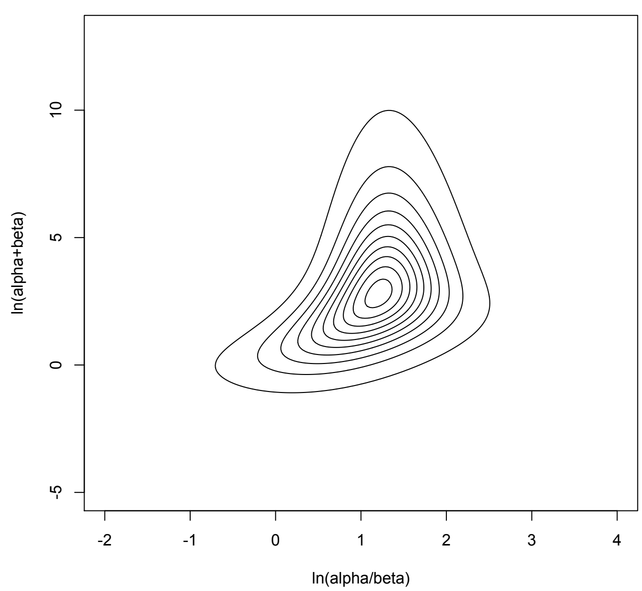
Assignment 10 (2 pages)

1. Consider the data presented in Table 7.3

   (a) Let θ_1, θ_2, and θ_3 represent the proportion of houses that have basements in Blue Earth, Clay, and Goodhue counties respectively. Fit a hierarchical model the data, and use it to obtain posterior summaries for θ_1, θ_2, and θ_3. For the hyperprior, we will use the same one as in the rats example, that is uniform on $(\alpha/\alpha+\beta, (\alpha+\beta)^{-1/2})$. Using the grid method from chapter 5, we identify a grid $(\ln(\frac{\alpha}{\beta}), \ln(\alpha+\beta)) \in [-2,-4] \times [-5,13]$ (See Figure)
   As in the rats question from the last assignment, convert these into uniform hyperprior distributions for U and V where $U = \frac{\alpha}{\alpha+\beta}$ and $V = (\alpha+\beta)^{-1/2}$ with appropriate ranges.



   (b) Fit a linear regression to the natural log of the radon measurements, with indicator variables for the three counties and for whether a measurement was recorded on the first floor or basement, do not include an intercept term. Present posterior summaries for parameters and summarize your posterior inferences in non-technical terms (hint, for each parameter β, what does e^{β} represent?).

(c) Suppose another house is sampled at random from Clay County, simulate values from the posterior predictive distribution for its radon measurements anf give an 95% predictive interval. Express the interval of the original unlogged scale. (Hint: You must consider whether or not the randomly chosen house has a basement)

2. The file `drinks.txt` contains the amount of time needed by a company employee to refill an automatic vending machine. For each refill, the number of cases of product and the distance walked (in feet) is also recorded.

   (a) Fit a linear regression model for the time taken, with number of cases and distance walked as explanatory variables (include an intercept term).

   (b) In Classical Statistics, one way the quality of a regression model can be analyzed is by calculating something called the $AdjustedR^2$ value, it is basically the proportion of variation in the response variable that is explained by the explanatory variables, adjusted for the number of variables. Obviously, the closer this number is to 1, the better. The Bayesian equivalent R_B^2 is defined as

   $$R_B^2 = 1 - \frac{\sigma^2}{s_Y^2}$$

   where σ^2 is the variance of the regression model and s_Y^2 is the sample variance of the response variable data. Note that since in the Bayesian framework σ^2 is a random variable, so is R_B^2. Obtain a 95% credible interval for R_B^2. Does it seem like the model is a good for the data?

   (c) Obtain a 95% predictive interval for how long it would take to restock the vending machine if the number of cases and distance were at their average (mean) values.

3. Revisit the data in Question 1. Fit a Two-Way ANOVA to the natural log of the radon measurements. For each county in separately, construct a 95% credible interval for the difference between the average (unlogged) radon measurement in houses with basements and in houses without basements. Test the hypothesis that the average radon measurement in houses with basements is greater than in houses without basements.