

# mix-proj

*Rob McCulloch*

*November 12, 2019*

## Computational Statistics, A Suggested Project

We have looked at mixture modelling using the EM algorithm.

Let's see how we can use MCMC.

Here is our model,

The variable  $I_i$  is the latent variable indicated which mixture component the  $i^{th}$  observation comes from.

$$I_i \in \{1, 2, \dots, k\}, \quad P(I_i = j) = p_j.$$

Each mixture component is  $N(\mu_j, \sigma_j^2)$ .

$$\mu = (\mu_1, \mu_2, \dots, \mu_k), \quad \sigma = (\sigma_1, \sigma_2, \dots, \sigma_k)$$

Given the means  $\mu$  and standard deviations  $\sigma$ , and the mixture component for the  $i^{th}$  observation, we know the distribution of  $Y_i$ .

$$Y_i | I_i = j, \mu, \sigma \sim N(\mu_j, \sigma_j^2)$$

To specify a full Bayesian model we need to put priors on  $p$ ,  $\mu$ , and  $\sigma$ .

$$\mu_j \sim N(\bar{\mu}, \sigma_\mu^2), \quad \sigma_j^2 \sim \nu\lambda/\chi_\nu^2.$$

$$p = (p_1, p_2, \dots, p_k) \sim \text{Dirichlet}(\alpha)$$

The Gibbs sampler is:

$$p | I, \quad I | \mu, \sigma, p, y, \quad \mu | \sigma, I, y, \quad \sigma | \mu, I, y$$

where,

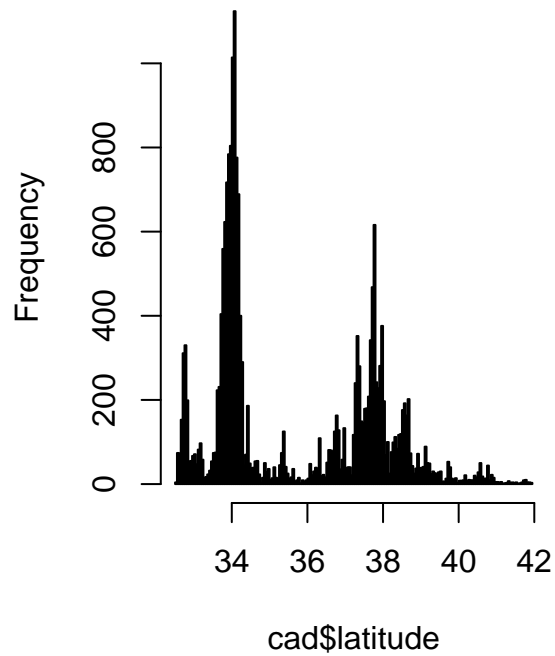
- the draw of  $p$  is a Dirichlet
- the draw of each  $I_i$  is an independent multinomial
- the draw of each  $\mu_j$  is an independent normal
- the draw of each  $\sigma_j$  is an independent inverted chi-squared.

Project:

- read (don't worry about getting everything) Chapter 22 of "Bayesian Data Analysis", third Edition, by Gelman et. al.
- code up the EM algorithm for the mixture model
- code up the Gibbs sampler
- compare EM with Gibbs on real and simulated data.

```
cad = read.csv("http://www.rob-mcculloch.org/data/calhouse.csv")
par(mfrow=c(1,2))
hist(cad$latitude,nclass=200)
hist(cad$longitude,nclass=200)
```

**Histogram of cad\$latitude**



**Histogram of cad\$longitude**

