Week 5 – Bayesian Inference Exercises

Instructions: This exercise requires an Internet connection and a browser unless you have JAGS software installed on your computer and the BEST package loaded in R (as described in Chapter 5). Use the website http://www.sumsar.net/best_online/ to run the BEST procedure on data from mtcars. Type or cut and paste these data into the boxes on the web page:

Automatic:
21.4 18.7 18.1 14.3 24.4 22.8 19.2 17.8 16.4 17.3 15.2 10.4 10.4 14.7 21.5 15.5 15.2 13.3 19.2

Manual:
21.0 21.0 22.8 32.4 30.4 33.9 27.3 26.0 30.4 15.8 19.7 15.0 21.4

If you want to accomplish the same thing in R, use the following code:

```
library(BEST)
bestOut <- BESTmcmc(mtcars$mpg[mtcars$am==0], mtcars$mpg[mtcars$am==1])
bestOut
plot(bestOut)
```

1. Using R, display a boxplot that depicts the distributions of automatic and manual transmissions.

2. Write R code calculating the value of the "point estimate" of the mean difference in mpg between the two types of transmissions.

3. After running the BEST procedure, either on the website or within R (as described above), write a comment indicating the lower limit and upper limit of the highest density interval (HDI).

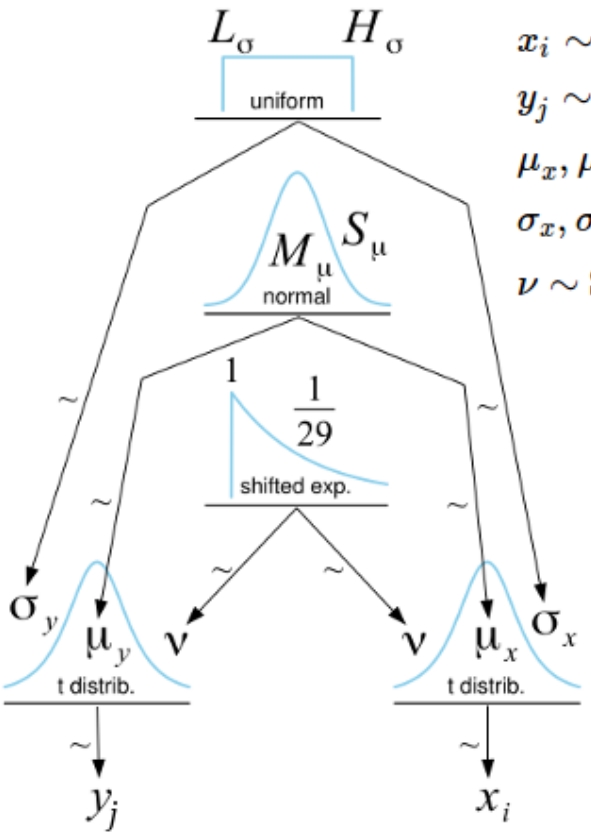
4. Use a comment to describe what an HDI is? Hint: what can you say about the probability that the population value for the mean difference lies between the upper and lower limits of the HDI?

5. Does the HDI overlap with zero? What percentage of the estimates under the HDI curve is above zero? What implications does this have for the credibility of a non-zero difference in mpg between automatic and manual transmissions?

6. The web page output shows "Density – SD Group 1" and "Density – SD Group 2" plots. You can also view these plots in R-Studio with plotAll(bestOut). Which transmission group is likely to have greater variability in the population. Can you explain this?

7. Based on the results of the Bayesian t-test, do you prefer automatic transmissions or manual transmissions? Assume for a moment that a typical owner puts a total of about 66,000 miles on a car before selling or trading it and that average gas costs (inflation adjusted) are $2.60 per gallon. What would you say to automotive designers or buyers?

The next exercise requires that the JAGS software is installed on your computer. Go to http://mcmc-jags.sourceforge.net if you have not yet installed JAGS. You will know if you have JAGS successfully installed if you can library() the rjags package without generating an error. In addition to this handout, make sure to download from Blackboard the file JagsBreakoutExercise.R, a file with R code. The R code includes a long string variable that contains the code for the model that we pass to JAGS. Examine this graphic before you proceed with the exercise:



$$x_i \sim \mathrm{t}(\mu_x, \sigma_x, \nu)$$
$$y_j \sim \mathrm{t}(\mu_y, \sigma_y, \nu)$$
$$\mu_x, \mu_y \sim \mathrm{Normal}(M_\mu, S_\mu)$$
$$\sigma_x, \sigma_y \sim \mathrm{Uniform}(L_\sigma, H_\sigma)$$
$$\nu \sim \mathrm{ShiftedExp}(\tfrac{1}{29}, \mathrm{shift} = 1)$$

This is a so-called Krushcke plot, that graphically displays the contents of a JAGS model. This image was developed by Rasmus Baath. Working from the bottom up, posterior predicted values of x and y are modeled as a t-distribution. The t-distribution has a value nu to model the degrees of freedom as well as mu and sigma values for x and y. Priors for nu are modeled as an exponential distribution. Priors for the mus are modeled as a normal distribution. Priors for the sigmas are modeled as a uniform distribution. Instead of directly modeling sigmas, the JAGS code tracks "tau" values. Tau is the inverse of the variance.

8. Review the JAGS code in JagsBreakoutExercise.R and match the various elements of the JAGS code to the model displayed above.

9. Open the R code file JagsBreakout.R in R-Studio and review the code. Make sure that your setwd( ) call points to the folder where you have stored the jags model file.

10. Step through the R code one line at a time, making note of any questions that arise. As you create the various objects needed to run the JAGS model, feel free to inspect them with str() or View( ) so that you understand what is happening.

11. Where do you find the HDI for the mean differences in mpg? Does the HDI overlap with zero? What percentage of the estimates for samples[,"mu_diff"] are above zero? What implications does this percentage have for the credibility of a non-zero difference in mpg between automatic and manual transmissions?

12. Once you reach the end of the R code and have reviewed the plots and the summary of the JAGS model run, add some code to conduct an independent samples t-test on the original mpg data. Compare the results of the t-test to the results of the JAGS model.

13. Time permitting, find another data set with at least two groups (such as PlantGrowth) and repeat the process of running this JAGS model *and comparing the results to the output of the frequentist t-test*.

14. Time permitting, create two random samples of data using the same distribution. You can use any non-normal distribution you like (e.g., by using rexp(), rpois(), etc.). Run the JAGS models and the t-test on these new data and interpret the results.