

LAB 3. BAYES LAW

ESTIMATING PARAMETERS WITH BAYES LAW

The goal of these exercises is to assure that you can do moment matching and that you understand the individual components of Bayes law. Here is some advice. The key to success is to create code that does the following:

- 1) create vectors using the probability density functions in the numerator of Bayes law (the likelihood and the prior),
- 2) multiply these vectors to obtain the joint,
- 3) integrate the joint, and
- 4) divide the joint by its integral.

You will need to learn how to use the *integrate()* function to accomplish the third step. Because this integral is a numerical approximation, it will be important to choose the appropriate range over which you integrate. As with all modelling challenges, break the problem down into individual pieces, solve each piece, then put them together. Don't move from one piece to the next until you are satisfied that the first one is working and makes sense.

EXERCISE 1. Write a function that takes the mean and the variance of a gamma distribution as arguments and returns the shape and scale parameters and a second one that takes the shape and scale parameters and returns the mean and the variance.

EXERCISE 2. You have prior information taken from the literature on aboveground net Primary production (ANPP) of forbs in a savanna-woodland. The mean production is 41.4 g / m^2 , with a standard deviation of 2 g / m^2 . You clip 30 plots and obtain new data. The mean of the new data is 53.7 with a standard deviation of 10.1.

Calculate the posterior distribution of the average production. To start—look at the lecture on Bayes law and go through its derivation, particularly the section on the likelihood, the prior, the joint, and the probability of the data. Keep the equation in front of you as you do this exercise. Think about what each piece means as you write your code.

Steps:

Set up a collage of 6 panels for plots using `par (mfrow=c(2,3))`.

- 1) Simulate 30 data points from a normal distribution with mean 53.7 and sd 10.1---this portrays the data that you obtained in your clipped plots, but it is a lot less work.
- 2) Set up a vector containing a sequence of values for θ , the mean ANPP, using code like `theta = seq(25,75,.01)`. Call this sequence `theta`.
- 3) Output a vector representing the prior using the vector θ and the prior mean and standard deviation as input. It will be useful to create a function to accomplish this. Plot the prior, i.e. $P(\theta)$ as a function of θ . Does your plot make sense?
Plot a histogram of the data. Use the `freq = FALSE` option to get the density rather than the frequency.
- 4) Create a function for the likelihood. The function must use all 30 new observations to output the total likelihood (not the log likelihood) of the data. Create a vector of $P(y|\theta)$ using the likelihood function and the vector `theta`. Plot the likelihood as a function of `theta`. Does it make sense?
- 5) Create a function for the joint as the product of the prior and the likelihood. It should have `theta` as its first (or only) argument. Call this function `joint`. Output a vector

containing the product of the likelihood and the prior. Does this seem reasonable?

6) Integrate the function for the joint over the values of theta, i.e.,
`Py=integrate(joint,25,75, abs.tol=1e-100)`

The function `integrate()` returns a list of two values: 1) the scalar result of the integration and 2) an estimate of the absolute error. To use the scalar in future calculations, you will need to use something like `Py$value` to extract the scalar result.

7) Divide the vector output by the joint function by the definite integral obtained in 6. This is the posterior distribution. Plot the posterior, i.e. $P(\theta|y)$ as a function of θ .

8) Check to be sure that your posterior distribution was correctly calculated using the formula for the mean and standard deviation of the posterior based on the conjugate of a normal distribution, given by:

$$\mu_{post} = \frac{\mu_{prior} / \sigma_{prior}^2 + \mu_{data} n / \sigma_{data}^2}{1 / \sigma_{prior}^2 + n / \sigma_{data}^2}$$

$$\sigma^2 = \frac{\sigma_{prior}^2 \sigma_{data}^2 / n}{\sigma_{prior}^2 + \sigma_{data}^2 / n}$$

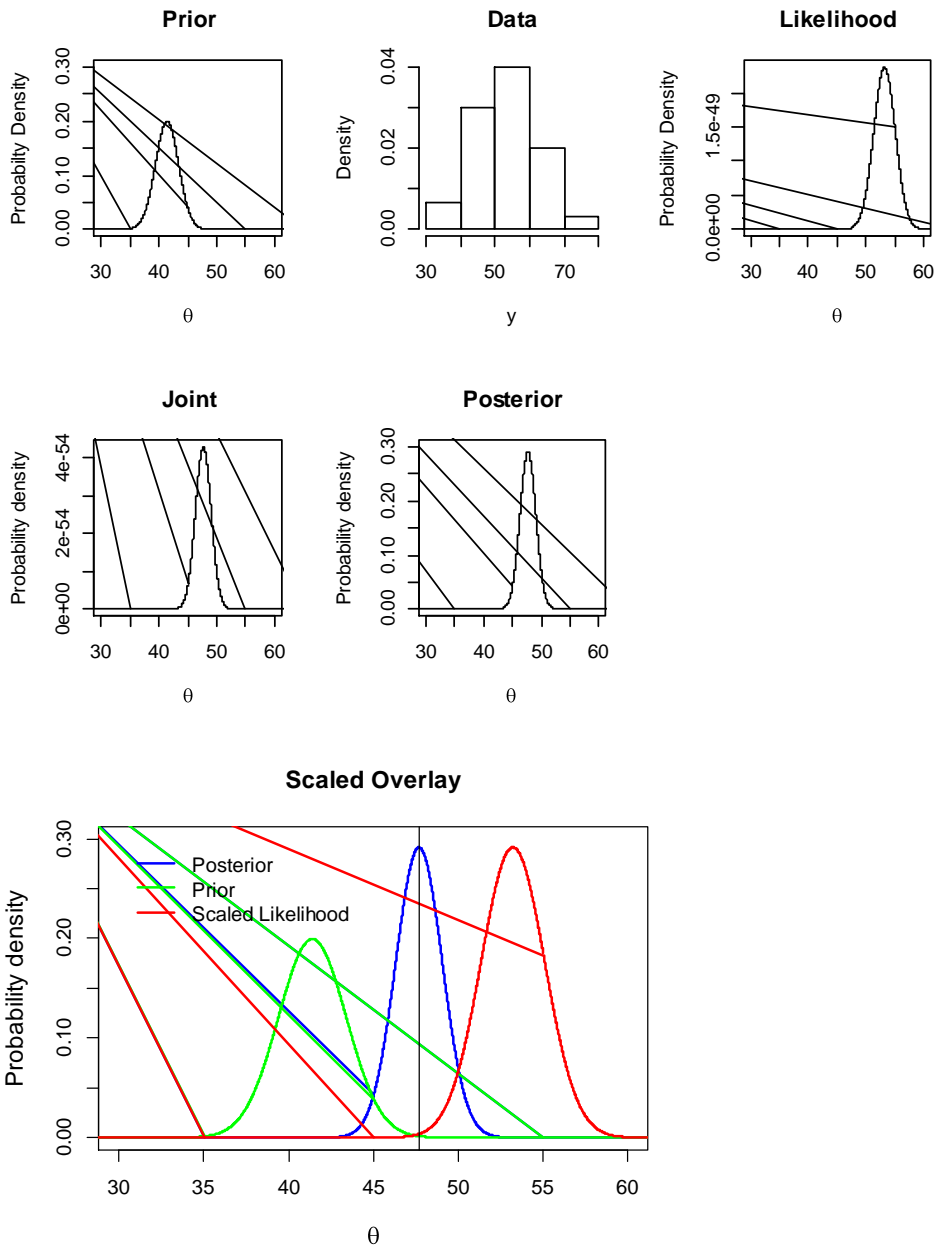
Overlay the plot of the posterior that you calculated using integration with a plot of the posterior derived from the conjugate formula. (Hint--Make the line width of the first plot = 4, type = dashed, and color= red. Use the default line width and color for the second plot)

9) Scale the likelihood vector so that the MLE value has the same probability density as the mean of the posterior (hint, divide the likelihood vector by the maximum of the likelihood vector and multiply by the maximum of the posterior vector). Think about what you are doing here and discuss it with your lab mates.

10) Change your plot output to a single panel. Overlay the prior, the posterior, and the scaled likelihood on the same plot, color coding them so that you know which is which.

Recalculate the posterior distribution after increasing the standard deviation on the prior. What happens to the posterior relative to the likelihood? Now return the prior standard deviation to its original value and increase the standard deviation in the simulated data. What happens to the posterior relative to the likelihood and the prior?

If you set up your code properly, these last two steps should require changing only the values for the standard deviations of the old and the new data and rerunning your code. Your plots should resemble the ones on the next page. Why are they not exactly the same as the ones I made?



EXERCISE 3. You have prior information on the prevalence of chronic wasting disease in female mule deer in the area north of Fort Collins [prevalence = number of infected / (number infected + number healthy)]. The average prevalence in an earlier study was 0.09 (i.e., 9%) with a standard deviation of 0.02. You take new data on the population and find 2 females are sick out of a total sample of 80 females.

Compute the posterior distribution of prevalence including the old and the new data. (Hint: remember moment matching. What types of data is prevalence? What about the data?) Generate individual plots of the prior, the likelihood, the joint and the posterior. Check your estimate of the posterior using the formula for the shape parameters of the posterior distribution of a binomial sample with a beta prior using overlaid plots, as above. Overlay a plot of the posterior, the prior, and the scaled likelihood. You should be able to modify the code you used above to accomplish these tasks easily, particularly if you created functions for the prior, the likelihood, and the joint. Note that the beta is an improper distribution so it does not integrate to 1.