Univariate Mixture of Normals: MCMC and EM algorithms

by Hedibert Freitas Lopes
Applied Econometrics, Spring 2005
Graduate School of Business
University of Chicago

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 and $p_N(y|\mu, \sigma^2)$ is the density of a normal distribution with mean μ and variance σ^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 $\mu_j \sim N(m, \tau\sigma_j^2)$, $\sigma_j^2 \sim IG(a/2, b/2)$, $m \sim N(m_0, \tau_m)$, $\tau \sim IG(c/2, d/2)$, and $w \sim D(\alpha)$, with $a, b, c, d, \mu_0, \tau_m$, and $\alpha = (\alpha_1, \ldots, \alpha_k)'$, known hyperparameters. Let $n_j = \text{card}(I_j)$, $n_j \bar{y}_j = \sum_{i \in I_j} y_i$, and $n_j s_j^2 = \sum_{i \in I_j} (y_i - \bar{y}_j)^2$. The full conditional distributions are as follows.

- $[\sigma_j^2|\mu, z, y] \sim IG\left(\frac{a+n_j+1}{2}, \frac{1}{2}\left[b + n_j s_j^2 + n_j(\mu_j - \bar{y}_j)^2 + \frac{1}{\tau}(\mu_j - m)^2\right]\right)$
- $[\mu_j|\sigma^2, m, \tau, z, y] \sim N\left(\frac{\tau n_j \bar{y}_j + m}{\tau n_j + 1} \; , \; \frac{\tau\sigma_j^2}{\tau n_j + 1}\right)$
- $[\tau|\sigma^2, \mu, m, y] \sim IG\left(\frac{c+k}{2}, \frac{1}{2}\left[d + \sum_{j=1}^{k} \frac{(\mu_j - m)^2}{\sigma_j^2}\right]\right)$
- $[z_i] \in \{1, \ldots, k\}$, with $p(z_i = j|\theta, y_i) = \frac{\omega_j}{\omega_1 + \cdots + \omega_k}$ 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 $n = (n_1, \ldots, n_k)$.
- $[m|\sigma^2, \tau, \mu] \sim N\left((\tau_m^{-1} + \tau^{-1}\sum_{j=1}^{k}\sigma_j^{-2})(\tau_m^{-1} m_0 + \tau^{-2}\sum_{j=1}^{k}\sigma_j^{-2}\mu_j), (\tau_m + \tau^{-1}\sum_{j=1}^{k}\sigma_j^{-2})\right)$

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

$$\begin{aligned} w_j^{(l+1)} &= n^{-1} \sum_{i=1}^{n} \delta_{ij} \\ \mu_j^{(l+1)} &= \frac{\sum_{i=1}^{n} y_i \delta_{ij}}{n w_j^{(l+1)}} \\ \sigma_j^{2(l+1)} &= \frac{\sum_{i=1}^{n} (y_i - \mu_j^{(l)})^2 \delta_{ij}}{n w_j^{(l+1)}} \end{aligned}$$

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)

- Dempster, Laird and Rubin (1977) Maximum likelihood from incomplete data via the EM algorithm, *JRSS-B*, 39, 1-38.
- Titterington, Smith and Makov (1984) *Statistical Analysis of Finite Mixture Distributions*, New York: Wiley.
- Roeder (1990) Density estimation with confidence sets exemplified by superclusters and voids in the galaxies, *JASA*,85,617-624.
- Crawford (1994) An application of the Laplace method to finite mixture distributions, *JASA*,89, 259-267.
- Chib (1995) Marginal likelihood from the Gibbs output, *JASA*,90,1313-1321.
- Carlin and Chib (1995) Bayesian model choice via Markov chain Monte Carlo methods, *JRSS-B*,473-484
- Escobar and West (1995) Bayesian density estimation and inference using mixtures, *JASA*,90,577-588.
- Phillips and Smith (1996) Bayesian model comparison with jump diffusions, in Markov Chain Monte Carlo in Practice (eds Gilks, Richardson and Spiegelhalter), chapter 13, pp. 215-239. London: Chapman and Hall.
- Richardson and Green (1997) On Bayesian analysis of mixtures with an unknown number of components (with discussion), *JRSS-B*, 59, 731-792.
- Stephens (2000) Dealing with label switching in mixture models, *JRSS-B*, 62, 795-809.
- Stephens (2000) Bayesian analysis of mixtures with an unknown number of components: an alternative to reversible jump methods. *Annals of Statistics*, 28.