Lab 6: Gibbs sampling for mixture models using data augmentation – STA 360/602

Rebecca C. Steorts

March 27, 2017

A Three Component Mixture Model

Consider a three component mixture of normal distribution with a common prior on the mixture component means, the error variance and the variance within mixture component means. The prior on the mixture weights w is a three component Dirichlet distribution. (The data for this problem can be found in "Lab6Mixture.csv").

$$
\begin{aligned}
p(Y_i|\mu_1, \mu_2, \mu_3, w_1, w_2, w_3, \varepsilon^2) &= \sum_{j=1}^{3} w_i N(\mu_j, \varepsilon^2) \\
\mu_j|\mu_0, \sigma_0^2 &\sim N(\mu_0, \sigma_0^2) \\
\mu_0 &\sim N(0, 3) \\
\sigma_0^2 &\sim IG(2, 2) \\
(w_1, w_2, w_3) &\sim Dirichlet(\mathbf{1}) \\
\varepsilon^2 &\sim IG(2, 2),
\end{aligned}
$$

for $i = 1, \ldots n$.

Specifically,

- w_1, w_2 and w_3 are the mixture weight of mixture components 1,2 and 3 respectively
- μ_1, μ_2 and μ_3 are the means of the mixture components
- ε^2 is the variance parameter of the error term around the mixture components.

Since we're building a hierarchical model for the means of the individual component, we have a common hyperprior, where, μ_0 is the mean parameter of this hyperprior, σ_0^2 is its variance parameter. Both of these have priors as well, but the parameters of those priors are fixed, where μ_0 has a Normal prior with mean 0 and variance 3, σ_0^2 has an Inverse-Gamma prior with shape and rate parameter of (2,2) respectively. Similarly, ε^2 has an Inverse-Gamma prior with shape and rate parameter of (2,2) respectively. While they have the same parametrisation, they do not share a prior). The mixture weights w_1, w_2, w_3 jointly come from a Dirichlet distribution, with parameter vector $(1, 1, 1)$. $w_1, w_2, w_3, \mu_1, \mu_2, \mu_3, \varepsilon^2, \mu_0$ and σ_0^2 are all random variables that we will estimate when we fit the model.

Task 1

Derive the joint posterior $p(w_1, w_2, w_3, \mu_1, \mu_2, \mu_3, \varepsilon^2, \mu_0, \sigma_0^2 | Y_1, ..., Y_N)$ up to a normalizing constant.

Task 2

Derive the full conditionals for all the parameters up to a normalizing constant.

- $p(w_1, w_2, w_3 | \mu_1, \mu_2, \mu_3, \varepsilon^2, Y_1, ..., Y_N) \propto$
- $p(\mu_1 | \mu_2, \mu_3, w_1, w_2, w_3, Y_1, ..., Y_N, \varepsilon^2, \mu_0, \sigma_0^2) \propto$
- $p(\mu_2 | \mu_1, \mu_3, w_1, w_2, w_3, Y_1, ..., Y_N, \varepsilon^2, \mu_0, \sigma_0^2) \propto$
- $p(\mu_3 | \mu_1, \mu_2, w_1, w_2, w_3, Y_1, ..., Y_N, \varepsilon^2, \mu_0, \sigma_0^2) \propto$
- $p(\varepsilon^2 | \mu_1, \mu_2, \mu_3, Y_1, ..., Y_N) \propto$
- $p(\mu_0 | \mu_1, \mu_2, \mu_3, \sigma_0^2) \propto$
- $p(\sigma_0^2 | \mu_0, \mu_1, \mu_2, \mu_3) \propto$

The data augmentation scheme

Since neither the joint posterior nor any of the full conditionals involving the likelihood are of a form that's easy to sample, we introduce a data augmentation scheme. A common solution is to introduce an additional set of random variables $\{Z_i\}_{i=1}^N$ that assign each observation to one of the mixture components with the proabilitiy of assignment being the respective mixture weight. If we condition on Z_i we can then write the likelihood of Y_i as

$$p(Y_i | Z_i, \mu_1, \mu_2, \mu_3, \varepsilon^2 = \sum_{j=1}^{N} N(\mu_j, \varepsilon^2)\delta_j(Z_i) = N(\mu_{Z_i}, \varepsilon^2)$$

$$P(Z_i = j) = w_j.$$

This means that conditional on Z_i we no longer have a sum of Normal pdfs in our likelihood, resulting in a significant simplification. Conditional on the $\{Z_i\}$

updates will be straightforward, only depending on the mixture component that any given Y_i is currently assigned to. The drawback is that we also have to update $\{Z_i\}_{i=1}^{N}$ as well, introducing extra steps into our sampler. Also note that the Dirichlet distribution is a conjugate prior for categorical variables.

Task 3

Where necessary, (re)derive the full conditionals under the data augmentation scheme.

Task 4

In task 3 you derived all the full conditionals, and due to data augmentation scheme they are all in a form that is easy to sample. Use these full conditionals to implement Gibbs sampling using the data from "Lab6Mixture.csv".

Task 5

Given tasks 1-4 and the provided solutions,

- Show traceplots for all estimated parameters
- Show means and 95% credible intervals for the marginal posterior distributions of all the parameters

Now suppose you re-run the sampler using 3 different starting values, are your results in a,b the same? Justify your reasoning by with visualizations.