Lab 2: Components of Bayes' theorem

The lab is designed to help you understand the relationship among the component of Bayes' Theorem. You will need to: 1) creates vectors using the probability density functions in the numerator of Bayes' theorem (the likelihood and the prior), 2) multiplies these vectors to obtain the joint distribution of the parameters and the data, 3) integrates the joint distribution and 4) divides the joint distribution by its integral.

Problem

You will estimate the posterior distribution for the mean number of individuals of an invasive plant species per m^2 in a disturbed grassland. We will call that mean θ. You have prior information telling you that the average number of these plants per m^2 is 10.2 with a standard deviation of the mean = .5. You have a set of fifty observations in hand obtained by sweaty labor in the field. Execute the following steps.

1. Simulate 50 data points from a Poisson distribution with mean θ = 6.4 to represent the data set. What is the variance? Be sure to put the R function set.seed(3) before the call to rpois() to assure that we all get the same results. Call the data vector y.
2. Plot a histogram of the data with density on the y-axis. It turns out that the histogram function in R is not really appropriate for discrete data (why?). Here is a chance to write a function that plots discrete data properly! Hint– the count() function in the plyr package and the type="h" argument in the plot function might prove useful. (You can skip this discrete histogram bit with no loss of value from the rest of the exercise.)
3. Set values for the prior mean (mu.prior) and standard deviation (sigma.prior).
4. Set up a vector containing a sequence of values for θ, the mean number of invasive plants, You want this vector to approximate a continuous θ so be sure it contains values that are not too far apart. Use code like this: theta = seq(0,15,step) where you set step = .01. Setting a value for step with global scope is important. You will use it later when you integrate.

The prior distribution of θ

5. Write a function for the prior on θ. Use a gamma distribution for the prior (refer to the distribution cheat sheet). The function for the prior should return a vector of gamma probability densities, one for each value of θ. It should have arguments 1) the vector for θ you created in the previous step as well as 2) the prior mean and 3) the prior standard deviation. The mean and the standard deviation, of course, will need to be moment-matched to the proper parameters of the gamma distribution.

```
prior <- function(theta, mu = mu.prior, sigma = sigma.prior){#code implementing function}
```

6. Plot the prior distribution of θ, the probability density of θ as a function of the values of θ.
7. Check your moment matching by generating 100,000 random variates from a gamma distribution with parameters matched to the prior mean and standard deviation. Now

compute the mean and standard deviation of the random variates. They should be very close to 10.2 and .5.

The likelihood

8. Write a function for the likelihood. The function must use all 50 observations to compute the *total* likelihood across all of the data points (not the log likelihood) for each value of the vector θ. It should have arguments for the vector θ and the data. The function should create and return a vector with elements [y|θi]. Note that this is the total probability density of all of the data for *each* value of θi, not the probability density of a single data point. In reality, θ is a continuous random variable, the mean of the Poisson distribution. We are discretizing it here into small intervals. The function template will be something like:

```
like <- function(theta, y){#code to calculate total likelihood of the data conditional on each valu
e of theta}
```

9. Plot the likelihood of the parameter value conditional on the data, L(θi|y) as a function of θi. Recall L(θi|y) is proportional to [y|θi]. We assume the constant of proportionality = 1. What is this plot called? Can you say anything about the area under the curve? What happens to inference we can make based on likelihood if we multiply the curve by a constant?

The joint distribution

10. Create a function for the joint distribution of the parameters and the data as the product of the prior and the likelihood functions. Call this function joint. The function should simply call the previous two functions and multiply them. Recall that when a function is composed of a single statement as it is here, the statement can simply follow the function template on the same line; curly brackets are not needed. Plot joint(theta) as a function of theta. Does this seem reasonable? Why are the values on the y axis so small? Think about what is going on here.

The marginal probability of the data

11. Approximate the integral of the likelihood multiplied by the prior to obtain a normalization constant [y]. How would you accomplish this integration? (Hint–Recall the first principles definition of the definite integral.) What is this mathematical expression for this integral? Explain the output of this integration, a scalar. Why do we call [y]a "distribution" if it evaluates to a scalar?

The posterior distribution

12. Compute the posterior distribution by dividing each element of the vector of output produced by the joint function by the integral of the joint function. Plot the posterior as a function of θ.

Putting it all together

13. Plot the prior, a histogram of the data, the likelihood, the joint, and the posterior in a six panel layout.