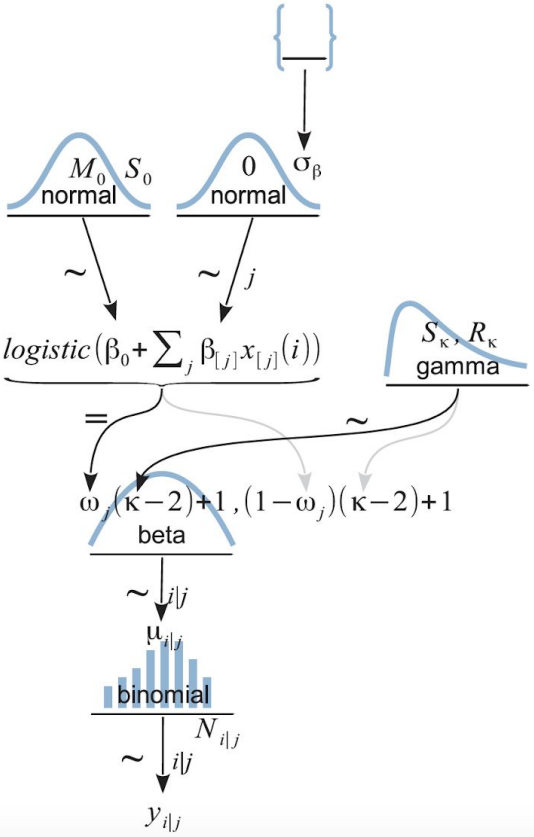
Bayesian Logistic Regression

We are interested in estimating the probability of getting a Republican vote based on the income category of the state. Income category is a nominal predictor (poor, middle, mid-upper, rich).

Figure 1. Hierarchical diagram for logistic ANOVA-like model



The top part of the structure is based on ANOVA-like model. It shows that the modal ability of income category j, denoted ωj, is a logistic function of a baseline ability β0 plus a deflection for the income category, β[j]. The baseline and deflection parameters are given the usual priors for an ANOVA-style model.

The category income j, denoted ωj, is the mode of a beta distribution that describes the spread of individual votes within the income category. The concentration of the beta distribution, denoted κ, is estimated and given a gamma prior.

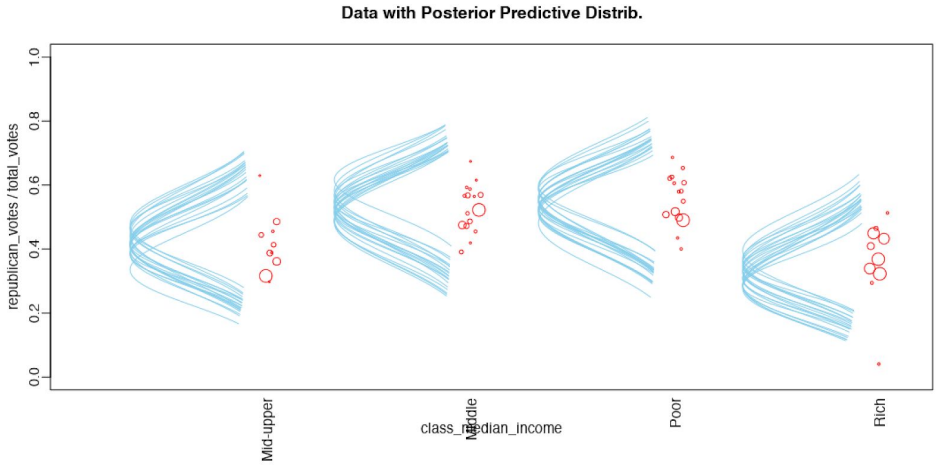
The vote of each state i in position j is denoted μi|j. The number of Republican votes by state i, out of Ni|j total votes, is denoted yi|j, and is assumed to be a random draw from a binomial distribution with mean μi|j.

The binomial distribution is used as a computational shortcut for multiple independent dichotomous events. In the model, y=1 indicates Republican vote and y=0 indicates Democratic vote.

Results

The present model focuses on descriptions at the group level, not at the individual level. Therefore the program has contrasts between the group income categories, not between states.

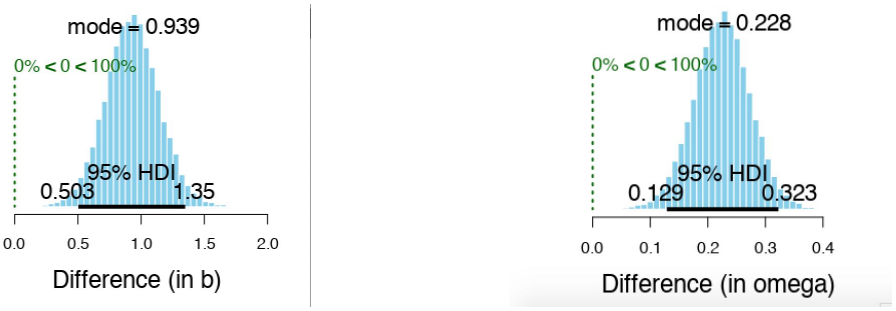
Figure 2. Hierarchical diagram for logistic ANOVA-like model



Each state's ratio of Republican votes to total votes is shown as a dot in Figure 2. The dot size indicates the number of total votes. The posterior predictive beta distribution is superimposed along the side of each group's data. The beta distributions have their tails clipped so they span 95%. The graphs are wide, indicating that the estimate of the group modes is uncertain. One possible reason is that the data set is small. Nonetheless, the graph confirms that the ratio of Republican votes/ Total votes is larger for Middle and Poor income categories.

The posterior predictive distributions are credible beta distributions assuming homogeneous concentration across positions. Figure 3 shows the contrasts for the log odds deflection parameters (βj or b) and the group modes (ω or omega) of Poor-Rich corroborating that the ratio of Republican votes/ Total votes is larger for Poor income categories with respect to Rich.

Figure 3. Posterior predictive distributions



Reference: Kruschke, John K. (2015). Doing Bayesian Data Analysis: A tutorial with R, JAGS, and Stan (2nd ed.). London: Elsevier.