

STA 602L: BAYESIAN AND MODERN STATISTICS

SPRING 2020
DUKE UNIVERSITY

INSTRUCTOR:	OLANREWAJU MICHAEL AKANDE, PH.D.
EMAIL:	✉ olanrewaju.akande@duke.edu
OFFICE:	256 Gross Hall
OFFICE HOURS:	Wed 9:00 - 10:00am and Thur 11:45am - 12:45pm, 256 Gross Hall. (Still subject to change!)
LEAD TA	Jordan Bryan
TEACHING ASSISTANTS:	Zhuoqun Wang. Office hours: Tues 3:00 - 5:00pm, Old Chem 025. Bai Li. Office hours: Wed 3:00 - 5:00pm, Old Chem 025.
LECTURES:	Wed/Fri 11:45am - 01:00pm, Old Chemistry 116.
LABS:	Section 01: Mon 11:45am - 01:00pm, Sociology Psychology 127. Section 02: Mon 01:25pm - 02:40pm, Old Chemistry 101.
COURSE PAGE:	https://sta-602l-s20.github.io/Course-Website/
REQUIRED TEXTBOOK:	<i>A First Course in Bayesian Statistical Methods</i> Peter D. Hoff, 2009, New York: Springer. (Available online from Duke library.)
OPTIONAL TEXTBOOKS:	<i>Bayesian Data Analysis</i> (Third Edition) by Andrew Gelman, John Carlin, Hal Stern, David Dunson, Aki Vehtari, and Donald Rubin.
IMPORTANT DATES:	<ul style="list-style-type: none">• Fri, January 10 First class for STA 602L (not January 8!!!)• Mon, January 20 Martin Luther King Jr. Day; no classes!• Wed, January 22 Drop/add ends• Fri, March 6 Tentative date for midterm exam
	Spring break begins 7:00pm
	<ul style="list-style-type: none">• Mon, March 16 Spring break ends; classes resume 8:30am
	<ul style="list-style-type: none">• Wed, April 15 Graduate classes end
	<ul style="list-style-type: none">• Sat, May 2 Final exam

1 Course Overview

Bayesian methods are increasingly important in both industry and academia. This is a graduate-level course, within the Department of Statistical Science, 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 outside of class. 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 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 L^AT_EX or a Markdown language by the end of the course.

4 Class Materials

Lecture notes and slides, lab exercises and other reading resources will be posted on the course website. In-class black/white boards will also be used frequently so class attendance is required. Finally, we will closely follow the main textbook so students should make sure to always read the corresponding textbook chapters per topic, outside of class.

5 Graded Work

Graded work for the course will consist of homework, lab exercises, quizzes, a midterm exam and a final exam. Regrade requests for homework 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. See the instructor in advance of relevant due dates to discuss possible alternatives.
- ▀ Students' final grades will be determined as follows:

Component	Percentage
Final Exam (in-class, online or take home?)	30%
Midterm (in class)	20%
Homework assignments (at least one per week)	15%
Quiz I (in class)	10%
Quiz II (in class)	10%
Lab exercises	10%
Class Participation	5%

- 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.

6 Descriptions of graded work

6.1 Homework and lab exercises

Homework assignments will be given on a weekly basis. They will be based on both the lectures and labs and will be announced every Friday at the most – please always check the website! **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 homework problems or to the instructor/TA. However, the write-up, solution, 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.

Homework solutions 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 solutions set with each answer being attributed to the student who wrote it. **If you would like to OPT OUT of having your homework solutions used for the class solutions, please let the Instructor and TAs know in advance.**

Finally, your lowest homework score will be dropped!

6.2 Quizzes

Two quizzes will be administered through the semester, one before the midterm (sometime in February) and the other after the midterm (sometime early April). The exact dates will be announced later. There will be no make-ups for quizzes so make sure to be in class on the dates to take them.

6.3 Midterm Exam

The midterm exam will hold either March 4 or March 6 (March 6 is the current tentative date). See the course schedule for the updated date for the midterm exam closer to the end of February. After the midterm, you will be given a midterm grade assessing your overall performance. If you are an

undergraduate, this will also be sent to the registrar. Note that this does not go on your transcript, with the main purpose being to let you know how you are doing in the class.

6.4 Final Exam

If you miss any of the quizzes or midterm, your grade will depend more on the final exam score since there are no make-up exams. You cannot miss the final exam! Detailed instructions on the final exam will be made available later.

7 Late Submission Policy

- You will lose 50% of the total points on each homework if you submit within the first 24 hours after it is due, and 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 do not show up to take them or submit later than the due dates.

8 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 attend most of the classes and participate like everyone else. There are very limited seats for auditing this course and most students will be expected to take the course for credit, therefore, you must contact the instructor in advance if you wish to audit the course.

9 Tentative Course Schedule

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

- ▣ Introduction to course
- ▣ Introduction to Bayesian methods
 - ▣ Bayesian vs frequentist paradigms
 - ▣ Review of probability and exchangeability
 - ▣ Conjugacy
 - ▣ Monte Carlo approximation
 - ▣ Credible intervals
- ▣ One parameter models
 - ▣ Binomial models
 - ▣ Poisson models
 - ▣ Other one-parameter exponential family models
- ▣ The normal model
 - ▣ Unknown mean, known variance
 - ▣ Unknown mean, unknown variance
 - ▣ Prior specification based on expectations

- ▀▶ Markov chain Monte Carlo
 - ▀ Gibbs Sampling
 - ▀ Metropolis and Metropolis-Hastings
 - ▀ Tuning and convergence
- ▀▶ Multivariate Gaussian models
 - ▀ Unknown mean, known covariance matrix
 - ▀ Unknown mean, unknown covariance matrix
 - ▀ Missing data and imputation
- ▀▶ Linear regression
 - ▀ Frequentist inference vs Bayesian inference
 - ▀ Prior specification
 - ▀ Model selection (mini-intro)
 - ▀ Model averaging (mini-intro)
- ▀▶ Finite mixture and latent variable models
 - ▀ Data augmentation
 - ▀ Mixtures of normal distributions
 - ▀ Mixtures of multinomial distributions
 - ▀ Probit regression
- ▀▶ Introduction to multilevel/hierarchical models
- ▀▶ Wrap up (would spend more time on model selection and averaging should time permit).

10 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 homework assignments, 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>.

11 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.

12 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@aes.duke.edu as soon as possible to better ensure that such accommodations are implemented in a timely fashion.

13 Other Information

It can be a lot more pleasant oftentimes to get in-person answers and help. Make use of the teaching team's office hours, we're here to help! Do not hesitate to come to my office during office hours or by appointment to discuss a homework problem 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.

14 Professionalism

Please refrain from texting or using your computer for anything other than coursework during class. You are responsible for everything from lecture, mentioned in class and lab, and in the Hoff book. You will be expected to follow along the Hoff book as we go along in the lectures.