Hierarchical modeling

The US can be split up into four main regions, the east coast, the south the Midwest and west.

The complex historical reasons that I won’t go into here the states and within these regions are more similar to one another than states in other regions.

To give an example, the south tends to be more religious than the rest of the country and religious people tend to vote more conservatively.

So we can the fact that each district is in a state and each state is in a region to model hierarchically.

Model Assumptions

Since we are dealing with a binary outcome we believe that a logistic regression is best suited for modeling

We also assume that for our hierarchical components that that each district is exchangeable within a state and that each state is exchangeable within it reason.

Finally since most elections also tend to be quite close, we assume that given no other information the theta in our logistical regression is 0.5.

We mention this because in all of our models we choose not the have intercept vary by state or region, meaning our baseline theta is the same across all districts.

We have four models all with 4 chains and 2000 iterations the first model has many covariates, with two level non nest hierarchy the second model has nested hierarchy, the third model is using a state only hierarchy and the fourth model has only regional hierarchy

Model 4

M4 Equations

Here we choose to model the probability of voting democrat with a only two covariate, urban index and pct centage retirees.

As we already talked about in the assumptions we assume that the baseline is a probability of roughly 0.5 so we choose the intercept prior to be centered at zero with a small standard deviation, putting more importance on the other two covariates.

We then also choose a non centered parameterization for urbanization where the slope varies by region.

As with the other models we think that urbanization is positive and has large variability by states which is why we choose a normal and then a half Cauchy with k=10 for the sigma parameter.

Similarly as before we believed that a higher percentage of retirees would have a negative correlation with probability of voting democratic, so choose a student t prior with a center at -2.

We choose student t over normal to represent our increased uncertainty, and a general desire for fatter tails.

M4 Trace Plots

As we can see in the trace plots we didn’t have any major divergences and seemed to have pretty good estimation.

Next we can see the posterior distribution of urban index, which looks roughly normal centered around 1.25.

M4 Coeff

Moving on to the coefficient summary we see that the estimates for the variables of interest are relatively similar to models two and three and both the rhat and large ESS show that this model was able to coverge properly.

RMSE

Moving on the model comparison, how did the models compare to one another?

Within a logistic regression framework, root mean square error is a bit more difficult to interpret since we are predict a probability rather than an outcome.

However in ideal model for a district that voted democrated we would want theta as close as possible to 1 and vice versa.

We also know that our RMSE is bounded between 0 and 1.

Then looking at the distributions of the RMSE for different models, we see that two and three look pretty similar and model 1 performed slightly better,

however when looking at the scale of model 4 we see it that the majority of the error begins where it stops in other models indicating poor performance.

Log Likelihood

Here we see the distribution of the log likelihood for observing our particular data given our model. What we want to see here are larger values, in all plots we see that there might a few outliers, however we can see that while model 1 2 and 3 look almost identical, model 4 has more mass in small values, again indicating a worse fit as compared to the other models.

Best Model

So in the end the posterior probability looks like this makes sense marginally, because we essentially have two classes that we want to predict, one where we would expect that theta is one and the other where we expect that that theta is zero.

We like this shape because it shows that our model was able to differentiate the two different models.

It also makes sense that they have roughly the same amount of density on both sides since our observed data was roughly evenly split among the two classes.

Also all the model posterior distributions looked nearly identical.