Leila Peitsch
12/15/23
Mentor: Erin Lipman

Bayesian Perspectives on Probability and Statistics (Pt. 2)

Introduction

In this Directed Reading Program (DRP), I continued my study of Bayesian statistics to deepen my understanding. Unlike the commonly used frequentist statistics, Bayesian statistics grants more interpretable results and a framework that allows for understanding to be influenced by both incoming data and prior information. In Spring 2023, I learned about Bayesian conjugate priors that make it so we can calculate the posterior distribution directly - a helpful tool in simple Bayesian models. However, in real-world analyses, conjugate priors are unable to be utilized if we have more complicated models with many parameters. For these cases, we would find the posterior distribution using posterior approximation via a method called Markov Chain Monte Carlo. We can also model these multi-leveled models using hierarchical modeling - two concepts we focused on within this DRP.

Posterior Approximation

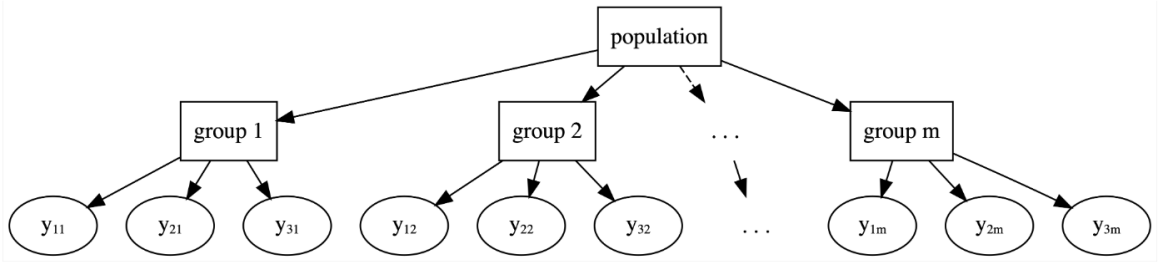
In order to approximate the posterior via *simulation*: two methods that were explored in the textbook *Bayes Rules! An Introduction to Applied Bayesian Modeling* were grid approximation and Markov Chain Monte Carlo - though we definitely focused on the latter, as we utilized it in our applied analysis project during the second half of the quarter. Grid approximation produces a set of *N* independent values of some parameter from a discretized approximation of the posterior pdf. In certain models, grid approximation suffers the curse of dimensionality - a phenomenon where this method becomes less informative beyond a certain dimensionality - as grid approximation can become very computationally expensive. However, Markov chains Monte Carlo methods are much more flexible for posterior approximation.

Markov chain Monte Carlo (MCMC) simulation produces a *chain* of *N* dependent values from the parameter that are *not* drawn from the posterior pdf. For MCMC methods, there are a multitude of different algorithms such as Hamiltonian Monte Carlo (used in the *rstan* package in R that we used in our project) and Metropolis-Hastings. In a further study of MCMC, with the guidance of the textbook, I was able to implement a Metropolis-Hastings algorithm in base R. Metropolis-Hastings is a two step iterative process that (1) proposes a new location μ', conditioned on the current location μ, then (2) calculates an acceptance probability which decides whether or not to go to the new location μ' or stay at the current location μ. In evaluating

if a MCMC simulation is “good”, we use a few diagnostics such as trace plots and parallel chains to visually observe if there are any inconsistencies in how the MCMC explores the posterior distribution. Furthermore, there are a few numerical checks that can be calculated such as: effective sample size, autocorrelation, and R-hat ($\widehat{R}$). Effective sample size of some chain quantifies the number of independent samples it would take to produce an equivalently accurate posterior approximation, given that *N* is the actual sample size/length of the dependent Markov chain. Autocorrelation looks at how different pairs of Markov chain values are correlated, a numeric way of looking at how the Markov chain is mixing. Mixing refers to how the MCMC simulation is exploring the posterior distribution, fast mixing being “good” as it fully explores the distribution randomly (ex. not getting stuck in one place or staying in one portion of the distribution), similar behaviorally to independent sampling, and slow mixing being “bad” as it doesn’t produce an accurate posterior approximation. R-hat allows us to calculate the ratio between the variability across all parallel chains and the variability within a particular chain.

Hierarchical Modeling

Hierarchical models can often refer to many different types of models such as multileveled models, mixed effects models, and random effects models. In this DRP, we used Bayesian hierarchical models to utilize partial pooling. Partial pooling is a way of using parts of no pooled and completed pooled models - a way that separates data into individual groups (instead of one big group or all individual values) that grant us access to observing trends within subgroups and subgroups providing information about other subgroups.



An general framework for hierarchical models

Models often involve incoming data at the top level, values we extract from some dataset we are observing, followed by parameters that are inferred about the subgroups (ex. state population, general population), and sometimes involve *hyperparameters*. Priors on parameters are called *hyperpriors*, the hyperparameters are parameters of the hyperprior. We usually choose values for hyperparameters, but they are helpful for sensitivity analysis and explicitness of the overall

model. It also is used for computational convenience, like conjugate priors, and it does *not* affect the larger Bayesian inference.

Conclusion

Overall, this DRP was a great experience and I was very grateful to be able to continue my study of Bayesian philosophy in probability and statistics from the previous quarter. Not only has it deepened my preference and admiration for Bayesian statistics but it gave me an opportunity to build upon my previous experience (similar to Bayesian inference). I was able to create a very interesting analysis using a dataset that covered a topic I am heavily invested in, mental health in the US. Using Bayesian inference helped aid the understanding of possible outcomes regarding the data within the real-world (such as asking questions like, “why does California have such a low percentage of people who take medication for mental health?” and “do these numbers possibly reflect state populations’ accessibility to medication?”). I would definitely recommend this experience to *anybody* who has any interest in statistics!