

# STAT 238 - Bayesian Statistics

## Lecture Twenty Eight

Spring 2026, UC Berkeley

Aditya Guntuboyina

08 April 2026

### 1 Gibbs Sampler for Mixture Models

We observe real-valued data  $y_1, \dots, y_n$  and consider the model:

$$y_i \stackrel{\text{i.i.d.}}{\sim} pN(\mu_1, 1) + (1 - p)N(\mu_2, 1) \quad (1)$$

with  $p$  known and fixed (e.g.,  $p = 0.3$  as in the simulations) and  $\mu_1, \mu_2$  being unknown. The Gibbs sampler formulae can be easily generalized to more realistic cases (e.g., when  $p$  is unknown, variances are different from 1 etc. as we will see later). The log-likelihood corresponding to (1) is:

$$\sum_{i=1}^n \log(p\phi(y_i, \mu_1, 1) + (1 - p)\phi(y_i, \mu_2, 1)),$$

where  $\phi(x, \mu, \sigma^2)$  is the normal density with mean  $\mu$ , variance  $\sigma^2$  evaluated at  $x$ .

We will take a standard prior for  $\mu_1, \mu_2$  e.g.,  $\mu_1, \mu_2 \stackrel{\text{i.i.d.}}{\sim} N(0, C)$  for a large  $C$ . The posterior of  $\mu_1, \mu_2$  is then:

$$\pi(\mu_1, \mu_2 \mid \text{data}) \propto \phi(\mu_1, 0, C)\phi(\mu_2, 0, C) \prod_{i=1}^n (p\phi(y_i, \mu_1, 1) + (1 - p)\phi(y_i, \mu_2, 1)).$$

This posterior cannot be evaluated in closed form and numerical methods need to be used. The Gibbs sampler is one computational tool that is applicable here. The standard approach of using the Gibbs sampler in this problem proceeds via augmentation. This is explained below.

First observe that the model (1) can be rewritten in the following way:

$$z_i \stackrel{\text{i.i.d.}}{\sim} \text{Bernoulli}(p) \quad \text{and} \quad y_i \mid z_i = 1 \sim N(\mu_1, 1) \quad \text{and} \quad y_i \mid z_i = 0 \sim N(\mu_2, 1).$$

It should be clear that, under the above model, the marginal distribution of  $y_i$  coincides with (1).  $z_1, \dots, z_n$  can be thought of as unobserved latent variables which represent which of the two populations (corresponding to the distributions  $N(\mu_1, 1)$  and  $N(\mu_2, 1)$  respectively) the observation  $y_i$  comes from.

Gibbs sampler is implemented for jointly sampling from the posterior of  $\mu_1, \mu_2, z_1, \dots, z_n$  given the data. This requires being able to sample from the full conditionals

$$z \mid \mu_1, \mu_2, y \quad \text{and} \quad (\mu_1, \mu_2) \mid z, y.$$

where  $y = (y_1, \dots, y_n)$  is the data and  $z = (z_1, \dots, z_n)$ . It is easy to see that these full conditionals can be written in closed form as follows. Given  $\mu_1, \mu_2, y_1, \dots, y_n$ , the variables  $z_1, \dots, z_n$  are independent with

$$z_i \mid \mu_1, \mu_2, y \sim \text{Bernoulli} \left( \frac{p\phi(y_i, \mu_1, 1)}{p\phi(y_i, \mu_1, 1) + (1-p)\phi(y_i, \mu_2, 1)} \right).$$

In the limit  $C \rightarrow \infty$ , given  $z_1, \dots, z_n, y_1, \dots, y_n$ , the variables  $\mu_1, \mu_2$  are independent with

$$\mu_1 \mid z, y \sim N \left( \frac{\sum_{i=1}^n y_i z_i}{\sum_{i=1}^n z_i}, \frac{1}{\sum_{i=1}^n z_i} \right) \quad \text{and} \quad \mu_2 \mid z, y \sim N \left( \frac{\sum_{i=1}^n y_i (1 - z_i)}{\sum_{i=1}^n (1 - z_i)}, \frac{1}{\sum_{i=1}^n (1 - z_i)} \right)$$

Based on these full conditional distributions, the Gibbs sampler can be easily implemented. We shall give the exact form of the algorithm in the next lecture. It also turns out that the EM algorithm for computing the MLE is similar to the Gibbs sampler. We shall also look at this in the next lecture.