The observations $y_1, \ldots, y_n$ form a sample from the following finite mixture of normal distributions:

$$p(y_i | \theta) = \sum_{j=1}^{k} w_j p_N(y_i | \mu_j, \sigma_j^2)$$

where $\theta = (\mu, \sigma^2, w, \mu = (\mu_1, \ldots, \mu_k)')$, $\sigma^2 = (\sigma_1^2, \ldots, \sigma_k^2)'$, $w = (w_1, \ldots, w_k)'$, and $p_N(y | \mu, \sigma^2)$ is the density of a normal distribution with mean $\mu$ and variance $\sigma^2$ evaluated at $y$. Therefore,

$$p(y | \theta) = \prod_{i=1}^{n} \left( \sum_{j=1}^{k} w_j p_N(y_i | \mu_j, \sigma_j^2) \right)$$

Using latent indicators $z_1, \ldots, z_n$, such that $z_i \in \{1, \ldots, k\}$ and $p(z_i = j | \theta) = w_j$, the augmented model for $(y, z)$ has the following joint density:

$$p(y, z | \theta) = p(y | z, \theta) p(z | \theta) = \left[ \prod_{j=1}^{k} \prod_{i \in I_j} p_N(y_i | \mu_j, \sigma_j^2) \right] \prod_{i=1}^{n} p(z_i | \theta)$$

where $I_j = \{ i : z_i = j \}$.

**Bayesian Inference (MCMC)**

The priors are $\mu_j \sim N(m, \tau_j \sigma_j^2)$, $\sigma_j^2 \sim IG(a / 2, b / 2)$, $m \sim N(m_0, \tau_m)$, $\tau \sim IG(c / 2, d / 2)$, and $w \sim D(\alpha)$, with $a, b, c, d, m_0, \tau_m$, and $\alpha = (\alpha_1, \ldots, \alpha_k)'$, known hyperparameters. Let $n_j = \text{card}(I_j)$, $\bar{y}_{j} = \sum_{i \in I_j} y_i$, and $n_j \sigma_j^2 = \sum_{i \in I_j} (y_i - \bar{y}_j)^2$. The full conditional distributions are as follows.

- $[\sigma_j^2 | \mu, z, y] \sim IG \left( \frac{a + n_j + 1}{2}, \frac{1}{2} \left[ b + n_j \sigma_j^2 + n_j (\mu_j - \bar{y}_j)^2 + \frac{1}{\tau}(\mu_j - m)^2 \right] \right)$
- $[\mu_j | \sigma_j^2, m, \tau, z, y] \sim N \left( \frac{m \bar{y}_j + m}{\tau \sigma_j^2 + 1}, \frac{\tau \sigma_j^2}{\tau \sigma_j^2 + 1} \right)$
- $[\tau | \sigma_j^2, \mu, m, y] \sim IG \left( \frac{c + k}{2}, \frac{1}{2} \left[ d + \sum_{j=1}^{k} \frac{(\mu_j - m)^2}{\sigma_j^2} \right] \right)$
- $[z_i] \in \{1, \ldots, k\}$, with $p(z_i = j | \theta, y_i) = \frac{\omega_j}{\omega_1 + \ldots + \omega_k}$ and $\omega_l = w_l p_N(y_i | \mu_l, \sigma_l^2)$ for $l = 1, \ldots, k$.
- $[w | \mu, \sigma^2, z, y] \sim D(\alpha + n)$, where $n = (n_1, \ldots, n_k)$.
- $[m | \sigma^2, \tau, \mu] \sim N \left( \frac{\tau - 1}{\tau} \sum_{j=1}^{k} \sigma_j^{-2}, \frac{\tau - 2}{\tau} \sum_{j=1}^{k} \sigma_j^{-2} \mu_j \right)$
Maximum Likelihood Inference (EM)

The Expectation-Maximization (EM) algorithm finds $\hat{\theta}$ that maximizes the (incomplete) log-likelihood, ie.

$$\hat{\theta} \equiv \arg \max_\theta l(\theta|y)$$

where

$$l(\theta|y) = \sum_{i=1}^n \log \left[ \sum_{j=1}^k w_j (2\pi \sigma_j^2)^{-1/2} \exp \left\{ \frac{1}{2\sigma_j^2} (y_i - \mu_j)^2 \right\} \right]$$

by iteratively cycling through the following two steps:

- **E-step**: Compute the integral $Q(\theta, \theta^{(l)}) = \int \log \{p(y, z|\theta)\} p(z|y, \theta^{(l)})dz$

- **M-step**: Find $\theta^{(l+1)}$ such that $\theta^{(l+1)} = \arg \max_\theta Q(\theta, \theta^{(l)})$

The EM algorithm for the mixture of normal model case, with $\theta^{(0)}$ as starting value, cycles through $l = 1, \ldots, L$ as follows.

For $i = 1, \ldots, n$ and $j = 1, \ldots, k$ compute

$$\delta_{ij} = p(z_i = j|y_i, \theta^{(l)}) = \frac{w_j^{(l)} p_N(y_i|\mu_j^{(l)}, \sigma_j^{2(l)})}{p(y_i|\theta^{(l)})}$$

For $j = 1, \ldots, k$, compute

$$w_j^{(l+1)} = n^{-1} \sum_{i=1}^n \delta_{ij}$$

$$\mu_j^{(l+1)} = \frac{\sum_{i=1}^n y_i \delta_{ij}}{nw_j^{(l+1)}}$$

$$\sigma_j^{2(l+1)} = \frac{\sum_{i=1}^n (y_i - \mu_j^{(l)})^2 \delta_{ij}}{nw_j^{(l+1)}}$$

It can be shown that the sequence $\{\theta^{(1)}, \theta^{(2)}, \ldots\}$ converges to $\hat{\theta} = \arg \max_\theta l(\theta|y)$ as $l \to \infty$ (for more details about the EM algorithm, see Dempster, Laird and Rubin, 1977).
References (chronological order)

- Richardson and Green (1997) On Bayesian analysis of mixtures with an unknown number of components (with discussion), *JRSS-B*, 59, 731-792.