**Homework 7**
**Statistics 200B**
**Due Mar 21, 2019**

1. Consider a Bayesian model in which, conditional on unknown parameter $\lambda$, $X_1, \ldots, X_n$ are iid with exponential PDF

$$f(x|\lambda) = \frac{1}{\lambda} e^{-x/\lambda}$$

for $x > 0$, and the prior distribution is $InverseGamma(a, b)$, with PDF

$$f(\lambda) = \frac{b^a}{\Gamma(a)} \lambda^{-a-1} e^{-b/\lambda}$$

for $\lambda > 0$.

(a) Find the posterior distribution for $\lambda$, conditioning on $X_1, \ldots, X_n$. It is fine to write the family and specify its parameters; you do not need to write out the CDF or PDF.

(b) Show that the posterior mean can be written as a weighted average of the prior mean and the MLE for $\lambda$. You may use the fact that the mean of an Inverse Gamma distribution with parameters $a$ and $b$ is $\frac{b}{a-1}$. What happens as $n \to \infty$?

2. Here is an example for which we can use the model in problem 1. Let $\lambda$ represent the average time (in units of days) between earthquakes in the Berkeley area. To make this more precise, let's consider only earthquakes with magnitude 3 or greater on the Richter scale, and whose epicenter is within a 10 mile radius of downtown Berkeley, whose coordinates I have as $37°52'18''N$ and $122°16'22''W$.

(a) Consider using an Inverse Gamma prior for $\lambda$. We need to choose the parameters $a$ and $b$. You may have some prior knowledge about $\lambda$, but it may be difficult to translate this into a choice of $a$ and $b$. To facilitate this, write expressions for $a$ and $b$ in terms of the prior mean $m$ and the prior variance $v$, using that $m = \frac{b}{a-1}$ and $v = \frac{b^2}{(a-1)^2(a-2)}$ when $a > 2$.

(b) Based on your current knowledge, choose parameters $a$ and $b$, and make a plot of the prior PDF. You may find it useful here and in the rest of the problem to modify the `R` code in the file `BetaBinomial.R`, which is on bCourse. (There is a `dinvgamma` function in the `R` package `MCMCpack`, which you can install and load using `install.packages("MCMCpack")` and then `library(MCMCpack)`, or you can just code the mathematical form of the prior PDF directly.) Turn in a sentence of explanation with your plot regarding how your prior knowledge (or lack of it) informed your choice of prior distribution.

(c) The file `BerkeleyEarthquakes.RData` on bCourse contains a data frame called `earthquakes` with information about earthquakes within a 10 mile radius of Berkeley, from 1969-2008. Load it in and take a look at the first few lines, then extract the waiting time between each earthquake using

```
load("BerkeleyEarthquakes.RData")
head(earthquakes)
x <- earthquakes$Lag[-1] # First element is NA
```

Using the results you found in problem 1, calculate the posterior distribution for $\lambda$, conditional on the observed waiting times. Make a plot comparing your posterior PDF to your prior PDF. Turn in a sentence of explanation with your plot regarding any changes in your knowledge about $\lambda$ after seeing the data.

3. Suppose $X_1, \ldots, X_n | \theta \overset{iid}{\sim} Unif(0, \theta)$. Show that the Pareto distribution is the conjugate prior distribution for $\theta$. The PDF for a random variable $Y \sim Pareto(\psi, \alpha)$ is

$$f(y; \psi, \alpha) = \frac{\alpha \psi^\alpha}{y^{\alpha+1}} I\{y > \psi\}.$$

4. Consider rejection sampling when the target density $h(\theta) = f(\theta | x^n)$. In class we considered taking the proposal density $g(\theta) = f(\theta)$, i.e., the prior PDF. We set $M = \mathcal{L}(\hat{\theta}_n)$, where $\hat{\theta}_n$ is the MLE for $\theta$. Explain why we should not take $M$ to be any larger than this. Explain why we should not take it to be any smaller.

5. Suppose $\theta$ has truncated normal distribution with parameters $\mu$, $\sigma^2$, $\alpha$, and $\beta$. That is, when $\alpha < \theta < \beta$,

$$f(\theta; \mu, \sigma^2, \alpha, \beta) \propto (2\pi\sigma^2)^{-1/2} \exp\left\{ -\frac{1}{2\sigma^2} (\theta - \mu)^2 \right\}$$

and $f(\theta; \mu, \sigma^2, \alpha, \beta) = 0$ otherwise.

(a) Consider using rejection sampling to sample from this distribution, using the PDF of a $Normal(\mu, \sigma^2)$ distribution as your proposal density. What are the steps of the algorithm in obtaining $B$ iid samples?

(b) Write down an expression for the probability of acceptance in any iteration of the algorithm. When will the algorithm be efficient (i.e., have high probability of acceptance)?

(c) Write R code to generate 1000 iid samples from the truncated normal distribution with $\mu = 10$, $\sigma^2 = 4$, $a = 9$, and $\beta = 13$. Run it and make a histogram of the samples. (Turn in both your code and your plot.)

6. Suppose you are in the following setting, which is a simple but realistic clinical trial. 100 people in a control group receive a placebo (or standard treatment), and 100 receive a new experimental treatment. In the control group, $x_1 = 33$ survive at least one year, while in the experimental group, $x_2 = 38$ survive at least one year. Let $p_1$ be the probability of one-year survival under control and $p_2$ under treatment in the hypothetical population from which these patients were drawn.

   (a) Find the posterior distribution for $p_1$ and $p_2$, assuming the two groups of patients are independent and that we use prior distribution $p_1, p_2 \overset{iid}{\sim} Unif(0, 1)$. You may use the result in Example 11.1 of Wasserman, but make sure you understand the steps in obtaining it.

   (b) Using the method suggested on page 406 of Wasserman, sample from the posterior distribution of $\delta = p_2 - p_1$. Make a histogram of the sampled values.

   (c) Suppose a doctor tells you that a value of $\delta$ greater than 0.04 would represent a breakthrough in treatment. Use your sample to approximate the posterior probability that this drug represents a breakthrough.