

Midterm Project (due 03/31/2017)

Download data X_1, \dots, X_n from

<https://www2.math.binghamton.edu/p/people/shang/teach/605>.

Consider the following Bayesian hierarchical model:

$$\begin{aligned} X_i | \gamma_i, \mu_1, \mu_2, \sigma_1, \sigma_2 &\stackrel{iid}{\sim} \gamma_i N(\mu_1, \sigma_1^2) + (1 - \gamma_i) N(\mu_2, \sigma_2^2), \quad i = 1, \dots, n, \\ \gamma_1, \dots, \gamma_n | \theta &\stackrel{iid}{\sim} \text{Bernoulli}(\theta), \\ \theta &\sim \text{unif}(0, 1), \\ \mu_1, \mu_2 &\stackrel{iid}{\sim} N(0, 1), \\ \sigma_1^2, \sigma_2^2 &\stackrel{iid}{\sim} \text{IG}(1/2, 1/2). \end{aligned}$$

Answer the following questions:

1. Plot histogram (with a density curve) of the data. Hint: use

```
hist(X, prob=TRUE, col="grey")
lines(density(X), col="blue", lwd=2)
```

2. Based on the histogram, can you classify the observations?
3. Derive the full conditionals for **all** parameters.
4. Propose a Gibbs sampler for generating samples from the posterior distribution.
5. Use PSRF to monitor chain convergence.
6. Estimate $\mu_1, \mu_2, \sigma_1^2, \sigma_2^2$, and construct 90% credible intervals.
7. Classify the observations based on γ_i samples.
8. Find the proportion of the observations belonging to the first class, i.e., $N(\mu_1, \sigma_1^2)$.