STA 360: Bayesian and Modern Statistics

Summer Term II 2020
Duke University

Instructor:	Olanrewaju Michael Akande, Ph.D.
Email:	✉ olanrewaju.akande@duke.edu
Course Page:	https://sta-360-602l-su20.github.io/Course-Website/
Meeting Times:	There will be no "*fully synchronous*" meetings. There will be pre-recorded lecture videos for each topic. There will be Zoom discussion sessions **every day from 11am-12pm**, but those will be recorded and shared with all students. Students are **strongly encouraged** to attend the live sessions and ask questions but students who prefer not to can watch the recordings. Zoom meeting IDs TBD. *Times for the discussion sessions are tentative and will be decided based on a survey of students' preferences.*
Office Hours:	**Tuesdays and Thursdays: 3pm - 4pm**. Zoom meeting IDs TBD.
Teaching Assistants:	Christine Shen. Zoom office hours: TBD. Bo Liu. Zoom office hours: TBD. *TAs are both for STA 360 and STA 602.*
Online Labs:	**Mondays and Wednesdays: 2pm - 3:15pm**. These are live "optional" lab sessions on Zoom for students who wish to attend. Recordings will be made available afterwards for students who prefer not to. Zoom meeting IDs TBD.
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:	• Mon, June 29 — Classes begin • Wed, July 1 — Drop/Add for Term II ends • Fri, July 3 — Independence Day holiday observed • Tue, July 7 — Quiz I day (*tentative*) • Fri, July 17 — Midterm exam day (*tentative*) • Mon, July 27 — Last day to withdraw with W • Wed, July 29 — Quiz II day (*tentative*) • Thurs, August 6 — Classes end • Fri, August 7 - Sun, August 9 — Final exam period

1   Course Overview

Bayesian methods are increasingly important in both industry and academia. This is an upper-level undergraduate 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   Course Format

This is primarily an asynchronous online course, meaning that there will be NO standard meeting time, so that students are able to participate according to their own schedule. That said, the course is designed to align with the standard six-week Term II schedule of the summer school calendar. Therefore, there will be set deadlines to ensure that all course materials and assessments are completed in six weeks.

2.1   Asynchronous Activities

There will be pre-recorded lecture videos on Zoom (via Sakai), so that students can watch them and also do the pre-assigned readings according to their own schedule. However, students will also need to attend one-hour Zoom discussion sessions every day or watch the recorded versions, to stay on top of the materials. Students can also do the problem sets at their own time but within a one-week window from when they are posted on the website.

2.2   Synchronous Activities

There are no "fully synchronous" activities for the course. There will be two live lab sessions on Zoom (to be held by the TAs) per week, for students who wish to attend but recordings will be made available afterwards for students who prefer not to. Lab exercises must be turned in within a 48 hour window from when they are handed out. 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. Likewise, students will be able to take the midterm within any two-hour slot in a 24-hour window. Finally, there will be a final exam, which students will be able to take within any three-hour slot in a 48-hour window.

Additional live sessions include office hours for the instructor and TAs. The exact times will be decided based on a class survey on the availability of students. Students who are unable to attend will be allowed to send in their questions in advance, so that the instructor or TAs can provide answers during the recorded discussion sessions.

3   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.

4   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 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.

5   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.

6   Workload

You are expected to put in approximately 15-20 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.

7   Graded Work

Graded work for the course will consist of problem sets, lab exercises, two quizzes, a midterm exam and a final exam. **Please note that most of the problem sets and lab exercises assigned to students enrolled in STA 602 will contain at least one question more than those assigned to students enrolled in STA 360.**

Regrade requests for problem sets and lab exercises must be done via Gradescope AT MOST **48 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.

➠ Students' final grades will be determined as follows:

➠ 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 after the final exam.

 If grades are curved, grades for students in STA 360 will be curved differently in comparison to the grades for students in STA 602.

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

8 Descriptions of graded work

8.1 Problem sets and lab exercises

There will be five problem sets which will be handed out on a weekly basis. They will be based on all topics covered in the lecture videos and assigned readings will often be posted on the website every Monday, so please check the website regularly! **Also, 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 her or his 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!

8.2 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.

8.3 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.

8.4   Midterm Exam

There will be a midterm exam on July 17. 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.

8.5   Final Exam

There will be a final exam, which you will be able to take between August 7 and August 9. 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.

9   Late Submission Policy

- You will lose

  1. 25% of the total points on each problem set if you submit within the first 24 hours after it is due,
  2. 50% of the total points if you submit within the first 48 hours after it is due, and
  3. 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.

10   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.

11   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.

➠ **Week 1:**

  ▍ Course overview
  ▍ Introduction to Bayesian inference
  ▍ Bayesian vs frequentist paradigms
  ▍ Probability review
  ▍ Conjugacy

▎One parameter models I

⇒ **Week 2:**

▎Loss functions and Bayes risk

▎One parameter models II

▎Monte Carlo approximation and sampling

▎Rejection sampling

▎The normal model I

⇒ **Week 3:**

▎The normal model II

▎Gibbs Sampling

▎The multinomial model

▎Multivariate normal model I

⇒ **Week 4:**

▎Multivariate normal model II

▎Missing data and imputation

▎Hierarchical models

⇒ **Week 5:**

▎Linear Regression Models

▎Metropolis and Metropolis-Hastings

▎Generalized Linear Regression Models

⇒ **Week 6:**

▎Finite mixture models

▎Wrap up and review

12 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 on exams or plagiarism on problem sets, 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. Please review the academic dishonesty policies at `https://studentaffairs.duke.edu/conduct/z-policies/academic-dishonesty`.

13   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, please 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.

14   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.

15   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. Please make sure to check your email daily.

Most of the course components, including discussion sessions, labs and all office hours, will be held online using Zoom meetings. Meeting IDs will be provided closer to the beginning of the course. If you have any concerns, issues or challenges, please 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.

16   Professionalism

Try as much as possible to refrain from texting or using your computer for anything other than coursework while watching the lecture videos. 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, labs, lecture notes/slides, and in the Hoff book. You will be expected to follow along with the Hoff book as we go along in the lectures.