STATS/DATASCI 451: Bayesian Data Analysis

Yang Chen, Department of Statistics, University of Michigan, Ann Arbor

WINTER TERM, 2024

Course Overview

The course is an introduction to both principles and practice of Bayesian inference for data analysis. At the end of this course students will be familiar with the Bayesian paradigm, and will be able to analyze different classes of statistical models. The course gives an introduction to the computational tools needed for Bayesian data analysis and developes statistical modeling skills through a hands-on data analysis approach. Topics include: prior/posterior distributions, Bayes rule, Markov Chain Monte Carlo computations, linear and generalized linear models, mixed effect models, hierarchical models, analysis of spatial data, model selection and comparison, model checking.

Grades are based on weekly quizzes on Canvas, weekly homework (submitted electronically), one midterm exam (closed-book), and one final project (groups of 2-4). For the final project, students are expected to apply Bayesian methods to analyze real data and write up a thorough report on the modeling, computation, and interpretation of the statistical analysis.

Prerequisites: (STATS 412 or STATS 425) and (STATS 306 or EECS 280).

Textbook and References

- Kruschke, John. Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan. Academic Press, 2014.
- Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2014). Bayesian data analysis (Third Edition). Chapman & Hall/CRC Texts in Statistical Science.
- Hoff, Peter D. A first course in Bayesian statistical methods. Springer Science & Business Media, 2009.

Q & A: Piazza

This term we will be using Piazza for class discussion. The system is highly catered to getting you help fast and efficiently from classmates, the GSI, and myself. Rather than emailing questions to the teaching staff, I encourage you to post your questions on Piazza.

Find our class page at: `https://piazza.com/umich/winter2024/statsdatasci451`
Access Code: ilovestatistics

Office Hours and Contact Information

- Instructor: Yang Chen, 445E West Hall, 1085 S. University Ave., Ann Arbor.
  In-person Office hours: TBD.
  Zoom Office hours: TBD, `https://umich.zoom.us/j/97184630329`.
- GSIs: Yuxuan Ke and Jiuqian Shang.
  Office hours: TBD.
- Emails: **All emails are handled via Canvas, for both instructor and GSI.**

Lecture Capture

Each students can request up to two lecture videos for this entire semester. You will have to email the instructor on Canvas with the subject being "STATS 451 lecture capture (date) (your full name)" to request it. For more requests, please have either the SSD email the instructor directly or your graduate/undergraduate advisor email the instructor directly to request. Emails directly sent to the instructor's personal mailbox will not be processed.

Course Outline

This course covers, but is not restricted to, the following materials.

- Review of Probability Theory
  - Likelihood principle, conditional probability, Bayes formula
- Monte Carlo Methods
  - Monte Carlo Simulations
- Basics of Bayesian Modeling and Inference
  - Basic Concepts: Prior, Likelihood, Posterior
  - Examples: Single and Multi Parameter Models
  - Bayesian Hierarchical Model and Shrinkage Estimators
  - Bayesian regression models
- Computational Methods
  - Markov chain Monte Carlo Methods
  - Usage of the STAN Package
  - Variational Inference
  - Model checking and visualization techniques
- Bayesian Inference in Practice

At the end of this course, students (*who make efforts to learn the materials*) will be familiar with the Bayesian paradigm, and will be able to analyze different classes of statistical models. The course gives an introduction to the computational tools needed for Bayesian data analysis and develops statistical modeling skills through a hands-on data analysis approach.

Grading Guideline

A total of 100 points is broken down into four parts.

- Weekly quizzes: 10. Denote M quiz scores by $y_1, y_2, \cdots, y_M$ where each $y_i \in [0, 100], 1 \leq i \leq M$. Let the ordered sequence be $y_{(1)} \leq y_{(2)} \leq \cdots \leq y_{(M)}$. The final quiz score is calculated as follows.

$$(M-5)^{-1} \sum_{1 \leq i \leq M-5} y_{(M-i+1)} \cdot 10\%.$$

  The quizzes will be done at the beginning of the lectures, Monday or Wednesday.

- Homework: 20. Denote N homework scores by $x_1, x_2, \cdots, x_N$ where each $x_i \in [0, 100], 1 \leq i \leq N$. The final homework score is calculated as follows.

$$(N-1)^{-1} \sum_{1 \leq i \leq N, i \neq n_{\min}} x_i \cdot 20\%,$$

  where $n_{\min} = \{i\,;\, x_i = \min\{x_1, x_2, \cdots, x_N\}\}$.

- Closed-book Midterm: 40.

- Final project proposal and presentation: 10. Final project report: 20.
  - See below for detailed descriptions on requirements and grading rubics for final projects.

- Bonus points: up to 2.
  - Active class participation: discussions (Q&A) on Canvas & Piazza.
  - Filling out class surveys.

Formula for letter grades:

- A+ = 97-100, A = 93-97, A- = 90-92
- B+ = 87-89.9, B = 83-86.9, B- = 80-82.9
- C+ = 77-79.9, C = 73-76.9, C- = 70-72.9
- D+ = 67-69.9, D = 63-67.9, D- = 60-62.9

Course Policy

- Q & A are handled **ONLY** on Piazza and office hours, thus please avoid asking technical questions via email. Students are encouraged to help solving each other's problems on Piazza. Students are encouraged to share notes on Canvas or Piazza. Bonus points (up to 2%) will be given according to students' involvement and contributions on Canvas and Piazza.

- Please come to the GSI or the instructor's office hours if you have any questions that you don't want to post on Canvas discussions. Emails sent to the GSI or instructor about (a) contents that can be found on Canvas or Piazza, (b) technical questions including bugs in code, or (c) asking for homework / final project extensions (except medical conditions) won't be replied.

- Proper collaboration is highly recommended. However, you must recognize your collaborators by writing down their names clearly in the front when handing in your homework or final project. Refer to the Academic Misconduct for policies of plagiarism. The instructor will follow the university honor code policy strictly. Lack of knowledge of the academic honesty policy is not a reasonable explanation for a violation. If you discuss with other students about homework, you are supposed to write down your own homework/code independently after the discussions. Under any circumstances, two exact copies of homework/code are considered to be plagiarism. Students are under their own risk of any type of penalty from the university in violating academic honesty.
- All assignments should be fully prepared by the student. Developing strong competencies in the skills associated with this course, from student-based brainstorming to project development, will prepare you for success in your degree pathway and, ultimately, a competitive career. Therefore, **the use of generative AI tools to complete any aspect of assignments for this course are not permitted and will be treated as plagiarism.**

Homework Requirements

- Please put all your answers in a single pdf file and submit it through Gradescope (it should be easy to access through Canvas). If you are using rmd, simply knit it to a pdf file. For those using Jupyter Notebook or JupyterLab, the easiest way to export it as pdf is to download it as html, open the html in your browser, and then print it to pdf.
- The following grading criteria for homeworks is based on 100% maximum. Correctness of derivations: 20%. Correctness of results: 40%. Code: 30%. Writeup: 10%.
  - Dispute of homework score MUST be within 7 business days of homework return.
  - Homework-grading-related questions are resolved at GSI or instructor's office hours.
  - Late homework policy. If "Submission Date - Homework Deadline $>$ 7 days": Grade is 0. Otherwise, each student is allowed one late homework ($\leq$ 7 days) throughout the semester without penalty. Extra late homework (late for less than 7 days) will receive half of the total grade. Medical reasons for late homework: a doctor's note and a signature from your graduate (or undergraduate) advisor are needed.

Final Project Requirements

- **Proposal**.
  - Clarify the purpose of data analysis:
    - What is the problem of interest?
    - What is the final goal you want to achieve?
  - To the best of your knowledge, explain the data as thoroughly as possible:
    - How is the data collected?
    - Is the data a simple random sample from some population?
    - What is the potential defect in the data?
    - If you were able to collect the data, what would you have done differently?

    * Are you going to account for the data collecting process in your analysis?
    * Describe simple facts about the data: size, data structure & source, etc.
    * If you are not going to use real data, explain how you are going to generate "realistic" synthetic data that represents your knowledge of the problem.
  - Sketch a plan about how you are going to work on a collaborative project.
    * If you are working on a final project on your own, give a solid reason.
    * How are you going to divide the work? Do you have a group leader, or convener? How often are you going to meet with each other?
    * For each group member (if you work on group projects): What is your role in the group project? What is your contribution so far? How do you plan to contribute further throughout the whole project?
  - Draft a plan for data analysis:
    * What statistical model are you going to use for the data analysis?
    * Have you done any preliminary analysis of the data?
    * What is your group's time line for (1) propose and fit a preliminary model, (2) examine initial results and adjust models further if needed, (3) prepare for a presentation, (4) summarize and write a final report.
    * What are the (potential) difficulties with your data analysis?

  Each group only needs to submit one single proposal with the names of all members.

- **Presentation**. For collaborative projects, all group members are supposed to work on the presentation together. However, you can choose to have one of the members deliver the whole presentation. The grading criteria for final project presentation is as follows.
  - Clear description of the problem (25%).
  - Clear description/visualization of the data (25%).
  - Clear and thorough description of statistical analysis (35%).
  - Clear and thorough interpretation of results (15%).

- **Report**. Each group needs to submit one thorough report (less than 10 pages, including figures and references), similar to the format of an academic paper. Annotated code which generates the results from the final project should also be submitted. Each member needs to submit a brief description of his/her contributions to the final project. The following grading criteria for the final project report is based on 100% maximum.
  - Solid understanding and description of problem (**+5%**).
  - Quality of data analysis.
    * Appropriateness of the statistical model (**+20%**).
    * Efficiency of the computational algorithms adopted (**+20%**).
  - Presentation of results, including visualization (**+15%**).
  - Proper interpretation of results (**+20%**).
  - Well-organized and clean code (**+5%**) .

- Clarity of written report (**+15%**).
- Irrelevance with course materials (**-40%**).
- Plagiarism of code or written report (**-100%**).
- Lack of contribution to group project (varies from **-80%** to **-20%**).

A single grade is given to all members of a group except the case when one or two members of a group is reported as "lack of contribution", see the last point above. **NO late final projects. Students who submit final projects late will receive a score of 0.** In case of severe medical reasons for late final project, a doctor's note and a signature from your graduate (or undergraduate) advisor is needed.

Miscellaneous

Academic Misconduct

The University of Michigan community functions best when its members treat one another with honesty, fairness, respect, and trust. The College of LSA promotes the assumption of personal responsibility and integrity, and prohibits all forms of academic dishonesty and misconduct. All cases of academic misconduct will be referred to the Office of the Assistant Dean for Undergraduate Education. Being found responsible for academic misconduct will usually result in a grade sanction, in addition to any sanction from the College. For more information, including examples of behaviors that are considered academic misconduct and potential sanctions, please see `www.lsa.umich.edu/academicintegrity`.

Accommodations for Students with Disabilities

If you think you need an accommodation for a disability, please let me know at your earliest convenience. Some aspects of this course, the assignments, the in-class activities, and the way the course is usually taught may be modified to facilitate your participation and progress. As soon as you make me aware of your needs, we can work with the Services for Students with Disabilities (SSD) office to help us determine appropriate academic accommodations. SSD (734-763-3000; http://ssd.umich.edu) typically recommends accommodations through a Verified Individualized Services and Accommodations (VISA) form. Any information you provide is private and confidential and will be treated as such.

Other resources

As a student, you may experience a range of issues that can negatively impact your learning, such as anxiety, depression, interpersonal or sexual violence, difficulty eating or sleeping, loss/grief, and/or alcohol/drug problems. These mental health concerns or stressful events may lead to diminished academic performance and affect your ability to participate in day-to-day activities. In order to support you during such challenging times, the University of Michigan provides a number of confidential resources to all enrolled students, including Counseling and Psychological Services (CAPS) (734.764.8312), Sexual Assault Prevention and Awareness Center (SAPAC) (24-Hour Crisis Line: 734.936.3333), Psychiatric Emergency Services (734-996-4747), and Services for Students with Disabilities (734.763.3000; 734.615.4461 [TDD]; 734.619.6661 [VP]).