Bayes Assignment 5 of 2025

Sean van der Merwe

Instructions

Goal: To implement a robust Bayesian regression model on a dataset that has extensive missing values in multiple explanatory variables, using multiple imputation as an intermediate step.

The observations are (non-fasting) blood glucose measurements taken recently. You should look up the meaning of such measurements before interpreting the results. The primary explanatory variable is a previous measurement, usually taken roughly a year prior, sometimes 2 years.

1. Isolate your assigned data set and visualise it via plots. Identify the key properties, without jumping to conclusions.
2. Use an SRMI style multiple imputation approach to create 10 completed copies with all missing values imputed. Either do this using a tool such as MICE, or your own chained pseudo-Gibbs Bayesian predictions, or similar. **Ensure that the dependent variable and previous measurement are separated out first. Do not include them at all in the imputation process!**
3. Calculate the between imputation standard deviation of the mean of any one continuous variable for which imputations were required.
4. Apply a robust Bayesian regression on the conditional median of your dependent variable (given your chosen explanatory variables) by using a Laplace (double-exponential) distribution. The suggested log prior is $\log\pi(\sigma) = -\log\sigma + c_1$. **You must apply this model on 1 of the completed data sets only.**
5. Adjust your model to accommodate censoring at the identified censoring point. Apply this model first to the same single data set as above and calculate the probability that each coefficient has changed by more than 2%.
6. Apply the robust censoring model to all 10 completed data sets and store the sets of parameter simulations.
7. Concatenate the sets of parameter simulations to achieve one large set of simulations. Summarise those simulations. **Give the posterior mean and 95% credibility interval for each coefficient at least.**
8. Interpret the coefficients of the combined fit for the conditional median in detail, paying careful attention to what is actually being explained by the coefficients. *Hint: consider that the coefficients affect the previous measurements*.
9. Select two people in your data set, each having exactly one missing value, and tell me a story about their past - how did their observed values come to be? Adjust the values of their explanatory variables slightly and share a narrative about those changes. *For example, you could replace their previous measurement with their current measurement, increase their age by a year, tweak their BMI, etc*. Lastly, give their posterior predictive distributions based on the new explanatory values and all simulations, then describe their possible future journeys **based on your predictions** from their data.

Bonus marks for clever handling of the censoring of the prior year measurements.