MBA5011 Multivariate Analysis: Model-Based Statistics 2020 Spring

Assignment 4: assessment percentage with 20.83%; deadline: 11:59pm, 2020-05-24 (Sunday)

Question 1(45%)

Consider again the data(rugged) data on economic development and terrain ruggedness, examined in this chapter. One of the African countries in that example, Seychelles, is far outside the cloud of other nations, being a rare country with both relatively high GDP and high ruggedness. Seychelles is also unusual, in that it is a group of islands far from the coast of mainland Africa, and its main economic activity is tourism.

One might suspect that this one nation is exerting a strong influence on the conclusions. In this problem, I want you to drop Seychelles from the data and re-evaluate the hypothesis that the relationship of African economies with ruggedness is different from that on other continents.

(a) Begin by using Stan to fit just the interaction model:

$$y_i \sim \text{Normal}(\mu_i, \sigma)$$
$$\mu_i = \alpha + \beta_A A_i + \beta_R R_i + \beta_{AR} A_i R_i$$

where y is log GDP per capita in the year 2000 (log of rgdppc_2000); A is cont_africa, the dummy variable for being an African nation; and R is the variable rugged. Choose your own priors. Compare the inference from this model fit to the data without Seychelles to the same model fit to the full data. Does it still seem like the effect of ruggedness depends upon continent? How much has the expected relationship changed? (10%)

(b) Now plot the predictions of the interaction model, with and without Seychelles. Does it still seem like the effect of ruggedness depends upon continent? How much has the expected relationship changed? (5%)

(c) Finally, conduct a model comparison analysis, using WAIC. Fit three models to the data without Seychelles using Stan: (20%)

$$\begin{aligned}
\text{Model 1} : y_i &\sim \text{Normal}(\mu_i, \sigma)\\
\mu_i &= \alpha + \beta_R R_i\\
\text{Model 2} : y_i &\sim \text{Normal}(\mu_i, \sigma)\\
\mu_i &= \alpha + \beta_A A_i + \beta_R R_i\\
\text{Model 3} : y_i &\sim \text{Normal}(\mu_i, \sigma)\\
\mu_i &= \alpha + \beta_A A_i + \beta_R R_i + \beta_{AR} A_i R_i
\end{aligned}$$

(d) Using the "**Metropolis Hastings**" algorithm with R to fit Model 3 in (b), see Core Statistics sec. 6.2.5 and the MCMC_code.R under topic 6.(10%)

Question 2(15%)

Run the model below and then inspect the posterior distribution and explain what it is doing. (This is a very simple one, no trap.)

```
data{
    int<lower=1> N;
}
parameters{
    real a;
    real b;
}
model{
    b ~ cauchy( 0 , 1 );
    a ~ normal( 0 , 1 );
}
```

Question 3(40%)

The data contained in library(MASS);data(eagles) are records of salmon pirating at- tempts by Bald Eagles in Washington State. See ?eagles for details. While one eagle feeds, some- times another will swoop in and try to steal the salmon from it. Call the feeding eagle the "victim" and the thief the "pirate." Use the available data to build a binomial GLM of successful pirating attempts.

(a) Consider the following model:

$$\begin{aligned} y_i &\sim \text{Binomial}(n_i, p_i) \\ \log \frac{p_i}{1 - p_i} &= \alpha + \beta_P P_i + \beta_V V_i + \beta_A A_i \\ \alpha &\sim \text{Normal}(0, 10) \\ \beta_P &\sim \text{Normal}(0, 5) \\ \beta_V &\sim \text{Normal}(0, 5) \\ \beta_A &\sim \text{Normal}(0, 5) \end{aligned}$$

where y is the number of successful attempts, n is the total number of attempts, P is a dummy variable indicating whether or not the pirate had large body size, V is a dummy variable indicating whether or not the victim had large body size, and finally A is a dummy variable indicating whether or not the pirate was an adult. Fit the model above to the eagles data, using Stan. (15%)

(b) Now interpret the estimates. Plot the posterior predictions. Compute and display both (1) the predicted probability of success and its 89% interval for each row (i) in the data, as well as (2) the predicted success count and its 89% interval. What different information does each type of posterior prediction provide? (15%)

(c) Now try to improve the model. Consider an interaction between the pirate's size and age (immature or adult). Compare this model to the previous one, using WAIC. Interpret. (10%)