PBHS 43010: Homework #3 Bayesian hierarchical models and model selection

1 (10pts) Poisson/Gamma model

a) Suppose that x_i is number of snow falls in Chicago in year i, $i = 1, 2., ..., n = 10$. Assume $x_i \mid \lambda \sim Poisson(\lambda)$. Find the conjugate prior distribution of λ, and the posterior distribution in closed forms.

b) Suppose the data are $x = (5, 4, 6, 5, 5, 6, 5, 7, 3, 6)$, find the posterior mean and variance of λ assuming the conjugate prior in a) with prior mean 5 and variance 5.

2 (10pts) Let $x_1, ..., x_n$ be i.i.d. data from a normal distribution with known mean 0 and unknown variance τ.

a) Write the likelihood $L(\tau|\mathbf{x})$, where $\mathbf{x} = (x_1, \ldots, x_n)$.

b) Derive the Jeffreys prior for τ.

c) Suppose we observe the 6 data values $x_1 = 2.72, x_2 = 1.65, x_3 = 0.44, x_4 = -1.62, x_5 = 0.27, x_6 = -2.22$. Write the posterior distribution, using your Jeffreys prior from part (b). Do you recognize the form of this posterior? Specify exactly what distribution it is, including the parameter values.

3. (15pts) The "California rain data set" contains data for 30 California cities. Read in the data set by (don't copy paste the command in R, as the breaks will create problems).

```
rain.data<-read.table("https://ccte.uchicago.edu/Bayes2017/Homework/calirain.txt",
header=F, col.names = c("number", "city", "precip",
 "altitude", "latitude","distance")); attach(rain.data)
```

will read in the data correctly. The variables are: x_1 = 'altitude', x_2 = 'latitude', x_3 = 'distance from coast', y = 'annual precipitation (in inches)'. Use "precipitation" as the response variable in the regression model.

(a) Adapt the model selection code on the course web page to perform a Bayesian model selection based on response variable Y and candidate predictor variables X_1, X_2, X_3. Which model or models appear best based on their posterior probabilities?

(b) Fit your "best" model (with noninformative priors on β and σ^2) using a Bayesian approach and write the estimated linear regression function for predicting precipitation. Use the posterior mean of β and σ^2 for the estimated linear regression.

(c) Now consider interaction (cross-product) terms X_1X_2, X_1X_3, X_2X_3 as other candidate predictors. Perform a Bayesian model selection using all six candidate predictors (first-order and interaction terms), using the convention that no interaction term should appear in the model without each of its component variables appearing as first-order terms. Does the "best" model change from the one chosen in part (a)? Explain.

4. (15pts) Suppose $X \mid \theta \sim N(\theta, \sigma^2)$ with unknown θ but known σ^2. We want to test $H_0 : \theta = 0$ vs. $H_1 : \theta \sim N(0, \Sigma^2)$. Assume we observe $x = \lambda \cdot \sigma$ for some positive λ, derive the closed-form solution of BF_{01}? Which model is favored if $\lambda \gg 1$, or if $\lambda \approx 1$ and $\sigma/\Sigma \gg 1$?