STA 602: Bayesian and Modern Statistics

Spring 2020
Duke University

Instructor:	Olanrewaju Michael Akande, Ph.D.
Email:	✉ olanrewaju.akande@duke.edu
Office Hours:	**Mondays (6pm - 7pm) and Thursdays (9am - 10am)**. Zoom meeting ID: **See Sakai**.
Course Page:	https://sta-602l-s21.github.io/Course-Website/
Meeting Times:	**Tuesdays and Thursdays (10:15am - 11:30am)**. Zoom meeting ID: **See Sakai**.
Teaching Assistants:	Jennifer Kampe. **Times TBD**. Zoom meeting ID: **See Sakai**.
	Ziang Wang. **Times TBD**. Zoom meeting ID: **See Sakai**.
Labs:	**Section 01: Mondays (12pm - 1:15pm)**. Zoom meeting ID: **See Sakai**.
	Section 02: Mondays (1:45pm - 3pm). Zoom meeting ID: **See Sakai**.
Required Textbook:	*A First Course in Bayesian Statistical Methods* Peter D. Hoff, 2009, New York: Springer. (Available online from Duke library.)
Optional Textbooks:	*Bayesian Data Analysis* (Third Edition) by Andrew Gelman, John Carlin, Hal Stern, David Dunson, Aki Vehtari, and Donald Rubin.
Important Dates:	• Wed, January 20 — Classes begin • Tue, February 2 — Drop/Add ends • Thur, February 11 — Quiz I day (*tentative*) • Tue - Wed, March 9 - 10 — No classes held • Mon, March 15 — Midterm exam day (*tentative*) • Thur, April 8 — Quiz II day (*tentative*) • Mon, April 12 — Wellness day • Fri, April 23 — Classes end • Fri - Sat, April 30 - May 1 — Final exam period

1 Course Overview

Bayesian methods are increasingly important in both industry and academia. This is a graduate-level course that introduces students to the basics of Bayesian inference and provides students with the tools needed to fit Bayesian models.

In this course, you will learn the importance of Bayesian methods and inference. You will be introduced to Bayesian theory, with particular emphasis on conceptual foundations as well as implementation and model fitting. You will learn the essential distinctions between classical and Bayesian methods and become familiar with the origins of Bayesian inference. You will also learn about conjugate families of distributions and why they are very convenient, and how to conduct Bayesian inference with intractable posterior distributions, when you do not have conjugate distributions.

Although this course emphasizes the mathematical theory behind Bayesian inference, data analysis and interpretation of results are also important components. Students who wish to explore the mathematical theory in more detail than what is covered in class are welcome to engage with and request further reading materials from the instructor. Also, all students must have the theoretical background covered in the prerequisites to be able to keep up with and understand the materials.

2 Learning Objectives

By the end of this course, students should be able to

- Understand the basics of Bayesian inference, that is, be able to define likelihood functions, prior distributions, posterior distributions, prior predictive distributions and posterior predictive distributions.
- Derive posterior distributions, prior predictive distributions and posterior predictive distributions, for common likelihood-prior combinations of distributions.
- Interpret the results of fitted models and conduct checks to ascertain that the models have converged.
- Use the Bayesian methods and models covered in class to analyze real data sets.
- Assess the adequacy of Bayesian models to any given data and make a decision on what to do in cases when certain models are not appropriate for a given data set.

3 Course Format

This is an online course designed to be primarily synchronous. However, there will also be some asynchronous activities. Students will be required to do pre-assigned readings, go through lecture slides, watch pre-recorded lecture videos, and take the quizzes embedded in the videos, all before each synchronous meeting time. The meeting times, which will be held on Zoom, are thus designed to be live demonstration, discussion and Q&A sessions. Each live meeting session will also be recorded and made available to all students afterwards. Additional live sessions include office hours for the instructor and TAs. Those will not be recorded. Students who are unable to attend the office hours can send their questions in advance of the live meeting sessions, so that the instructor or TAs can provide answers during those recorded sessions.

4 Course Info

4.1 Playposit

To gain access to the pre-recorded lecture videos, you will have to create a Playposit account. There are participation quizzes embedded within the videos. These quizzes make up a part of your final grade (see: course policies) so take them seriously. To join the class on Playposit, you need to create a new account as a student here. Next, you will use the class link, which I will send out via email, to join the class site. While you need not create an account with your Duke email, I strongly suggest you do.

4.2 Zoom meetings

The easiest way for you to join the different Zoom meetings is to log in to Sakai, go to the"Zoom meetings" tab, and click "Upcoming Meetings". For the recordings (for lab and discussion sessions), also log in to Sakai, go to the "Zoom meetings" tab, and click "Cloud Recordings". Those will be available few minutes after the sessions.

5 Prerequisites

ALL students are expected to be familiar with all the topics covered within the required prerequisites to be in this course. That is, STA 611 or the following: STA 210 and (STA 230 or 240L) and (MATH 202, 202D, 212, or 222) and (MATH 216, 218, or 221, any of which may be taken concurrently). Students are also expected to be very familiar with R and are encouraged to have learned LaTeX or a Markdown language by the end of the course.

6 Class Materials

Lecture notes and slides, lab exercises and assigned readings will be posted on the course website, while lecture and lab videos will be posted on Sakai. White boards will also be used frequently in the lecture videos, so please pay special attention to those. Finally, we will closely follow the main textbook so students should make sure to always read the corresponding textbook chapters in the assigned readings.

7 Workload

You are expected to put in approximately 18-25 hours of work per week. The work hours will include time spent going through the preassigned readings, watching the lecture videos, watching or attending the lab sessions, and doing all graded work. Please note that the more focused and engaged you are, the quicker you will be able to get through all the materials.

8 Graded Work

Graded work for the course will consist of problem sets, lab exercises, two quizzes, a midterm exam and a final exam. Regrade requests for problem sets and lab exercises must be done via Gradescope AT MOST **24 hours** after grades are released! Regrade requests for quizzes, midterm, and final exams must be done via Gradescope AT MOST **12 hours** after grades are released!

- There are no make-ups for any of the graded work except for medical or familial emergencies or for reasons approved by the instructor BEFORE the due date. Contact the instructor in advance of relevant due dates to discuss possible alternatives.
- Grades may be curved at the end of the semester. Cumulative averages of 90% – 100% are guaranteed at least an A-, 80% – 89% at least a B-, and 70% – 79% at least a C-, however the exact ranges for letter grades will be determined at the end of the course.
- Students' final grades will be determined as shown by the breakdown in Table 1.

9 Descriptions of graded work

9.1 Problem sets

Problem sets will be handed out on a weekly basis. They will be based on both the lectures and labs and will be announced every Thursday or Friday – be sure to check the website regularly! **Also,**

Table 1: Breakdown of graded work

Component	Percentage
Final Exam	25%
Midterm	20%
Problem Sets	20%
Quiz I	10%
Quiz II	10%
Lab exercises	10%
Participation Quizzes	5%

please note that any work that is not legible by the instructor or TAs will not be graded (given a score of 0). Every write-up must be clearly written in full sentences and clear English. Any assignment that is completely unclear to the instructors and/or TAs, may result in a grade of a 0. For programming exercises, you are required to use R and you must submit ALL of the code as an appendix.

Each student MUST write up and turn in their own answers. You are encouraged to talk to each other, regarding problem sets or to the instructor/TA. However, the write-up, solutions, and code must be entirely your own work. The assignments must be submitted on Gradescope under "Assignments". Note that you will not be able to make online submissions after the due date, so be sure to submit before or by the Gradescope-specified deadline.

Solutions to the problem sets will be curated from student solutions with proper attribution. Every week the TAs will select a representative correct solution for the assigned problems and put them together into one solution set with each answer being attributed to the student who wrote it. **If you would like to OPT OUT of having your solutions used for as a representative solution, please let the instructor and TAs know in advance.**

Finally, your lowest score on the problem sets will be dropped!

9.2 Lab exercises

The objective of the lab assignments is to give you more hands-on experience with Bayesian data analysis. Join the live session or watch the recorded videos and learn a concept or two and some R from the TAs, and then work on the computational part of the problem sets. Each lab assignment should be submitted in timely fashion. You are REQUIRED to use R Markdown to type up your lab reports.

9.3 Quizzes

There will be two quizzes, both timed and online. Students will be able to take the quiz within **any one-hour slot in a 24-hour window**. Detailed instructions on the quizzes will be made available later.

9.4 Midterm Exam

There will be a midterm exam mid-semester. Students will be able to take the midterm within **any two-hour slot in a 24-hour window**. Soon after the midterm, you will be given a midterm grade assessing your overall performance. Note that the main purpose of this is to let you know how you are doing in the class. Detailed instructions on the midterm will be made available later.

9.5 Final Exam

There will be a final exam after the reading week. If you miss any quiz or the midterm, your grade will depend more on the final exam score since there are no make-up exams. You cannot miss the final exam! Students will be able to take the final within **any three-hour slot in a 48-hour window**. Detailed instructions on the final will be made available later.

10 Late Submission Policy

- You will lose

  1. 40% of the total points on each problem set if you submit within the first 24 hours after it is due, and
  2. 100% of the total points if you submit later than that.

  In addition, you will lose

  1. 50% of the total points on each lab exercise if you submit within the first 24 hours after it is due, and
  2. 100% of the total points if you submit later than that.

- You will lose 100% of the total points on quizzes, midterms and final exams if you miss the dates/times.

11 Auditing

Students who audit this course will be expected to complete most of the graded work with the goal of getting an overall score of at least 70%; you will only need to complete enough graded work to get to 70%. You are also expected to watch the videos, go through the readings, and generally, participate like everyone else. You must contact the instructor in advance if you wish to audit the course.

12 Tentative Course Schedule

We will cover the topics below. We may spend different amounts of time on each topic. For a detailed and updated outline, check on the updated course schedule on the course page regularly.

- **Introduction to Bayesian Inference:**
  - Building blocks of Bayesian inference
  - Probability review
- **One parameter models:**
  - Conjugacy
  - Beta-Bernoulli and beta-binomial models
  - Gamma-Poisson model
  - Other one-parameter exponential family models
  - Loss functions and Bayes risk
- **Monte Carlo and multiparameter models:**
  - Monte Carlo approximation and sampling

- Rejection sampling
- The normal model
- MCMC and Gibbs Sampling

Multivariate Data:

- Multivariate normal model
- Missing data and imputation

Hierarchical models:

- Hierarchical normal models with constant variance
- Hierarchical normal modeling of means and variances
- Introduction to other hierarchical models

Bayesian Linear Regression:

- Bayesian linear regression
- Bayesian hypothesis testing and model selection
- Bayesian generalized linear regression models

Metropolis and Metropolis-Hastings:

- Metropolis
- Metropolis-Hastings
- Metropolis within Gibbs

Categorical data and mixture models:

- The multinomial model
- Finite mixture models for categorical data
- Finite mixture models for continuous data

Wrap up and review

13 Academic Integrity

Duke University is a community dedicated to scholarship, leadership, and service and to the principles of honesty, fairness, respect, and accountability. Citizens of this community commit to reflect upon and uphold these principles in all academic and nonacademic endeavors, and to protect and promote a culture of integrity. To uphold the Duke Community Standard:

☞ I will not lie, cheat, or steal in my academic endeavors;

☞ I will conduct myself honorably in all my endeavors; and

☞ I will act if the Standard is compromised.

Cheating or plagiarism on any graded assessments, lying about an illness or absence and other forms of academic dishonesty are a breach of trust with classmates and faculty, violate the Duke Community Standard, and will not be tolerated. Such incidences will result in a 0 grade for all parties involved. Additionally, there may be penalties to your final class grade along with being reported to the Office of Student Conduct. Review the academic dishonesty policies at `https://studentaffairs.duke.edu/conduct/z-policies/academic-dishonesty`.

14 Diversity & Inclusiveness:

This course is designed so that students from all backgrounds and perspectives all feel welcome both in and out of class. Please feel free to talk to me (in person or via email) if you do not feel well-served by any aspect of this class, or if some aspect of class is not welcoming or accessible to you. My goal is for you to succeed in this course, therefore, let me know immediately if you feel you are struggling with any part of the course more than you know how to manage. Doing so will not affect your grades, but it will allow me to provide the resources to help you succeed in the course.

15 Disability Statement

Students with disabilities who believe that they may need accommodations in the class are encouraged to contact the Student Disabilities Access Office at 919.668.1267 or disabilities@aas.duke.edu as soon as possible to better ensure that such accommodations are implemented in a timely fashion.

16 Other Information

It can be a lot more pleasant oftentimes to get one-on-one answers and help. Make use of the teaching team's office hours, we're here to help! Do not hesitate to talk to me during office hours or by appointment to discuss a problem set or any aspect of the course. Questions related to course assignments and honesty policy should be directed to me. When the teaching team has announcements for you we will send an email to your Duke email address. Be sure to check your email daily.

Most of the course components, including live meeting sessions and all office hours, will be held online using Zoom meetings. If you have any concerns, issues or challenges, let the instructor know as soon as possible. Also, all students are strongly encouraged to rely on Piazza, for interacting among yourself and asking other students questions. You can also ask the instructor or the TAs questions on there and we will try to respond as soon as possible. If you experience any technical issues with joining or using Piazza, let the instructor know.

17 Professionalism

Try as much as possible to refrain from texting or using your computer for anything other than coursework while watching the lecture videos or during the live sessions. Again, the more engaged you are, the quicker you will be able to get through the materials. You are responsible for everything covered in the lecture videos, lecture notes/slides, and in the assigned readings.