Homework 2

1. For $y_i|\lambda \sim \text{Poisson}(\lambda)$, the conjugate prior is $\lambda \sim \text{Gamma}(a, b)$. The posterior given a sample of size n will also be Gamma. You will collect some count data in this homework. You will also specify two different priors, and work with both priors.

   (a) Algebraically derive the posterior $p(\lambda|y_1)$ given a single observation y_1. Specify your answer (a) in terms of a named distribution with parameters (ie Gamma(a, b), and specify a and b), and give the actual density formula.

   (b) What is the support (place where density/function is non-negative) of: (i) prior, (ii) posterior, (iii) sampling density, (iv) likelihood (v) prior predictive?

   (c) Algebraically derive the posterior given a sample $y_i, i = 1, \ldots, n$ of size n. Again specify it both in terms of a named distribution with parameters, and give the actual density formula.

   (d) In the prior gamma(a, b), which parameter acts like a prior sample size? (Hint: look at the posterior from problem (1c), how does n enter into the posterior density?) You will need this answer in the second part of problem 1(g)iv or you will do that problem incorrectly.

   (e) You will go to your favorite store entrance and count the number of customers entering the store in a 5 minute period. Collect it as 5 separate observations $y_1, \ldots, y_5$ of 1 minute duration each, this allows you to blink and take a break if needed. This will give you 5 data points.

   (f) Name your store, and the date and time.

   (g) We are now going to specify the parameters a and b of the gamma prior density. We will do this in two different ways, giving two different priors. We designate one set of prior parameters as a_1 and b_1; the other set of prior parameters are a_2 and b_2.

      i. <u>Before</u> you visit the store, make a guess as to the mean number of customers entering the store in one minute. Call this m_0. This is the mean of your prior distribution for λ.

      ii. Make a guess s_0 of the prior sd associated with your estimate m_0. This s_0 is the standard deviation of the prior distribution for λ. Note: most people underestimate s_0.

iii. Separately from the previous question 1(g)ii, estimate how many data points n_0 your prior guess is worth. That is, n_0 is the number (strictly greater than zero) of data points (counts of 5 minutes) you would just as soon have as have your prior guess of m_0.

iv. Solve for a_1 and b_1 based on m_0 and s_0.

v. Separately solve for a_2 and b_2 using m_0 and n_0 only. You usually will not get the same answer each time. This is ok and is NOT wrong. (Note: if you do get the same answer, then please specify a second choice of a_2, b_2 to use with the remainder of this problem!)

(h) Suppose we need to have a single prior, rather than two priors. Suggest 2 distinct methods to settle on a single prior.

(i) Go to your store and collect your data as instructed in 1e. Report it here.

(j) Update both priors algebraically using your 5 data points. Give the two posteriors.

(k) Give the posterior mean and variance for your two posteriors.

(l) Plot your two prior densities on one graph. Plot your two posterior densities in another graph. (Use the algebraic formula, or you can use the `dgamma` function in R). In one sentence for each plot, compare the densities (talk about location, scale, shape and compare the two densities).

(m) Plot each prior density/posterior density pair on the same graph. For each plot, compare the two densities in one sentence.

(n) Use WinBUGS (twice) to update your two priors with your data to get your two posteriors. Compare summary statistics between the two posteriors.

(o) How close are the WinBUGS numerical calculations to the actual algebraically calculated posterior means?

2. For this problem, treat the data as a single count y of customers that entered the store in 5 minutes. Define λ_1 as the 1 minute mean which you worked with previously. Define λ_5 as the 5 minute mean which you will work with now. Let a_5 and b_5 be the 5 minute prior parameters for λ_1 and similarly let a_1 and b_1 be 1 minute prior parameters from above.

   (a) Give algebraic formulas for the relationships between (i) λ_5 and λ_1, (ii) the *prior* mean of λ_5 and λ_1, (iii) prior variances, (iv) prior standard deviations, (v) prior a-parameters, and (vi) b-parameters. (Hint: Transformation-of-variables.)

(b) Give the two priors for the parameter λ_5 that correspond to your priors for λ_1.

(c) Give the two resulting posteriors for λ_5.

(d) Explain the relationship between the posterior means of λ_5 and λ_1. Repeat for the posterior variance, posterior standard deviation, posterior a-parameters and finally posterior b parameters.

(e) Do you need to redraw your plots (of priors and posteriors) that you drew in the previous problem? How could you alter them without redrawing to make them conform to the new data structure?

(f) Do your conclusions change if you consider your data as a single 5 minute observation or as 5 one minute observations? That is, do your recommendations to the store on staffing levels change?

In plotting gamma densities, you can use the `dgamma` function of `R`. You will need to calculate the densities for a range of λ running from 2 or 3 standard deviations below the mean to around 5 or 6 standard deviations above the mean. If you want to plot the density yourself, take a log of the density, calculate $a * \log b + (a-1) * \log \lambda - b * \lambda - \log \Gamma(a)$, then take the exponential. `R` has an `lgamma` function for calculating $\log \Gamma(a)$.

To sample from the posterior, set up your prior as lambda5 $\sim$ gamma(a5,b5), and single observation y as a dpois(lambda5). See the R2WinBUGS model and code below. First we sample from the posterior for lambda5, then from the prior.

```
library(R2WinBUGS)
setwd("C:\\Users\\Rob\\Courses\\Bayes\\homework\\hw_R_functions")
getwd()

Save to file "poissonmodel.txt":
model
{
        lambda5 ~ dgamma(a,b)
        y ~ dpois(lambda5)
        }

# data for posterior; assumes prior mean of 1, prior variance of 4,
```

```
# and a count of 7 customers in 5 minutes.

bugsdata = list(y=7, a=.25, b=.25)
#initial values for posterior.
bugsinits = list(lambda5=2)

bugsparms = list("lambda5")

resultspost = bugs(bugsdata, bugsinits, bugsparms,
    model.file="poissonmodel.txt", n.chains=1, n.iter=10001,
    n.burnin=1, n.thin=1)

print(resultspost)
# This model is so simple that we run only 1 chain and 1 iteration burnin.
# But this is not standard.

bugsdata2 = list(a=.25, b=.25)
#initial values for prior.
bugsinits2 = list(list(y=1, lambda5=2))

bugsparms2 = list("lambda5", "y")

resultsprior = bugs(bugsdata2, bugsinits2, bugsparms2,
model.file="poissonmodel.txt", n.chains=1, n.iter=10001,
n.burnin=1, n.thin=1, DIC=F)
# without DIC=F, WinBUGS blows up.

print(resultsprior)
```