

HW8

Neal Marquez

March 10th, 2019

As in previous homeworks, this question uses data from the 1975 U.S. Sustaining Effects Study of elementary education, available as the `egsingle` dataset in the `mlmRev` R package. This gives data on 1,721 students in 60 schools. We will take `math` (Mathematics achievement score) as the outcome variable, and `schoolid` (the code for the school the student attends) as the grouping variable. We will use only the data for year 0.5 (for which there are data on 1672 students).

Question 1

Estimate the Bayesian random effects one-way analysis of variance model for these data via Markov chain Monte Carlo using the prior distribution you specified in Homework 7.

Recall that we stated that if we would like for the data to inform the prior as much as possible we may set unit information priors such that $\nu_0 = 1, \eta_0 = 1, \kappa_0 = 1, \sigma_0^2 = s^2, \mu_0 = m^{-1} \sum_j^m \bar{y}_j, \tau_0^2 = \text{Var}(\bar{\mathbf{y}}_j)$. Here, μ_α refers to the mean of the school means, σ_α^2 refers to the between group variance, and σ_y^2 is the within group variance. Note that μ_0 and τ_0^2 refer to the mean of the school means and the variance of the school means respectively. Also note that these do not correspond to the MLE estimates of the parameters as each of the schools are weighted equally in the mean of means and variance of means calculations no matter how many students contributed to a schools calculation. We plot the stated priors in Figure 1. We then fit the model with STAN for MCMC inference which uses a No U-Turn Sampling (NUTS) algorithm.

Question 2

Assess the convergence of your algorithm and whether it has run for enough iterations. If not, run it for longer.

We plot the iterations of the chains for the parameters $\mu_\alpha, \sigma_\alpha^2$, and σ_y^2 . These parameters will often take longer to reach convergence than the school level effects and so convergence of these parameters often indicates that the model has run for a sufficient amount of iterations. We see that for the three parameters that there is a significant amount of overlap between the chains, as well as consistent parameter space exploration indicating that the model has run for a sufficient amount of time, Figure 2.

Question 3

Summarize the posterior distribution you obtain in graphical and tabular form.

Table 1 shows the estimates and Credible Intervals for the posterior distribution of our hyper parameters. Figure 3 shows the posterior Credible Intervals for the 60 school means arranged from highest to lowest mean effect.

Table 1: CI Estimates for Hyper-Parameters

Parameter	Estimate	2.5 CI	97.5 CI
b_Intercept	-0.3119609	-0.4480557	-0.1633180
sd_schoolid__Intercept	0.5041892	0.4047633	0.6334519
sigma	1.0979432	1.0605279	1.1362331

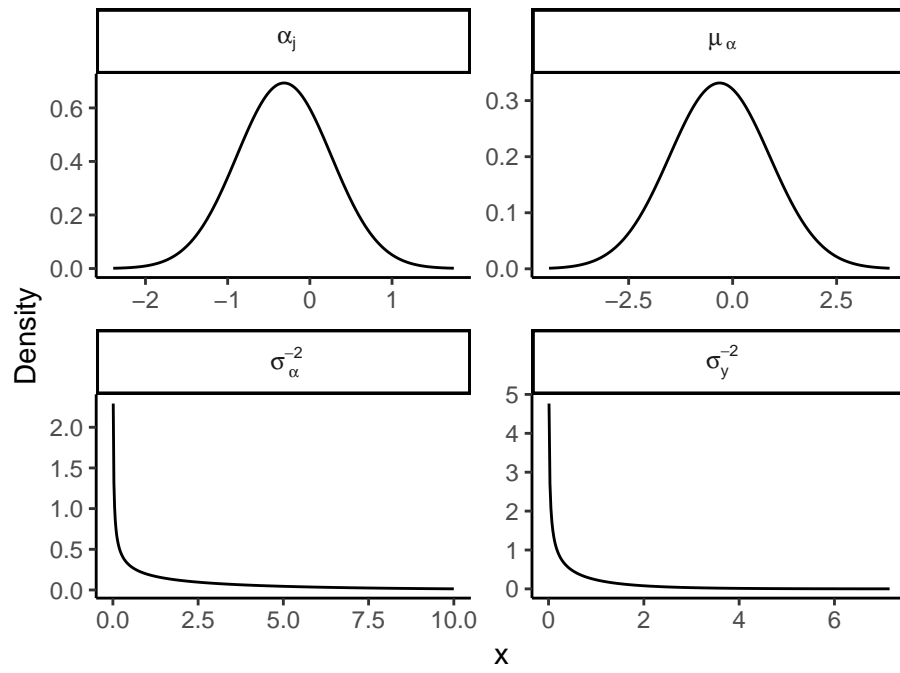


Figure 1: Prior Distributions of Parameters

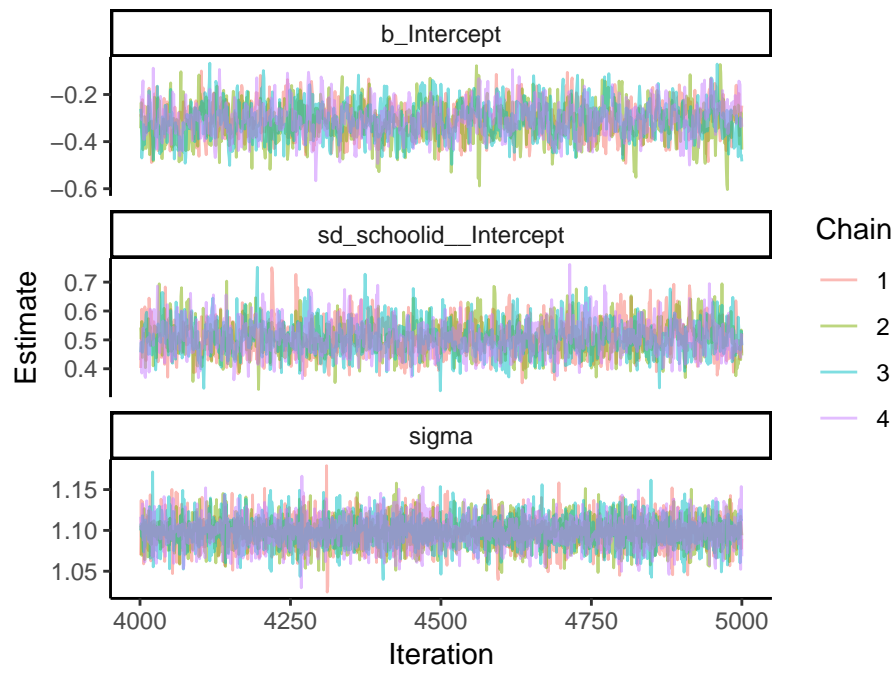


Figure 2: Chains for Hyper-Parameters

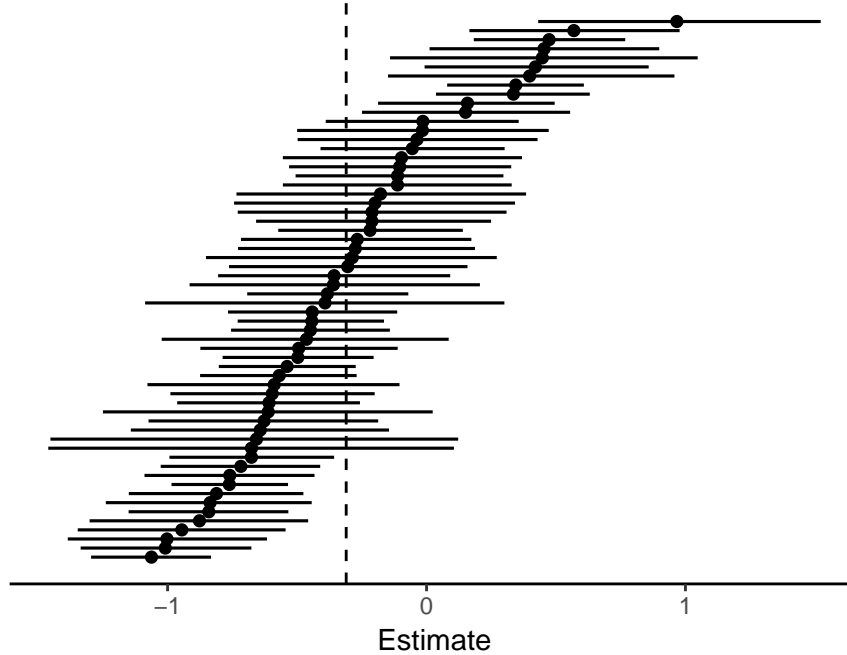


Figure 3: CI Estimates for School Means

Question 4

For which school is the difference between the non-Bayesian estimate and the Bayesian estimate of the group mean the largest? Why do you think this school has the largest difference?

When we run an MLE hierarchical model and compare the school mean estimates between that model and the Bayesian model we find that results are extremely similar. The priors that have been chosen allow the data to inform the estimate strongly and the pooling that occurs between the school means is similar to the pooling that happens in the MLE model. Table 2 shows the models with the “largest” difference in the estimate of the school means however the differences are on the order of the 10^{-3} . We do see however that these schools are on the lowest quartile of total observations of students allowing the priors to have more of an effect on the outcome.

Table 2: Sample Size and School Mean Estimate Difference

schoolid	Absolute Difference	n
4250	0.0040123	10
3170	0.0034612	12
2750	0.0032557	10
2850	0.0032088	10
3430	0.0022161	25