Second Homework Assignment

Bayesian Modeling

Due: Wednesday, April 2, 11:59pm

Assignment

In this assignment, you will work with an extended version of the sentencing data from session 6 and 7. The dataset now includes information on sentences by 118 judges over 5 years (`sentencing_data2`). You are going to rely on this dataset to investigate the probability of incarceration/percent incarcerated across judges and years using a Bayesian hierarchical regression model. The required dataset including all variables is available on Moodle from the homework assignment page. Read all the tasks thoroughly before starting on the assignment.

Task 1 (1.5 points):

Your outcome variable of interest y_i is a judge's sentence (`sentence`), which takes the value 1 if a defendant was incarcerated and the value 0 if a defendant was put on probation. Given this outcome, you want to specify a varying-intercepts model with intercepts varying across (1) judges (`j_id`) and (2) years (`year`). Additionally, you want to add varying intercepts that capture (3) the judge-year-specific probability of incarceration/percentage incarcerated. Write this model down formally, fully specifying the likelihood and all priors. Is this a nested or non-nested model? Explain your decision. Upload the answers to this task in a .pdf file.

Task 2 (1.5 points):

In R, import the `sentencing_data2` dataset. Process the dataset in such a way that you can use if for full Bayesian inference with Stan and assemble the data required to fit the varying-intercepts model specified in Task 1. How many varying intercept parameters are we estimating in total?

Task 3 (4 points):

In R, compile the data in the appropriate format for passing it to cmdstanr's sample method and specify the full Stan program that corresponds to your varying-intercepts model and

the data you prepared. Next, store the Stan file on your computer and compile your Stan program using the cmdstanr R package. Given the Stan program and data you prepared, sample from the posterior. Estimates should be directly stored on your computer. Do not use convenience R packages, such as brms or rstanarm.

Task 4 (3 point):

In R, choose 5 judges from the data for which we have observations on all five years. Using the posterior estimates for varying intercept parameters, compute the probability of incarceration/percent incarcerated for each of these 5 judges in every year. Visualize your posterior predictions.

Optional bonus task (1 point):

Alter your Stan program to incorporate additional intercepts varying across courts (`c_id`).

Submission:

Submit your R script, commented throughout, the .pdf file and other files on Moodle by the deadline given above.