Syllabus
EES 5891-xx: Bayesian Statistical Methods

Jonathan Gilligan
Vanderbilt University

Spring 2026

Contents

1 Nuts and Bolts

1.1 Class Meetings

MW 8:40–9:55 Classroom TBA

1.2 Professor

Jonathan Gilligan
Professor of Earth & Environmental Sciences
Professor of Civil & Environmental Engineering
Professor of Climate & Environmental Studies
jonathan.gilligan@vanderbilt.edu
www.jonathangilligan.org
Office Hours: TBA or by appointment.

1.3 Email

If you want to communicate with me be sure to begin the subject line of your email with "EES 5891" This helps assure that I will see your message quickly and respond to it.

I have set my email reader to flag all messages like this as important, so I will read them first. This also assures that I do not mistake your email for spam. I typically receive over 100 emails per day, so if you do not follow these instructions, I may not notice your email.

1.4 Course web site

In addition to Brightspace, I have set up a companion web site for this course at https://ees5891.jgilligan.org, where I post the reading and homework assignments, my slides from class, and other useful material. That web site will be the central place to keep up with material for the course during the semester. This web site will direct you to Brightspace if there is anything you need to find there.

2 Course Description

2.1 Concise Description

The class will begin with an introduction to Bayesian statistics and then focus on practical application of regression methods to data. We will use R together with the Stan software package (`https://mc-stan.org`) for Hamiltonian Monte Carlo methods. The course will combine practical applications of Bayesian methods to real (often messy) data with more philosophical discussions of Bayesian approaches to statistics and how to interpret results of statistical analyses. We will focus on regression methods, including hierarchical or multilevel regression modeling methods, which can be very powerful when you have data that has a nested structure (e.g., cities and counties within states or species within genera). Students will do projects applying Bayesian methods to their own data sets.

2.2 Prerequisites

You should be comfortable with differential and integral calculus and have some previous experience with standard statistics.

This course will be very mathematical and will make extensive use of the R software system, but I do not assume that you already know R or advanced mathematics beyond calculus.

2.3 Narrative Description

Bayesian statistics is a branch of statistics that has been around for almost 300 years, but for most of that time, it was very difficult to apply to practical problems because the mathematical equations were too difficult to solve. In the last 30 years, as computers have become much faster and more powerful, new computational methods have emerged that make Bayesian statistics practical for research and applications.

Bayesian statistical methods are valuable because they provide a systematic way to combine what you already know about a problem with new data from experiments or observations, and the results of Bayesian analyses are more straightforward to interpret than conventional statistics. These methods are widely used across a wide variety of research as well as practical applications.

Bayesian analysis is used by geologists to improve estimates of mineral distributions and radon hazards. It is widely used in meteorology to make weather forecasts and in climate science to combine data from many different sources and come up with quantitative predictions and detailed understanding of their associated uncertainties. One powerful application in climate science was to integrate model output with empirical paleoclimate data to put constraints on the earth's climate sensitivity. It is used to analyze results from high-energy particle physics experiments to discover new subatomic particles. It's used by biologists to identify and categorize variations in the genomes of humans and other species.

Bayesian methods are used extensively in medicine to analyze the results of clinical trials, to determine the pharmacokinetics of drug metabolism, and to assess the predictive value of tests for diseases such as cancer or COVID infection. It's used in political science and sociology to improve the accuracy of public opinion surveys and to understand patterns of voting. It has been applied to public health to estimate the prevalence of diseases and to make more effective treatment decisions when medical tests are uncertain.

Bayesian analysis is widely used in business and marketing to identify consumer preferences and improve the effectiveness of advertising. If you use Google, Amazon, Netflix, Stitchfix, or practically any large online platform for shopping or entertainment, advanced Bayesian methods form the basis of their recommendations.

Bayesian analysis has also been applied effectively to law and criminology to assess the value of evidence in proving guilt or innocence.

Bayesian methods are also widely used in computational applications, such as image analysis and reconstruction, computational text analysis, and natural language processing. One of the earliest practical applications of Bayesian textual analysis, in 1964, identified the anonymous authors of the Federalist Papers. More recent applications of Bayesian textual analysis are used to separate desired email from spam.

This course will provide a general introduction to Bayesian statistics and will combine practical instruction in how to do Bayesian data analysis and philosophical discussions about how to think about the assumptions that go into a Bayesian analysis and how to interpret the results that it produces.

You do not need to have any prior knowledge of computer programming, but I do expect that you are familiar with basic statistics and calculus (both derivatives and integrals).

3 Goals for the Course

By the end of the semester, you will:

- Understand Bayes's theorem and how to apply it.
- Understand problems with the traditional statistical emphasis on null-hypothesis significance testing (NHST), why Bayesian approaches to NHST don't solve these problems, and how Bayeian statistics offers superior alternatives to NHST.
- Understand how think about statistical models, how to choose an appropriate model for your problems, and understand the tradeoffs between different kinds of models.
- Be able to design and conduct a comprehensive Bayesian analysis of data from start to finish.
- Understand how to choose appropriate priors for your Bayesian analyses and how to test whether your choice of priors is sound.
- Understand how to set up, perform, assess the validity of, and interpret the results of Bayesian regression analysis.
- Understand why Markov Chain Monte Carlo (MCMC) sampling is used in Bayesian analysis, what the limits of MCMC are, and how to test your MCMC analyses for validity.
- Understand and be able to perform analyses using more complex statistical models, such as interaction models, generalized linear models, models of discrete (categorical and count) data.
- Understand what multilevel or hierarchical models are, when to use them, and how to interpret the results of a multilevel analysis.

4 Structure of the Course:

I divide the semester into three parts:

1. **Introduction to Bayes's Theorem and its Applications:** The first part of the course introduces the basic concepts of Bayesian statistics, using simplified approximations to calculate difficult equations. This section will focus on linear regression methods.
2. **Monte Carlo Methods:** Next, we study Monte Carlo methods, which help us solve more difficult problems that our earlier approximations are not powerful enough for. This section will introduce statistical models of discrete data (counts, categories, etc.), and generalized linear models.
3. **Hierarchical and Multilevel Statistical Modeling:** We will conclude the semester with multilevel statistical models, which can be very powerful methods for working with large and complex data sets. In these, we divide the data into several sets, or levels, and conduct simultaneous regression analysis on each level. Two famous early applications of multilevel modeling to earth & environmental sciences included estimating the risk of radon in homes and siting deep tube wells to provide safe, arsenic-free drinking water in Bangladesh.

4.1 Reading Material

There is one required textbook and two optional books. Supplementary reading on the Internet or in handouts will also be assigned during the term and posted on Brightspace.

Required Reading Materials

- Richard McElreath, *Statistical Rethinking: A Bayesian Course with Examples in R and Stan,* 2nd ed. (CRC Press, 2020; ISBN 978-0-367-13991-9). This will be the principal textbook for the semester. Be sure you get the second edition because it is significantly different from the first.

There is a companion web site to *Statistical Rethinking* at `xcelab.net/rm/statistical-rethinking/`, which has links a number of resources, including videos of the author's own lectures on the material.

For people who are familiar with R and like to work in the `tidyverse` dialect, there is a free companion e-book on the web at `bookdown.org/content/4857/`, that has translated almost all the R code in the book into the `tidyverse` dialect of R.

Optional Reading Materials

- John Kruschke, *Doing Bayesian Data Analysis,* 2nd ed. (Academic Press, 2015; ISBN 978-0-12-405888-0). This book has a more elementary introduction to Bayesian statistics, at an undergraduate level. It is very clear, but it focuses almost entirely on Monte Carlo sampling and doesn't go as deeply into other important aspects of Bayesian statistics as we will do in this course.

  It is a very useful resource to check out if Monte Carlo sampling seems confusing. I have asked the Science & Engineering library to put a copy on reserve so you can read it there without needing to buy it.

- Hadley Wickham and Garrett Grolemund, *R for Data Science* (O'Reilly, 2017; ISBN 978-1-491-91039-9). This book is a great introduction to the R statistical programming language. It uses the `tidyverse` dialect of R, developed by Hadley Wickham. There is a free web-based ebook version at `r4ds.had.co.nz/`, so you won't need to buy a paper copy.

Overview of Reading Materials

I will give out detailed reading that give specific pages to read for each class and notes on important things you should understand. **I expect you to complete the reading before you come to class on the day for which the reading is assigned**, so you can participate in discussions of the assigned material and ask questions if there are things you don't understand.

4.2 Graded Work

Basis for Grading

Class participation	5%
Homework	45%
Project	50%

Homework

Homework is due at the beginning of class on the day it is assigned.

Projects

You will do an extended research project in the second half of the semester, in which you will apply Bayesian methods to investigating a data set that you choose. This could be data from your dissertation research or another data set that interests you.

Tests and Examinations

There will be no tests or exams in this course.

5 Honor Code:

This course, like all courses at Vanderbilt, is conducted under the Honor Code.

Studying: As you study for this class, I encourage you to to seek help from me or from other classmates or friends.

Homework: I encourage working together. I also encourage you to talk with other classmates, as well as friends and acquaintances outside of class. You may discuss assignments, compare notes on how you are working a problem, and you may look at your classmates' work on homework assignments. But you must work through the problems yourself in the work you turn in: **Even if you have discussed the solution with others you must work through the steps yourself and express the answers in your own words. You may not simply copy someone else's answer.**

Research project: The research project will be conducted under the same ethical principals that apply to publishing papers in scientific journals. The work must be your own, but you may consult any other resources. If anyone else makes a substantial contribution, you must list them and their contributions in an Acknowledgements section.

If you ever have questions about how the Honor Code applies to your work in this course, please ask me. **Uncertainty about the Honor Code does not excuse a violation.**

6 Final Note:

I have made every effort to plan a busy, exciting, and instructive semester. I may find during the term that I need to revise the syllabus to give more time to some subjects or to pass more quickly over others rather than covering them in depth. Many topics we will cover are frequently in the news. Breaking news may warrant a detour from the schedule presented on the following pages. Thus, while I will attempt to follow this syllabus as closely as I can, you should realize that it is subject to change during the semester.