First Homework Assignment

Bayesian Modeling

Due: Wednesday, March 19, 11:59pm

Context

In his study on "The Politics of Respectability and Black Americans' Punitive Attitudes" Hakeem Jefferson seeks to explain Black Americans' support for racialized punitive social policies through individuals' adherence to respectability. The concept of respectability refers to the attitude among Blacks that negatively stereotyped Black behavior and attitudes must be avoided or overcome in favor of dominant norms and manners to reach racial equality. To measure respectability, Jefferson combines responses to several survey questions that are supposed to tap the concept in a scale—the RSP score. Higher values on this score indicate a stronger embrace of respectability politics.

Assignment

You are going to rely on Jefferson's data (`punitive_attitudes.RData`) to investigate how support for sagging pants ordinances among Black Americans differs along survey respondents' RSP score, sex, and age group. Sagging pants ordinances reflect a particular, discriminatory type of local policy aimed at regulating the public presentation and dress of residents. For this purpose, you will specify and estimate a Bayesian linear regression model of the following form:

$$
\begin{aligned}
y_i &\sim \text{Normal}(X_i\beta, \sigma) \\
\sigma &\sim \text{Log-Normal}(0, 1) \\
\beta &\sim \text{Normal}(0, 1),
\end{aligned}
$$

where the outcome y_i is the support for sagging pants policies (`sagging_pants`), X_i represents the covariate matrix including an intercept column of $1s$, respondents' RSP score (`respectability`), sex (`sex`), and age group (`age`). β is the corresponding coefficient vector and σ is the noise term. 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 (3 points):

In R, import the `punitive_attitudes.RData` dataset, process the dataset in such a way that you can use if for full Bayesian inference with Stan, and compile the data in the appropriate format for passing it to cmdstanr's sample method. Use whatever R functions and packages you deem appropriate for this task.

Task 2 (4 points):

In R, specify the full Stan program that corresponds to the above Bayesian linear regression model and the data you prepared. Following the above notation, specify the program using matrix notation. Store the Stan file on your computer. Do not use convenience R packages, such as brms.

Task 3 (2 points):

In R, compile your Stan program using the cmdstanr R package. Next, given the Stan program and data you prepared, sample from the posterior. In doing so, specify the seed `1457L`, use 4 chains, run in parallel, with 1000 warmup and 500 sampling iterations. Estimates should be directly stored on your computer. Do not use convencience packages, such as brms.

Task 4 (1 point):

In R, print a summary of the posterior samples. Then, make use of the custom function to import only the posterior for the coefficient β related to the RSP score. Visualize the full posterior distribution of the β coefficient. In addition to the custom function for importing posterior files, you can use whatever R functions and packages you deem appropriate for this task.

Optional bonus task (1 point):

Alter your Stan program to use the generated quantities block in order to conduct a prior predictive check with priors specified as above. Having run the prior predictive check, are the original priors for β and σ sensible in this context?

Submission:

Submit your R script, commented throughout, on Moodle by the deadline given above.