im} \log(\alpha_m f_m(x, \theta_m))$$

- ▶  $z$ 's define the clusters
- ▶ if had  $z$ 's and  $f_m(x) = \phi(x)$  (normal density), MLEs for  $\mu_m$  and  $\Sigma_m$  would be the within-cluster sample estimates
- ▶ unfortunately  $z$ 's are unknown

# E-M algorithm

$$l(\theta; x_1, \dots, x_N, z_1, \dots, z_N) = \sum_{i=1}^N \sum_{m=1}^M z_{im} \log(\alpha_m f_m(x, \theta_m))$$

want to maximize  $l$  w.r.t.  $\theta$  but  $z$ 's are unknown, algorithmic solution (EM algorithm):

1. initialize cluster assignments:  $z_i$
2. initialize  $\theta$  given  $z_i$  (maximize augmented likelihood)
3. Iterate until convergence:

E-step: given  $\theta$ , substitute  $z_{im}$  with  $E[z_{im}|\theta, x_1, \dots, x_N]$

M-step: given  $z_{im}$ , maximize augmented likelihood

# E-M algorithm

- ▶ “Expectation” step: substitute  $z_{im}$  with its expectation:

$$\gamma_{im} = E[z_{im}|\theta, x_1, \dots, x_N] = P(z_{im} = 1|\theta, x_1, \dots, x_N)$$

$$\begin{aligned} P(z_{im} = 1|\theta, x_1, \dots, x_N) &\propto P(x_i|z_{im} = 1)P(z_{im} = 1) \\ &= f_m(x_i, \theta_m)\alpha_m \end{aligned}$$

# E-M algorithm

- “Maximization” step: maximize the augmented likelihood:

$$l(\theta; x_1, \dots, x_N, \gamma_1, \dots, \gamma_N) = \sum_{i=1}^N \sum_{m=1}^M \gamma_{im} \log(\alpha_m f_m(x, \theta_m))$$

# E-M algorithm for normal mixtures

- ▶  $f(x, \theta_m) = \phi(x, \mu_m, \Sigma_m)$

- ▶ “Expectation” step:

$$\gamma_{im} = \frac{\phi_m(x_i, \mu_m, \Sigma_m)\alpha_m}{\sum_j \phi_m(x_i, \mu_j, \Sigma_j)\alpha_j}$$

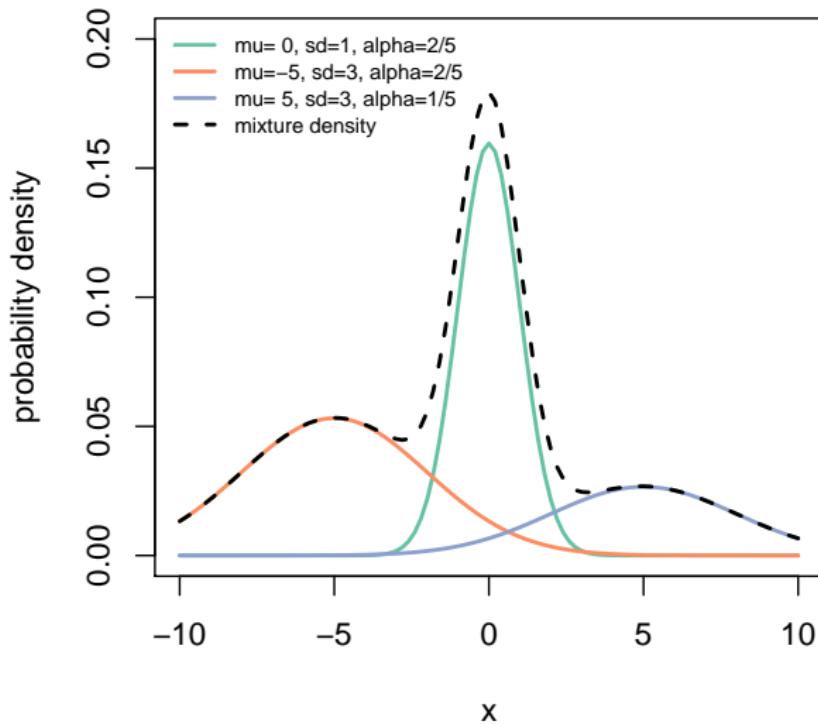
- ▶ “Maximization” step:

$$\mu_m = \frac{\sum_i \gamma_{im} x_i}{\sum_i \gamma_{im}}$$

$$\Sigma_m = \frac{\sum_i \gamma_{im} (x_i - \mu_m)^T (x_i - \mu_m)}{\sum_i \gamma_{im}}$$

$$\alpha_m = \frac{1}{N} \sum_i \gamma_{im}$$

# Example normal mixture w/3 components (M=3)



# Clustering with mixtures

- ▶ compute  $\gamma_{im}$  at final iteration
- ▶  $\gamma_{im} = P(z_{im} = 1|x_i)$
- ▶ assign  $x_i$  to cluster with largest  $\gamma_{im}$