Mixture Models and Bayesian Linear Regression

Fall 2022, MATH8050: Homework 8
Your Name, Section XXX

Due November 2, 12:00 PM

General instructions for homeworks: Please follow the uploading file instructions according to the syllabus. Each answer must be supported by written statements as well as any code used. Your code must be completely reproducible and must compile. For writing mathematical expressions in R Markdown, refer to the homework template posted on Canvas, a 30-minute tutorial, or LaTeX/Mathematics.

Advice: Start early on the homeworks and it is advised that you not wait until the last day. While the professor and the TA's check emails, they will be answered in the order they are received and last minute help will not be given.

No late homeworks will be accepted.

R Working Environment

Please load all the packages used in the following R chunk before the function `sessionInfo()`

```
# load packages


sessionInfo()
```

Total points on assignment: 10 (reproducibility) + 60 (Q1) + 30 (Q2)

Reproducibility component: 10 points.

1. (60pts total, equally weighted) 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 `Mixture.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.

(a) Let $\tau = 1/\varepsilon^2$ and $\phi_0 = 1/\sigma_0^2$. 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.

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

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

(c) 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 auxiliary random variables $\{Z_i\}_{i=1}^N$ that assign each observation to one of the mixture components with the probability of assignment being the respective mixture weight. Re-derive the full conditionals under the data augmentation scheme.

(d) In task (c) 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 "Mixture.csv".

(e) Given tasks (c)-(d), show traceplots for all estimated parameters, and compute means and 95% credible intervals for the marginal posterior distributions of all the parameters except the auxiliary variables. Now suppose you re-run the sampler using 3 different sets of starting values for the parameters, are your results the same? Justify your reasoning by with visualizations.

2. (30pts total, equally weighted) PH Exercise 9.1: The file `swim.dat` contains data on the amount of time in seconds, it takes each of four high school swimmers to swim 50 yards. Each swimmer has six times, taken on a biweekly basis.

(a) Perform the following data analysis for each swimmer separately: Write down a linear regression model of swimming time as the response and week as the explanatory variable. Complete the prior specification by using the information that competitive times for this age group generally range from 22 to 24 seconds.

(b) Implement a Gibbs sampler to fit each of the models. For each swimmer j, obtain a posterior predictive distribution for Y_j^*, the time of simmer j if they were to swim two weeks from the last recorded time.

(c) The coach has to decide which swimmer should compete in a swimming meet in two weeks. Use your posterior predictive distributions, compute $P(Y_j^* = \max\{Y_1^*, \ldots, Y_4^*\}|\mathbf{Y})$ for each swimmer j, and based on this make a recommendation to the coach.