Bayes BATS Tier 2 – Progress Report 1

Recap

Our previous discussions revolved around finding an activity that incorporates real-world data and Bayesian thinking in an accessible manner. This has admittedly proven to be a difficult task embedded in this larger task of designing a class that has never been taught previously and is unique in my department's context. So, I went back to the beginning and asked myself – what do I want my students to get out of my class? At the end of the day this is a class about "telling stories with data". How do we take a research question, collect some data, say something about this data, and contextualize this information given our own biases and prior experiences? This introspection led to a series of discussions with colleagues and the students whom I may be teaching in the future. Each of these conversations resulted in three connected themes.

1. What exactly are prior beliefs?
2. How do we combine prior beliefs with data?
3. Does the data matter more? Or the prior beliefs?

These 3 questions coalesce into the instructional aims of a proposed activity.

Instructional Context

While still under construction, the general outline of the course will be as follows: research design and data acquisition (what are data), connecting data to scientific inquiry (telling stories with data), and reaching a broad audience (reproducibility in science). The proposed activity will be completed in 2 parts. This activity will play off anthropology students' general aversion to statistics using the Beta-Binomial model. The aims of this extended activity are 1) to introduce probability using one's prior beliefs, and 2) to actively participate in the data-gathering step and connect the data to one's prior beliefs. I describe each aim below.

Activity 1

This activity will generate a Shiny application and lead to a classroom discussion. The expected outcome of this activity is that the student walks away with a better understanding of prior distributions, probability, and uncertainty. Each student is expected to have different backgrounds, so the expectation is we will be able to talk about how "prior" information is not static, but dynamic across contexts. During the 1st week of classes I will pose 2 questions: 1) What is your statistical background, and 2) Given that background, how well do you think you'd do if you took a statistics quiz today. The fact that this 2nd question is a conditional distribution will be hidden for now. These will be answered on a Shiny application. The answers to the 1st question will be tied to an unseen Beta distribution. This will represent the prior on the parameter p in the beta-binomial model. At this stage, students will not be introduced to the meaning of scale and shape, but instead we will focus on the shape of distribution. The goal is to demonstrate how each of our prior beliefs can be quantified probabilistically and how this can be different per person. Following a discussion about probability, uncertainty, and prior beliefs, I pose to the students that if we are going to use this prior belief in our analysis, we need data. This leads to the next activity.

Activity 2

This activity will utilize a previously unseen version of the Shiny application where I have aggregated all the previous responses. The thinking is that this activity will take place 2 to 3 weeks after the first. During this delay, I will cover basic research design, what is data, data wrangling, etc. By this point, the students will have been thinking about data and will be ready to connect this with some research question. This activity explicitly covers the formal introduction of Bayes to the class. Little did they know that they have actively participated in the data collection process via their answer to question 2. The probability of success here is defined as "passing" the perceived quiz. A portion of the Shiny application will show students the posterior, likelihood, and prior information but without any probabilistic language. Built into the application will be components designed for students to play around with the data and the prior to see how the posterior may change. The goal here is to demonstrate the balancing act of Bayesian analyses and to discuss how we can interpret Bayesian results.

Goals

My preliminary discussions and planning for this course have shown me that anthropology students are unsure about how their data can be "analyzed statistically". I felt what better way to demonstrate this connection by showing them a) data can be generated in any fashion, and b) we can connect it to something very personal or familiar. This will help this assignment reach a broader audience as opposed to focusing in on a certain dataset. I do assume upfront that students have little to know familiarity with statistical and/or mathematical language. The hope is that after this 2-part exercise, students will know what a parameter is, what a prior is, how to quantify uncertainty, and how to describe the balancing act of Bayesian inference. Following this activity, we will move into credible intervals / confidence intervals and more complex models.

Perceived Bottleneck

There is a small degree of risk involved with an extended activity. The students could forget our previous discussions, they may change their minds, or they may find the gap too large to connect the points. But, given the audience I am designing this for and the overall flow of the semester, I see some value in introducing priors first as a form of introducing probabilistic thinking. I believe it will also give students a personal stake in the analysis and will allow for a deeper connection once we formally delineate Bayes theorem. I foresee some confusion if I were to introduce prior, likelihood, and posterior, altogether. It may be best to begin the semester with prior beliefs before we even talk about data – in fact arguably, priors in our model may relate to if or how we collected data in the first place.

Revised Timeline

By our November check-in I will have the Shiny application complete and associated discussion questions. It may be best to forgo the Oct. 15th conference deadline and focus more on a conference deadline in the spring for the American Anthropological Association. By that point I will have some data and feedback to present to colleagues about the best way to learn these methods.