Simple Linear Regression with Mixture of Normal Errors

by Hedibert Freitas Lopes
Graduate School of Business
University of Chicago

Model, Prior and Data Augmentation

The observations y_i depend on the p-dimensional vector of regressors x_i as follows:

$$p(y|\mu, \sigma^2, w) = \prod_{i=1}^{n} \sum_{j=1}^{k} w_j p_N(y_i|\mu_j + x_i'\beta, \sigma_j^2)$$

where $\mu = (\mu_1, \ldots, \mu_k)', \sigma^2 = (\sigma_1^2, \ldots, \sigma_k^2)'$, $w = (w_1, \ldots, w_k)'$, x_i a p-dimensional vector of regressors with regression coefficients β, and $p_N(a|b, c)$ is the density of the univariate normal distribution with mean b and variance c evaluated at a. Throughout this note we keep $X = (x_1, \ldots, x_n)'$ fixed and known and, therefore, omitted from the notation hereafter. For the parameters in $\theta = (\mu, \sigma^2, w, m, \tau)$, we used the following prior specification

$$\pi(\theta) = \pi(\beta)\pi(\tau)\pi(m)\pi(w) \prod_{j=1}^{k} \pi(\mu_j|m, \tau, \sigma_j^2)\pi(\sigma_j^2|a, b)$$

with $\beta \sim N(\beta_0, V_\beta)$, $\mu_j \sim N(m, \tau\sigma_j^2)$ and $\sigma_j^2 \sim IG(a/2, b/2)$, for $j = 1, \ldots, k$, $w \sim D(\alpha)$,$m \sim N(m_0, \tau_m)$, and $\tau \sim IG(c/2, d/2)$. The hyperparameters a, b, c, d, m_0, τ_m, and $\alpha = (\alpha_1, \ldots, \alpha_k)'$, are known. Therefore, by Bayes theorem, the posterior distribution of θ is

$$\pi(\theta|y) \quad \propto \quad \pi(\theta) \prod_{i=1}^{n} \sum_{j=1}^{k} w_j dN(y_i|\mu_j + x_i'\beta, \sigma_j^2)$$

which is clearly analytically intractable. However, conditioning on latent variables $z = (z_1, \ldots, z_n)$, with $z_i \in \{1, \ldots, k\}$, and $I_j = \{i : z_i = j\}$, for $j = 1, \ldots, k$ and $i = 1, \ldots, n$, the joint posterior distribution of (θ, z) can be written as

$$\pi(\theta, z|y) \propto \pi(\theta)\pi(z|\theta)p(y|\theta, z)$$

with both

$$\pi(\theta|y, z) \quad \propto \quad \pi(\beta)\pi(\tau)\pi(m)\pi(w)\pi(z|w) \prod_{j=1}^{k} \left[\prod_{i \in I_j} \pi(\mu_j)p(\sigma_j^2)p_N(y_i|\mu_j + x_i'\beta, \sigma_j^2) \right]$$

and

$$\pi(z|\theta, y) \quad = \quad \prod_{i=1}^{n} \pi(z_j|w)p_N(y_i|\mu_j + x_i'\beta, \sigma_j^2)$$

much more straightforward to sample from. Therefore, θ and z can be easily and iteratively sampled within a Gibbs sampler, as described next.

Full conditionals of $\mu, \sigma^2, \tau, w, m$ and β

Conditionally on $z = (z_1, \ldots, z_n)'$, let $\varepsilon_i = y_i - \mu_{z_i} - x_i'\beta$, $y_i^* = y_i - \mu_{z_i}$, $y^* = (y_1^*, \ldots, y_n^*)'$, and $\Omega = \text{diag}(\sigma^2_{z_1}, \ldots, \sigma^2_{z_n})$, For $i = 1, \ldots, n$. Additionally, let $n_j = \text{card}(I_j)$, $n_j\bar{\varepsilon}_j = \sum_{i \in I_j} \varepsilon_i$, and $n_j s_j^2 = \sum_{i \in I_j}(\varepsilon_i - \bar{\varepsilon}_j)^2$, for $j = 1, \ldots, k$. Finally, if $n = (n_1, \ldots, n_k)$, then the full conditional distributions for $\mu, \sigma^2, \tau, w, m$ and β are as follows.

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

$$[\mu_j|\sigma^2, m, \tau, z, y] \sim N\left(\frac{\tau n_j \bar{\varepsilon}_j + m}{\tau n_j + 1} \ , \ \frac{\tau \sigma_j^2}{\tau n_j + 1}\right)$$

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

$$[\sigma_j^2|\mu, \beta, z, y] \sim IG(0.5(a + n_j), 0.5(b + n_j s_j^2 + n_j(\mu_j - \bar{\varepsilon}_j)^2 + \tau^{-1}(\mu_j - m)^2))$$

- $[\tau|\sigma^2, \mu, m, y] \sim IG(0.5(c + k), 0.5(d + \sum_{j=1}^k (\mu_j - m)^2/\sigma_j^2))$
- $[w|\mu, \sigma^2, z, y] \sim D(\alpha + n)$
- $[m|\sigma^2, \tau, \mu] \sim N((\tau_m^{-1} + \tau^{-1}\sum_{j=1}^k \sigma_j^{-2})^{-1}(\tau_m^{-1}m_0 + \tau^{-2}\sum_{j=1}^k \sigma_j^{-2}\mu_j), (\tau_m^{-1} + \tau^{-1}\sum_{j=1}^k \sigma_j^{-2})^{-1})$
- $[\beta|\mu, \sigma^2, z, y] \sim N((V_\beta^{-1} + X'\Omega^{-1}X)^{-1}(V_\beta^{-1}\beta_0 + X'\Omega^{-1}y^*), (V_\beta^{-1} + X'\Omega^{-1}X)^{-1})$

Full conditionals of $z_1, \ldots, z_n$

Conditionally on μ, σ^2, τ, w and β,

$$p(z|\mu, \sigma^2, \tau, w, \beta) = \prod_{i=1}^n p(z_i|\mu, \sigma^2, w, y_i, \beta)$$

so z_i is sampled from $\{1, \ldots, k\}$ with probabilities $\omega_{i1}, \ldots, \omega_{ik}$, where

$$w_{ij} \equiv Pr(z_i = j|\mu, \sigma^2, w, y_i, \beta) = \frac{w_j p_N(y_i|\mu_j + x_i'\beta, \sigma_j^2)}{\sum_{l=1}^k w_l p_N(y_i|\mu_l + +x_i'\beta, \sigma_l^2)}$$

for $i = 1, \ldots, n$ and $j = 1, \ldots, k$.

Mixture of normal distributions

It is worth noting that the case particular case where $\beta_0 = 0_p$ and $V_\beta = 0_{p\times p}$ corresponds to the traditional mixture of normal distributions.