The New School for Social Research
Advanced Econometrics 1
Fall 2017
Christian Schoder
Jangho Yang

Assignment 5

Due Oct 30 (Mon) 6:00 pm

1. Understanding Rstan (BDA3)

   (a) Load the attached data **light.csv**. The data records the amount of time required for light to travel a distance of 7442 meters (Simon Newcomb, 1882). The data are recorded as deviations from 24,800 nanoseconds.

   (b) Fit a normal model to the data using Stan. Use a non-informative prior for parameters μ and σ. Set up **generated quantities** block in your code in the way that you can get 20 posterior predictive distributions using μ and σ.

   (c) Use simulations from the posterior predictive distributions and check two test statistics 1)mean and 2)the smallest value. Visualize the result by plotting a histogram of each simulated posterior predictive distribution along with the mean and the smallest value of "data". Plot all 20 histograms in the same panel. Check if the posterior predictive distributions predict the mean and the smallest value of the fake data well.

2. Multi Linear Regression (Gelman, Hill, Problem 3 from Chapter 4): This data (link) contains three columns. The first column is children's test scores at age 3, the second is mother's education level (1 = no high school, 2 = high school grad, 3 = some college, and 4 = college grad), and the third column is mother's age at the time she gave birth for a sample of 400 children. Install **foreign** package to load the data and type `data <-read.dta("child.iq.dta")`

   (a) Fit a regression of child test scores on mother's age. Report the posterior means and variances of parameters. When do you recommend mothers should give birth? What are you assuming in making these recommendations?

   (b) Repeat this for a regression that further includes mother's education, interpreting both slope coefficients in this model. Have your conclusions about the timing of birth changed? Why do you think the coefficient of the existing predictor has changed?