Homework 2

(Your name here)

Due Sunday 10/16

Theory problem:

Suppose $y \mid \lambda \sim Poisson(\lambda)$. Find Jeffreys' prior density for λ, and then find values of α and β for which a proper Gamma(α, β) prior density is a reasonably close match to Jeffreys' prior.

Computing problem:

In the bioassay example of of BDA Section 3.7, replace the uniform prior density by a joint normal prior distribution on (α, β), with $\alpha \sim \text{Normal}(0, 2^2)$, $\beta \sim \text{Normal}(10, 10^2)$, and corr$(\alpha, \beta)$=0.5.

a) Repeat all the computations and plots of Section 3.7 with this new prior distribution.
b) Check that your contour plot and scatterplot look like a compromise between the prior distribution and the likelihood (as displayed in Figure 3.3).
c) Discuss (in a brief sentence) the effect of this hypothetical prior information on the conclusions in the applied context.

Applied problems:

Chicken Brains

An experiment was performed on the effects of magnetic fields on the flow of calcium out of chicken brains. Two groups of chickens were involved: a control group of 32 chickens and an exposed group of 36 chickens. The chickens were killed, their brains removed, and a measurement was taken on each chicken brain. The purpose of the experiment was to measure the average flow μ_c in untreated (control) chickens and the average flow μ_t in treated chickens. The 32 measurements on the control group had a sample mean of 1.013 and a sample standard deviation of 0.24. The 36 measurements on the treatment group had a sample mean of 1.173 and a sample standard deviation of 0.20.

a) Assuming the control measurements were taken at random from a normal distribution with mean μ_c and variance σ_c^2, what is the marginal posterior distribution of μ_c? Similarly, use the treatment group measurements to determine the marginal posterior distribution of μ_t. Assume a uniform prior distribution on $(\mu_c, \mu_t, log\sigma_c, log\sigma_t)$.

b) Generate a samples to approxmiate the posterior distribution of $\mu_t - \mu_c$[1]? To get this, you should sample from the independent distributions you obtained in part (a) above. Plot a histogram of your samples and give an approximate 95% posterior interval for $\mu_t - \mu_c$. Does it appear as if there is an effect of magnetic fields on the flow of calcium in chicken brains?

c) In a follow up experiment, each chicken brain was split into two. For each chicken, one of the brain halves was assigned to the magnetic treatment and the other half to the control. There were 32 chickens in the experiment and the sample mean in the control group was 0.987 and in the treatment group was 1.168. The sample standard deviation of the control group was 0.21 and in the treated group was 0.23. The sample correlation between the observations was 0.44. Use the conjugate inverse-Wishart family of prior distributions for the multivariate normal introduced in Section 3.6. As noted, the marginal posterior distribution of μ is a multivariate t distribution (See BDA Appendix A for how to sample from this distribution). Sample from the marginal posterior distribution of (μ_c, μ_t) and plot the joint density using an appropriate function from the ggdensity package.

Modeling Election Outcomes.

On November 4, 2014 residents of Kansas voted to elect a member of the United States Senate to represent the state. After the primaries, there were four major contenders in the race: 1) Republican incumbent Pat Roberts, 2) Democrat Chad Taylor, 3) Independent Greg Orman, and 4) Liberatarian Randall Batson.

For this problem we will reference polling data that can be found here.

In mid-August 2014 a SurveyUSA poll of 560 people found the following vote preferences:

Pat Roberts	Chad Taylor	Greg Orman	Randall Batson	Undecided
37%	32%	20%	4%	7%

Ignoring the "undecided votes' ', the maximum likelihood estimate for the true vote shares of each candidate assuming, assuming a multinomial distribution over the 4 candidates, is simply the fraction of people.

[1] the problem of estimating two normal means with unknown variances is a classic problem in statistics called the Behrens–Fisher problem.

a) Assume that you first interview the 7% of undecided voters. They claim they are equally likely to vote for any of the four candidates. Before reviewing the other survey data, you decide to use this information to construct a prior distribution for the true vote shares of the four candidates. What is the prior distribution and what are it's parameters (think pseudocounts)? Given the survey data above and the prior, specify the posterior distribution for the vote shares of the four candidates and the parameters of this distribution.

b) On September 3, 2014 Democratic nominee Chad Taylor withdrew from the race. Assume that amongst those who said they would vote for Taylor in the August survey, 70% of them changed their vote to Orman, 20% to Baston, and the remaining 10% for Roberts. The above information should be used construct a new prior distribution for the 3-candidate race again assuming that the undecided voters from the August poll will now vote equally among the remaining three candidates. Calculate the new posterior distribution over the vote shares for the 3 remaining candidates. Use Monte Carlo to find the posterior probability that more people in Kansas support Pat Roberts than Greg Orman.

c) From October 22-26, 2014 SurveyUSA released a poll of 623 found the following preferences:

Pat Roberts	Chad Taylor	Greg Orman	Randall Batson	Undecided
42%	–	44%	4%	10%

Pat Roberts	Chad Taylor	Greg Orman	Randall Batson	Undecided
42%	NA	44%	4%	10%

Use the posterior from the previous part as the prior for this new survey. Compute a new posterior given the new survey data above. Assume that the population consists of 100,000 eligible voters. However, not all eligible voters actually vote. In fact, roughly between 30-50% and eligible voters actually turn out in a midterm election https://www.fairvote.org/voter_turnout#voter_turnout_101. You express your uncertainty by assuming that the fraction of eligible voters who actually turn out is a Beta(40, 60) random variable. Assuming a random sample of eligible voters actually turn out, generate 10000 samples from the posterior predictive distribution. Use Monte Carlo to answer the following questions.

i. Greg Orman's team believes that if they can get at least 20,000 votes they will win the election. What is the posterior predictive probability that Greg Orman receives at least 20,000 votes *and* wins the election?

ii. Both leading candidates fear that the third party vote is taking away potential supporters. What is the posterior predictive probability that the difference between Greg Orman and

Pat Roberts is smaller than then vote total for Randall Batson.

d) Discuss the assumptions that were made to generate your predictions. If you think some assumptions were poor, how might you change the model to improve upon them? This is an open-ended question with no right or wrong answers and will be graded on thoughtfullness and effort only.