Stat 574B: Bayesian Methods, HW 4

- Make the presentation nice, following directions on HW 1. Please hand in one problem on Wednesday 11/15 and one on Wednesday 11/22. Class will be cancelled on 11/22 so email the HW to me by that day but bring a hard copy to class on Monday following Thanksgiving. I will leave it up to you which problem you decide to hand in each week.

Problem 1

(a) Suppose that $Y_i|\theta$ are independent exponential random variables with density $p(y_i|\theta) = \theta\exp(-\theta y_i)$, where $y_i > 0$ and $\theta > 0$, and define $Y = (Y_1, Y_2, ..., Y_n)'$. Note that the joint density of Y given θ is

$$p(y|\theta) = \prod_{i=1}^{n} \theta \exp(-\theta y_i) = \theta^n \exp(-\theta \sum_{i=1}^{n} y_i).$$

Assuming that $\theta \sim \text{Gamma}(\alpha, \beta)$, find the marginal density of Y, that is, find

$$p(y) = \int p(Y|\theta)p(\theta)d\theta$$

and show that this depends on $y = (y_1, y_2, ..., y_n)'$ only through $\bar{y} = \sum_{i=1}^{n} y_i/n$.

(b) Let y_{obs} be the observed value of Y for a sample of size n. Box's p-value for testing this model is defined to be

$$p = \Pr(p(Y) \leq p(y_{obs}))$$

where this probability is evaluated relative to the marginal distribution of Y. Letting $\bar{Y}$ and $\bar{y}_{obs}$ be the average of the elements of Y and y_{obs}, respectively, show that $p(Y) \leq p(y_{obs}) \iff \bar{Y} \geq \bar{y}_{obs}$. In essence, this says that larger values of $\bar{y}_{obs}$ correspond to smaller values of p, that is larger values of $\bar{y}_{obs}$ are less consistent with the model.

(c) The following data, from Fiegl and Zelen (1956), give the survival times in weeks for 17 leukemia patients: 65, 156, 100, 134, 16, 108, 121, 4, 39, 143, 56, 26, 22, 1, 1, 5, 65. At the time of the study, the typical survival time for patients given a similar treatment was 1 year (52 weeks). An early analysis of the data assumed that the individual survival times followed an exponential distribution, for which a reasonably non-informative prior is $\theta \sim \text{Gamma}(1, 52)$. Given the observed survival times, devise an algorithm to sample from the marginal distribution derived in (a), where $(n, \alpha, \beta) = (17, 1, 52)$. Use this algorithm to approximate p. Do the data seem inconsistent with the model?

(d) A (frequentist-based) way to qualitatively assess the exponential distribution assumption is to make a q-q plot of the data (read up on it). This can be easily done in the **cars** package, as illustrated below.

```
library(car)
y<- c(65, 156, 100, 134, 16, 108, 121, 4, 39, 143, 56, 26, 22, 1, 1, 5, 65)
qqPlot(y,dist="exp")
```

The observed and theoretical quantiles (which arise from a standard exponential with mean 1) should ideally be proportional, that is fall on a straight line. The function I used above gives confidence bands based on the expected order statistics - see the online help. The basic idea is that the bands quantify how much variation about the line might be expected from sample to sample variation. To understand the level of variation you might expect in such plots, you can randomly generate 17 observations from a standard exponential via $y \leftarrow \text{rexp}(17)$ and make the plot. Doing this repeatedly will show you that substantial variation from what is expected might be observed, even when the data are generated from the distribution in question. With this mind, make the q-q plot of the survival data and comment on whether the exponential assumption seems plausible, or not.

(e) Assuming the exponential model with a Gamma prior, we know that the posterior distribution of θ is also Gamma. Suppose that we consider using a Metropolis algorithm to simulate the posterior, where the proposal distribution is uniform. In particular, suppose we propose $\theta^* \sim U(\theta_i - \delta, \theta_i + \delta)$ for some $\delta > 0$, where θ_i is the current value in the chain. Given the survival data and the known posterior for θ, explore how the choice of δ impacts the mixing of the Markov Chain, by looking at characteristics such as the acceptance rate and the ACF. What values of δ seem to be good choices, and what values of δ lead to poor performance? Write a coherent summary (using words placed together in grammatically correct sentences).

Problem 2

I want you to revisit the ELS data analysis that we examined in class. Here we will think a bit more about how to rank schools and to describe uncertainty in the ranks.

Associated with the posterior distribution of $\theta = (\theta_1, \theta_2, ..., \theta_m)$ is the posterior distribution of $R = (R_1, R_2, ..., R_m)$, where R_i is the rank of the i^{th} school's θ_i in an ordered list of the θ_is. Assume that ranks $1, 2, ..., m$ correspond to increasing order on the θ_i's so being "number 1" is nothing to brag about. We wish to summarize the joint distribution of the ranks.

(a) Using Hoff's prior and Gibbs sampling routine, generate the posterior distribution for the ranks R. This is easy, as all you have to do is rank the θ_is for each posterior sample of $(\theta_1, \theta_2, ..., \theta_m)$. Compute the posterior mean, SD, and a 95% CI for R_i for each of the 100 schools. Rank the schools from worst to best (call this the Bayes rank) based on the mean of the posterior ranks. Plot the Bayes rank against the posterior mean of θ_i score for each school. Is there strong agreement between the Bayes ranks and the posterior means of the θ_is in the sense that rankings based on these two summaries are consistent?

(b) Plot $E(R_i|data)$ and the CI for R_i as a function of the Bayes rank - you might consider modifying Hoff's code that he used to plot the math scores for each school, where schools were ordered by mean score. Qualitatively describe what you see. For example, how distinct are the $E(R_i|data)$s (the summary used to order schools) for schools with fairly similar Bayes ranks and is there much variation in the distribution of the ranks for the schools? If so, is this variation observed across all schools?

Discuss.

(b) Another summary of interest is $p_{ij} = \Pr(\theta_i < \theta_j | data) = \Pr(R_i < R_j | data)$ for schools $i < j$. If p_{ij} is close to 0 or 1 then the separation between these schools is clear in the sense that their relative rankings do not flip-flop much in the posterior distribution. Provide a cleaver way to plot the p_{ij} where the axes account for the Bayes rank of the schools. Discuss what you see.

(c) Devise a metric (or metrics) that attempts to answer to the following (ill-posed) questions: How far apart do the Bayes ranks of two schools have to be for the difference in Bayes ranks of these school to be meaningful (in a pairwise comparison)? How far apart do the Bayes ranks of two schools have to typically be for the posterior means of the θ_i to differ by 5, which might be considered an important difference? Your answers may require a bit of nuance, as differences in ranks and means undoubtedly depend on location within the order.

(d) Write a short summary of your analysis.

Remark: CI refers to credible interval, which is another term for probability interval. It does not refer to confidence interval.....