NRES 779: Bayesian Hierarchical Modeling in Natural Resources

Lab 04: Bayes Theorem

Contents

I Motivation

Bayesian analysis is flexible and intuitive because it uses the same approach to all problems. We learn about unobserved quantities from observed ones using the laws of probability as an inferential scaffolding. We derived Bayes' Theorem in lecture from those basic laws. We learned that all Bayesian models have the same parts: the posterior distribution, the joint distribution composed of the likelihood(s) and priors, and and the marginal probability of the data, which serves as a normalizing constant after the data are collected. For most of this course, we will exploit the proportionality between the posterior and the joint distribution to allow us to dispense with the marginal probability of the data using simulation methods (MCMC). However, it is vital that you understand the relationship among the components of Bayes' Theorem before we do that. This understanding is the foundation for all of the powerful methods that follow. It is what sets Bayesian analysis apart from maximum likelihood.

This problem is a bit of a toy example because you will never estimate posterior distributions this way, at least I never have. But, it is an instructive toy example. Reliably using Bayesian methods requires understanding how Bayes'

theorem works. The best way to gain this understanding is to compute a posterior distribution from its component parts, the prior, the likelihood, and the marginal data distributions. It helps you understand how you multiply and divide by statistical distributions, something you probably don't do every morning after coffee. You will be thrilled to know that you only need to do this once.

To start, take out your notes on Bayes' theorem and go through its derivation we discussed at the beginning of lab, particularly the color-coded section on the likelihood, the prior, the joint distribution, 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.

The key to success is to create code that does the following:

- creates vectors using the probability density functions in the numerator of Bayes' theorem (the likelihood and the prior),

- multiplies these vectors to obtain the joint distribution of the parameters and the data,

- integrates the joint distribution and

- divides the joint distribution by its integral.

II Problem

You are interested in estimating 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 = 0.5. 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 $\theta = 6.4$ to represent the data set. (This portrays the data that you gathered from plots, but it is lots easier to obtain.) What is the variance? Be sure to put the R function `set.seed(2021)` 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, but if you do, I urge you to revisit it sometime.)

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 = 0.01`. Setting a value for `step` with global scope is important. You will use it later when you integrate.

III The prior distribution of θ

5. Write the mathematical expression for a gamma prior on θ. Be as specific as possible given the information at hand.

6. Write an R function for the prior on θ. The function for the prior should return a vector of gamma probability densities, one for each value of θ. It should have arguments

   - the vector for θ you created in the previous step
   - the prior mean
   - 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 in your function.

   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.

   So, in this case `mu.prior=10.2` and `sigma.prior=0.5`. You could hard-code these in the function template, but that is bad practice.

7. Plot the prior distribution of θ, the probability density of θ as a function of the values of θ.

8. 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 0.5.

IV   The Likelihood

9. What is the mathematical expression for the likelihood $[y|\theta]$, assuming that the data are conditionally independent? Be as specific as possible using the information at hand.

10. Write an R 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 $[y|\theta_i]$. Note that this is the total probability density of all of the data for each value of θ_i, 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.

11. Plot the likelihood $[y|\theta_i]$ holding the data constant and varying θ.

12. What is this plot called?

13. Can you say anything about the area under the curve?

14. What happens to the inference we can make based on likelihood alone if we multiply the curve by a constant?

V   The Joint Distribution

15. What is the mathematical expression for the joint distribution $[\theta, y]$?

16. Create an R 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.

17. Plot `joint(theta)` as a function of `theta`.

18. Does this seem reasonable?

19. Why are the values on the y-axis so small?

VI   The Marginal Probability of the Data

20. What is the mathematical expression for the marginal probability of the data $[y]$?

21. Approximate the marginal probability of the data, 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.)

22. Explain the output of this integration, a scalar. Why do we call $[y]$ a "distribution" if it evaluates to a scalar?

VII   The Posterior Distribution

23. What is the mathematical expression for the posterior distribution $[\theta|y]$?

24. 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.

25. Plot the posterior as a function of θ.

VIII   Putting It All Together

26. Plot the prior, a histogram of the data, the likelihood, the joint, and the posterior in a six panel layout. Your results should be the same as the plot below:

27. Overlay the prior, the likelihood, and the posterior on a single plot. To do this, you will need to re-scale the likelihood, which of course is OK because it is defined up to an arbitrary, multiplicative constant, i.e., $[y|\theta] = cL(\theta|y)$. It doesn't matter what c is. We can re-scale the likelihood to any value we like and the inference does not change because all evidence is relative in the likelihood framework. Do the following to obtain a scaled likelihood that can be plotted in a revealing way next to the posterior distribution. Divide each element in the likelihood vector by the maximum likelihood (thus, the maximum becomes 1). Multiply the resulting vector by the maximum value of the posterior density. Simple.

28. Check to be sure that everything is correct using the gamma-Poisson conjugate relationship. A gamma distribution is the conjugate for the Poisson likelihood, which means if we have a gamma prior and a Poisson likelihood, then the posterior is a gamma distribution with parameters $\alpha + \sum_{i=1}^{n} y_i$, and $\beta + n$ where α and β are the parameters of the gamma prior, and we have n observations (y_i) of new data. Overlay a plot of the posterior obtained using the conjugate on the plot of the posterior obtained by integration. Take a look at your scaled overlay of the posterior, the likelihood, and the prior. The likelihood profile for θ is based on the data but it shows much less dispersion than the distribution of the data shown in the histogram you constructed. Why?

29. Now that you have these lovely functions working and plots emerging from them, do some experiments to understand the effect of prior information on the posterior distribution of θ relative to the effect of the data. Increase the variance of the prior distribution to 2.5. Reduce it to 0.1. Increase the number of observations from 50 to 100. Examine the overlaid plots you produced above for each case.

30. Discuss the position of the prior, likelihood and posterior along the x axis and their variances.