Bayes-BATS: Tier 2 Proposal

Adam Gilbert Laura Lambert

August 2, 2024

Table of contents

Overview

We propose to create several activities which introduce and utilize Bayesian Inference techniques in a variety of contexts. These activities could be used as a vehicle for including data analysis and inference within a course from the corresponding context domain or as a collection of applied activities for introducing Bayesian topics within a statistics course.

Detailed Proposal

This project seeks to create several activities/modules which introduce Bayesian inference topics in a variety of contexts. The modules will be written as activities that could either be inserted into a course in the context domain or used as an activity in an applied statistics course.

We envision that each activity will be self-contained in order to encourage adoption within courses that aren't statistics-focused. For example an instructor for a course on *Field Methods and Ecological Principles* could utilize a finished version of our *Chytrid Fungus in Frogs* activity as a single day discussion on statistical analysis of field data. Similarly, an instructor for a

course in *Macroeconomics* could utilize a future activity on estimation and comparison of unemployment rates across different states or counties. For statistics courses, having activities written across a variety of contexts may permit students the flexibility of choosing to work within a context that they care about, while still gaining exposure to the desired topic.

To further increase the likelihood of adoption, the activities will be written using `{WebR}`, which allows users to write and execute R code from a web browser. That is, neither students nor the instructor would need a local installation of R in order to utilize these activities. This provides a no-stakes gateway into statistical computing. See our draft proof of concept on Chytrid Fungus in Frogs.

The hope is that these activities can be confidently used not only in a statistical/Bayesian course as learning support, but also in other STEM courses. By involving data sources from both our colleagues as well as 'outside world' sources we hope to increase departmental interest and buy-in to incorporating Bayesian thinking in STEM courses. Additionally, by lowering the barrier of access (using `{WebR}` rather than R/RStudio), we strive to allow instructors and students who are not R users to confidently incorporate these concepts into their courses.

In addition to the activities being planned, activity-specific learning objectives/goals will be written and shared, as well as a list of potential course-level objectives the activity would map well to. At the end of each activity, a suggested assessment plan is also offered for a variety of contexts.

Possible Bayesian Topics and Domain Contexts

We can envision eventually developing activities covering the following concepts and within the example domains listed.

From the *Prior* to the *Posterior*

This activity family explores allowing observed data to update our assumptions. These activities provide an intuitive first-exposure to Bayesian thinking.

- **Applications in Biology:** prevalence of a mutation within a population, prevalence of a disease within a population, physiological measures on animal populations
- **Applications in Chemistry:** molecular weights of polymers with a particular starting monomer but using different catalysts and additives (or vice-versa)
- **Applications in Environmental Science:** distributions of water purity metrics (turbidity, conductivity, pH, etc.)
- **Applications in Business and Economics:** unemployment rate, inflation rate, monthly/weekly expenditures on raw materials, yearly revenue, salary distributions
- **Applications in Social Science:** average galvanic skin response (GSR) at baseline or under a stressor for individuals

- **Applications in Criminal Justice:** measuring average sentence length recommendations for a particular class of felony convictions or the likelihood of at least one violent crime being reported in a particular neighborhood on a given day
- **Applications in Ethics:** measuring the probability of a particular response to a given instance of the *trolley problem*

Estimating a Population Parameter

This activity family explores the notion of extracting and interpreting credible intervals from a posterior distribution. Currently, this activity is lumped into our sample "proof of concept" activity. However, we would like to explore whether breaking that activity up into two parts would be preferable. We would plan to develop versions of this class of activity across multiple disciplines, similar to those outlined above.

Comparison of Population Parameters Across Populations

This activity family considers sampling from posterior distributions corresponding to different sub-populations and asks users to aggregate and compare the results of counter-factual draws. Again, we would plan to develop versions of this class of activity across multiple disciplines.