Bayesian Final Project Proposal

Joe Stoica, Ethan Violette, Qinzhe Wang

3/31/2019

What is the topic of the project?

The topic is predicting political orientation from two different regions of the brain. The goal of the project is to see if isolating the amygdala and anterior cingulate cortex results in measurable differences in an individual's tendency towards conservatism or liberalism.

The amygdala is a roughly almond-shaped mass of gray matter inside each cerebral hemisphere, involved with the experiencing of emotions. The Anterior Cingulate Cortex is a part of the brain located at the middle of the frontal lobe. It is located at the front of the corpus callosum which connects the left and right hemispheres of the brain. It is responsible for many psychological and physiological functions, including attention allocation, reward anticipation, decision-making, ethics and morality, impulse control (e.g. performance monitoring and error detection), and emotion. Our hypothesis is that these two regions of the brain are involved in political orientation.

What are the main issues or problems to be addressed?

The main issues/problems to be addressed for this project is completing the four steps for Bayesian Analysis, as described below, within the context of predicting political orientation:

1. Specify a joint distribution for orientation and all the unknowns, which typically takes the form of a marginal prior distribution for the unknowns multiplied by a likelihood for the outcome(s) conditional on the unknowns. This joint distribution is proportional to a posterior distribution of the unknowns conditional on the observed data
2. Draw from posterior distribution for orientation using Markov Chain Monte Carlo (MCMC).
3. Evaluate how well the model fits the data and possibly revise the model.
4. Draw from the posterior predictive distribution of the orientation given interesting values of the ACC and amygdala in order to visualize how a manipulation of one of these predictors affects a function of orientation.

What data will be used, including the variables measured?

The data set n90 pol.csv, taken from Shalizi's Advanced Data Analysis From an Elementary Point of View, contains information on 90 university students who participated in a psychological experiment designed to look for relationships between the size of different regions of the brain and political views. The variables amygdala and acc indicate the volume of two particular brain regions known to be involved in emotions and decision-making, the amygdala and the anterior cingulate cortex; more exactly, these are residuals from the predicted volume, after adjusting for height, sex, and similar anatomical variables. The variable orientation gives the subjects' locations on a five-point scale from 1 (very conservative) to 5 (very liberal). orientation is an ordinal but not a metric variable, so scores of 1 and 2 are not necessarily as far apart as scores of 2 and 3. As of right now, we're planning on encoding 1-2 as "conservative" and 3-5 as "liberal" to perform logistic regression on the dara.

What are your initial thoughts on appropriate models/distributions?

Our initial intuition is to use the stan_glm function from the rstanarm package, which essentially can be used as a logistic regression model with specified priors to sample from a posterior distribution using MCMC. As for the prior distributions, since we are unconfident that the coefficients for amygdala and ACC are close to 0, we'll probably use a heavier-tailed distribution like a Student t distribution with something like 7 degrees of freedom as the prior for our predictor variables. For logistic regression, articles online have pointed towards using a hierarchical shrinkage, or "horseshoe" prior for the response variable, so we will investigate this further when deciding a prior specification for orientation.

What questions and/or concerns do you have about the project?

The primary concern we have is that we haven't learned how to conduct logistic regression from a Bayesian perspective. Though it seems pretty straightforward to conduct the analysis in R, our lack of experience in this area may cause some small issues.