POLSCI 798
Advanced Topics in Quantitative Methodology

Winter 2019

Yuki Shiraito ByungKoo Kim

Department of Political Science
University of Michigan

This course is designed to introduce students to modern Bayesian data analysis with emphasis on applications in social sciences. While the course is largely an applied course, it is intended to provide modeling and computational tools to its students so that they will be able to develop new applied models for analyzing original data in future research. Topics covered include Bayesian regression models, item response theory models, topic models, sampling methods, approximate Bayesian inference, and Bayesian nonparametrics. Prerequisites are POLSCI 599 and 699, or familiarity with basic mathematical statistics and regression analysis equivalent to these courses.

1 Contact Information

	Yuki Shiraito (Instructor)	ByungKoo Kim (GSI)
Office:	4259 ISR	7750 Haven
Phone:	734–615–5165	
Email:	shiraito@umich.edu	kimbk@umich.edu

2 Logistics

- Lectures: Mondays and Wednesdays, 8:30AM – 10:00AM, LSA 3207
- Sections: Wednesdays, 6:00PM – 7:00PM, Mason Hall G437
- Yuki's office hours:
  - *Hacker Place Office Hours*: Mondays, 4:00PM – 6:00PM, ISR 1450 (1440 on March 11)
  - Stop by anytime at the ISR office
- ByungKoo's office hours:
  - Wednesdays, 2:30 – 4:30PM, Haven Hall 7750
  - By appointment

3 Questions and Announcements

In addition to sections and office hours, please use the *Piazza Discussion Board* when asking questions about lectures, problem sets, and other course materials. This allows all students to benefit from the discussion and to help each other understand the materials. Both students and instructors are encouraged to participate in discussions and answer any questions that are posted.

To join the POLSCI 798 Piazza site, click on "Piazza" from the modules in the Canvas course site. You will then be prompted to create your account and confirm enrollment. Once you create your account, the Piazza course page can also be accessed by logging in from `https://piazza.com`. In addition, all class announcements will be made through Piazza. Canvas will still be used for hosting all class materials.

4 Prerequisites

There are three prerequisites for this course:

- Foundations of mathematical statistics covered in POLSCI 599 or equivalent. In particular, students are assumed to be familiar with the topics covered in Chapters 1–5, 6.1–3, 7.5–6, 8.1–5, 8.7–8 of

  Morris H. DeGroot and Mark J. Schervish. *Probability and Statistics.* Pearson Education, 4th edition, 2012.

  For probability theory, an alternative reference is Chapters 1–5, 7, 8.1–2, 9–10 of

  Joseph K. Blitzstein and Jessica Hwang. *Introduction to Probability.* Chapman and Hall/CRC, 2014.

- Regression analysis covered in POLSCI 699 or equivalent. In particular, students are assumed to be familiar with ordinary least squares (OLS), maximum likelihood estimation (MLE) of linear and discrete choice models, instrumental variables, and fixed and random effects models for panel data. For those topics, the following references are recommended:

  William H. Greene. *Econometric Analysis.* Prentice Hall, 7th edition, 2012.

  Jeffrey M. Wooldridge. *Econometric Analysis of Cross Section and Panel Data.* The MIT Press, 2nd edition, 2010.

- Proficiency in **R** at the level of

  Kosuke Imai. *Quantitative Social Science: An Introduction.* Princeton University Press, 2017.

5 Course Requirements

The final grades are based on the following items:

- **Participation** (10%): The level of engagement in lectures, sections, and Piazza discussions.
- **Problem sets** (40%): There will be several problem sets throughout the semester. Each problem set will equally contribute to the final grade and contain both analytical and data analysis questions. The following instructions will apply to all problem sets:
  - *Collaboration policy*: Collaboration on the problem sets is encouraged, because it provides opportunities for learning from each other. However, to facilitate individual learning, you should try all problem set questions before working together with other students, and you are required to submit your own answers for each problem set. Copying any part of someone else's code or answers is considered as academic misconduct.

- *Submission policy*: Submit your answers individually through Canvas. All answers must be typed via LaTeX and your **R** source code should be included as part of the LaTeX document. Please ensure your code adheres to the Google's **R** Style Guide rules available at `https://google.github.io/styleguide/Rguide.xml`. No late submission will be accepted, unless the instructor's permission is given in advance.

- **Final project** (50%): The final project must be reanalysis and extensions of a published study in the field of your interest. You are required to use Bayesian methods (not limited to the methods covered in the course) for the final project. If the class enrollment is greater than 10, the project will be a collaborative project with another student. In the project, you start with reanalysis of the empirical results that you choose. The chosen empirical study does not need to involve Bayesian methods, but your reanalysis should approach the study's empirical question in the Bayesian manner. After the reanalysis, your goal is to improve the original analysis either methodologically or substantively (or ideally both). During the entire process, the instructor and the GSI are available to help you with both substantive and technical questions.

  Be aware of the following key deadlines. Late submission will be penalized.

  - **January 27 (Project and collaborator identification):** By this date, you should identify your project (and collaborator, if applicable). Upload one paragraph (up to 150 words) description of your project to Canvas.
  - **February 10 (Data acquisition):** By this date, you should acquire the data to be analyzed (e.g., by downloading the data from a replication archive or requesting the data from the original authors). Upload your data set and a one page description of the data to Canvas.
  - **February 24 (Descriptive analysis):** By this date, you should finish your descriptive analysis. Turn in a brief summary of your descriptive analysis of the data (up to three pages including figures and tables). This memo should not simply list tables of descriptive statistics and graphs of variables. Rather, you should describe the aspect of the data upon which you aim to improve the original analysis. Look into the details of the data, and if there are any discrepancies or irregularities in the data, you should find them at this stage.
  - **March 10 (Reanalysis and proposed extensions):** By this date, you should finish your Bayesian reanalysis of the data (e.g., the Bayesian replication of the original results) and come up with the proposed extensions of the analysis. Turn in a brief summary of your replication and proposed extensions of the original analysis (up to five pages including tables and figures). Meet with the instructor to get feedback for your memo.
  - **April 7 (Initial write-up):** By this date, you should finish all of your empirical analyses and create all the tables and figures for presenting the results. You must electronically submit an initial write-up of your report to Canvas. This write-up should consist of the title, the abstract, the introduction, and the tables and figures with informative captions (up to 10 pages). Over the next few days, you will be asked to comment on the initial write-up by another (group of) student(s) in the class so that the author can improve the analysis and writing.
  - **April 14 (Written feedback):** You should electronically submit through Canvas your written feedback (up to two pages) on the assigned initial write-up. Your written feedback will be graded based on its quality (10% of the course grade). Your feedback has

to be constructive—meaning that you should not only point out caveats in the assigned write-up but also provide suggestions for how to improve its analysis or presentation.

  - **April 28 (Final paper):** You should write up your final paper reflecting the comments you received on your initial write-up. A pdf copy of the paper should be submitted to Canvas. The final paper will be graded based on its overall quality. Specifically, we will look at the novelty of substantive and methodological contributions as well as the effectiveness of presentation and writing. The paper should be no longer than 10,000 words, including the main body of text, notes, references, and the headers of tables and figures (not including the title page, abstract, or supporting information).

- **Incomplete Policy**: No incompletes will be given.

6 Hacker Place Office Hours

In this class, we will explore a new type of office hours. Instead of standard one-on-one meetings at my office, we are going to engage in weekly hackathons to create an open, inviting and supportive space for computational social science programming. The weekly hackathons will convene at ISR 1450 (Institute for Social Research, 426 Thompson Street) on Mondays from 4:00-6:00 PM (on March 11 only, hackers will meet in ISR 1440). To successfully participate in these weekly hackathons, students should bring their own laptops and identify the programming tasks they intend to work on. As the instructor accompanying our hackathons, I will offer assistance with regard to data analysis including, but not limited to, assignments for this course. While the instructor is available for help, participating students are more than welcome to help each other. In addition, this new hacker space will be open to attend for anyone in the ISR community at University of Michigan. Our objective is to engage graduate students, research fellows, and faculty to share their expertise via peer programming as well as to receive assistance in their own data-intensive projects. As such, this weekly hacker space is designed to benefit anyone in the ISR community who engages in computational social science projects and seeks to advance their programming skills, be it with regard to parallel computing in R, OpenMP and Rcpp, web scraping using Python, using high performance computing clusters, and other computational methods.

7 Other Course Policies

- **Student Sexual Misconduct Policy:** Title IX prohibits sex discrimination to include sexual misconduct: harassment, domestic and dating violence, sexual assault, and stalking. If you or someone you know has been harassed or assaulted, you can receive confidential support and academic advocacy at the Sexual Assault Prevention and Awareness Center (SAPAC). SAPAC can be contacted on their 24-hour crisis line, 734–936–3333 and online at sapac.umich.edu. Alleged violations can be reported non-confidentially to the Office for Institutional Equity (OIE) at institutional.equity@umich.edu. Reports to law enforcement can be made to University of Michigan Police Department at 734–763–3434.[1]

- **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

[1] This statement is taken from: `https://sapac.umich.edu/article/faculty-resources-sample-syllabus-language`.

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.[2]

- **Religious-Academic Conflicts:** While the university does not observe religious holidays, it is the policy of the University of Michigan to make every reasonable effort to allow members of the university community to observe their religious holidays without academic penalty. Absensce from classes or examinations for religious reasons does not relieve students from responsibility for any part of the course work required during the period ob absence. Students who expect to miss classes as a consequence of their religious observance shall be provided with a reasonable alternative opportunity to make-up missed academic work. It is the obligation of students to provide faculty with reasonable notice of the dates on which they will be absent. When the absence coincides with an exam or other assignment due date, the options to make up that missed work may be limited and will be determined by the instructor within the boundaries of the respective class.[3]

- **Academic Misconduct:** The University of Michigan community functions best when its members treat one another with honesty, fairness, respect, and trust. The college 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 https://lsa.umich.edu/lsa/academics/academic-integrity.html.[4]

- **Student Mental Health and Wellbeing:** The University of Michigan is committed to advancing the mental health and wellbeing of its students. If you or someone you know is feeling overwhelmed, depressed, and/or in need of support, services are available. For help, contact *Counseling and Psychological Services (CAPS)* at (734) 764-8312 and https://caps.umich.edu/ during and after hours, on weekends and holidays, or through its counselors physically located in schools on both North and Central Campus. You may also consult *University Health Service (UHS)* at (734) 764-8320 and https://www.uhs.umich.edu/mentalhealthsvcs, or for alcohol or drug concerns, see https://www.uhs.umich.edu/aodresources. For a listing of other mental health resources available on and off campus, visit: http://umich.edu/~health.[5]

8 Textbooks

There is no single required textbook for this course, but lectures are based on relevant chapters and sections of the following books:

[2] This statement is taken from: https://ssd.umich.edu/article/syllabus-statement.
[3] This statement is taken from: *Handbook for Faculty and Instructional Staff 2018*, p. 17.
[4] This statement is taken from: *Handbook for Faculty and Instructional Staff 2018*, p. 16.
[5] This statement is taken from: *Handbook for Faculty and Instructional Staff 2018*, p. 16.

Christopher M. Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006.

Kevin P. Murphy. *Machine Learning: A Probabilistic Perspective*. The MIT Press, 2012.

Andrew Gelman, John B. Carlin, Hal S. Stern, David B. Dunson, Aki Vehtari, and Donald B. Rubin. *Bayesian Data Analysis*. CRC Press, 3rd edition, 2014.

Also, the following books are useful for supplementary readings on some of the topics covered in the class:

P. McCullagh and J.A. Nelder. *Generalized Linear Models*. Chapman & Hall/CRC, 2nd edition, 1989.

Andrew Gelman and Jennifer Hill. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press, 2007.

Peter Müller, Fernando Andrés Quintana, Alejandro Jara, and Tim Hanson. *Bayesian Nonparametric Data Analysis*. Springer, 2015.

Richard McElreath. *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*. CRC Press, 2015.

9 Course Outline

9.1 Part I: Introduction to Bayesian Data Analysis

Part I covers the foundations of applied Bayesian data analysis in social sciences. The topics of this part are the building blocks that anyone conducting Bayesian statistics needs—simple models and computational algorithms.

Bayesian Models:

1. Election predictions
2. Bayes' theorem
3. Prior and posterior distributions
4. Monte Carlo sampling methods
5. The two-parameter Gaussian model

Supplementary readings: Gelman et. al. 1–3, and 10.

Bayesian Inference for Linear Regression:

1. Linear regression with Gaussian errors
2. The Gibbs sampler
3. Posterior predictive distributions
4. Asymptotics and Gaussian approximations
5. Comparison with maximum likelihood inference

Supplementary readings: Gelman et. al. 14, 11.1, 6, and 4.

Discrete Choice Models:

1. Probit models
2. Data augmentation
3. Ordered probit models
4. The Metropolis-Hastings algorithm
5. Multinomial probit models (time-permitting)

Supplementary readings: Murphy 9.4; Gelman et. al. 11–12.

Hierarchical Models:

1. Overdispersion
2. The Poisson and negative binomial models
3. Fixed and random effects models for grouped data
4. Robust standard errors under model misspecification
5. The incidental parameter problem
6. Hierarchical regression models

Supplementary readings: Gelman et. al. 5, 15, and 17. Journal articles will be assigned.

Measurement Models:

1. Ideal point estimation using roll call data
2. Item response theory models
3. EM algorithm and variational inference
4. Ideal point estimation of the U.S. Supreme Court judges
5. Dynamic linear models

Supplementary readings: Bishop 9–10; Murphy 12.1. Journal articles will be assigned.

9.2 Part II: Extensions

Part II covers a selection of more advanced topics listed below, as time permits. Note that it is hard to find existing empirical applications of some of the topics in social sciences. This part of the course aims to open up potential for the students' future research using these methods.

More data augmentation:

1. Logit models
2. Bayesian support vector machines

Bayesian variable selection:

1. Bayesian Lasso
2. Spike-and-slab regression

Networks:

1. Mixed membership stochastic blockmodels
2. Latent space models

Text:

1. Latent Dirichlet allocation and its variants

Mixture models:

1. Finite mixture models
2. Hidden Markov models
3. Change point models

Bayesian nonparametrics:

1. The Dirichlet process
2. The beta process.