POL-GA 2251: Quantitative Methods III

Fall 2021
Tuesdays 10:00-11:50am EST
Recitation: Thursdays 6:00-7:00pm EST
19 W. 4th Room 217

Instructor Information

Professor	Email	Office Hours	Sign-up
Tara Slough	tara.slough@nyu.edu	F 10 am - 12 pm or by appointment	Sign-up

Teaching Assistant	Email	Office Hours
Jiawei Fu	jf3739@nyu.edu	F 4 - 6 pm

Tara's office hours can be virtual (via Zoom) or in person (19 W. 4th, Office 415). Please indicate your preference on the sign-up. Jiawei's office hours will be virtual (via Zoom) except through prior arrangement.

Course Overview and Prerequisites[1]

The course covers a range of techniques for data analysis and modeling: maximum likelihood, Bayesian inference, and non-parametric methods ("machine learning"). The class will enable students to engage with current political science literature that uses advanced methods and to apply these methods in their own work. The course assumes working knowledge of probability theory, matrix algebra, calculus, and statistical inference at the level of Quant I and Quant II. The course is restricted to NYU PhD students in the Department of Politics.

Expectations

Reading and participation (20%):
You are expected to read all required readings prior to the class in which they are discussed. If you do not understand the readings, please come prepared to discuss what you did and did not follow.

Problem sets (20%):
During the semester, there will be 10 problem sets. The problem sets are relatively short but will be assigned almost weekly. Problem sets will be released on Tuesdays immediately after class and are due the following Tuesday by 10:00 am (before class). You can find and submit your problem sets on the course Brightspace.

Replication and Extension (20%):
At the end of the semester, you should submit a short report (1,000-2,000 words) and code replicating a political science paper that uses any method covered in this class. You should sign up

[1] This class is designed after Arturas Rozenas' Quant III. Many aspects of this syllabus are inspired by his 2020 syllabus.

with the article you intend to replicate and confirm that replication data is available by **November 2**. We will cover some ideas for the replication and extension in class in November.

In-Class Midterm (20%)
There will be an in-class midterm on October 19. The exam is open book and open note, but will be timed. You will not be required to write any code for the midterm.

Take-Home Final (20%)
There will be a take-home exam administered on December 14 (in lieu of the last class session). This exam is open note and open book, but you may not communicate with class members. There may be questions that require you to use R on the final.

Course Materials

Textbooks

We will use two textbooks frequently throughout the course:

- **[ISL]** Gareth James et al. 2013. *An Introduction to Statistical Learning: with Applications in R*. New York: Springer.

- **[BDA]** Andrew Gelman et al. 2013. *Bayesian Data Analysis*. Third Edition. Boca Raton, FL: CRC Press.

ISL is a popular introductory text on non-parametric methods with examples in `R`. BDA is the most popular textbook on Bayesian statistics and data analysis. I recommend that you purchase these books if you like to have physical copies. You can access the full text of ISL online through NYU Libraries. BDA is available for non-commercial use at `http://www.stat.columbia.edu/~gelman/book/BDA3.pdf`. The following textbooks may also be useful for reference, particularly if you choose to delve further into topics covered in this course:

- **[Pawitan]** Yudi Pawitan. 2013. *In All Likelihood: Statistical Modeling and Inference Using Likelihood*. New York: Oxford University Press. This is an accessible introduction to maximum likelihood estimation methods with examples. You can access the full text online through NYU Libraries.

- **[Greene]** William H. Greene. 2010a. *Econometric Analysis*. Seventh Edition. Boston: Prentice Hall. This is a standard econometrics text with coverage of many of the topics in this course.

- **[McElreath]** Richard McElreath. 2016. *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*. Boca Raton, FL: CRC Press. This text covers Bayesian data analysis with many examples in `R` and Stan. You can access the full text online through NYU Libraries.

- Trevor Hastie, Robert Tibshirani, and Jerome Friedman. 2001. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Second Edition. Springer. This is the main textbook upon which ISL is widely used in the “machine learning” community. You can access the full text online through NYU Libraries.

Software

This class will use `R`, and prior experience at the level of Quant 1 and Quant 2 is assumed.

Schedule

The course meets Tuesdays from 4:00-5:50pm. The schedule for the semester is summarized in the table below:

			Problem sets	
Week	Date	Topic	Assigned	Due
1	September 7	Introduction	1	–
2	September 14	Maximum Likelihood Estimation: Theory	2	1
3	September 21	Maximum Likelihood Estimation: Applications I	3	2
4	September 28	Maximum Likelihood Estimation: Applications II	4	3
5	October 5	Bayesian Inference: Theory	5	4
6	October 12	Bayesian Inference: Applications	–	5
7	October 19	**In-class midterm**		
8	October 26	Bayesian Inference: Computation	6	–
9	November 2	Latent Variable Models	7	6
10	November 9	Mixture Models	8	7
11	November 16	High-Dimensional Data	9	8
12	November 23	Structural vs. Reduced-Form Approaches	–	9
13	November 30	Semi- and Non-Parametric Regression	10	–
14	December 7	Tree-Based Models	–	10
15	December 14	**Take-home final**		
	December 22	**Replication due**		

Course and assignment schedule. Assignments will be released after class and are due before class on the following Tuesday.

A detailed schedule of topics and readings, by class session, appears below. You are responsible for reading all readings in the "required readings" lists prior to the class section.

September 7: Introduction

Note: The first section will be virtual.

- Required readings:
  1. ISL chapter 2
  2. BDA chapter 1
- Additional readings:
  1. Michael D. Ward, Brian D. Greenhill, and Kristin M. Bakke. 2010. "The perils of policy by p-value: Predicting civil conflicts". *Journal of Peace Research* 47 (4): 363–375

September 14: Maximum Likelihood Estimation (MLE): Theory

- Required readings:

1. Pawitan chapters 2 and 4.

- Additional readings:

  1. Greene, chapter 14.

September 21: Maximum Likelihood Estimation (MLE): Applications I

- Required readings:

  1. ISL chapter 4
  2. Michael J. Hanmer and Kerem Ozan Kalkan. 2012. "Behind the Curve: Clarifying the Best Approach to Calculating Predicted Probabilities and Marginal Effects from Limited Dependent Variable Models". *American Journal of Political Science* 57 (1): 263–277

- Additional readings:

  1. Curtis S. Signorino and Kuzey Yilmaz. 2003. "Strategic Misspecification in Regression Models". *American Journal of Political Science* 47 (3): 551–566
  2. Christopher Zorn. 2005. "A Solution to Separation in Binary Response Models". *Political Analysis* 13:157–170
  3. Justin Esarey and Andrew Pierce. 2017. "Assessing Fit Quality and Testing for Misspecification in Binary-Dependent Variable Models". *Political Analysis* 20 (4): 480–500

September 28: Maximum Likelihood Estimation (MLE): Applications II

- Required readings:

  1. ISL chapter 4
  2. William D. Berry, Jacqueline H.R. DeMeritt, and Justin Esarey. 2010. "Testing for Interaction in Binary Logit and Probit Models: Is a Product Term Essential?" *American Journal of Political Science* 54 (1): 248–266

- Additional readings:

  1. William H. Greene. 2010b. "Testing hypotheses about interaction terms in nonlinear models". *Economics Letters* 107 (2): 291–296
  2. Ted Enamorado, Benjamin Fifield, and Kosuke Imai. 2019. "Using a Probabilistic Model to Assist Merging of Large-Scale Administrative Records". *American Political Science Review* 113 (2): 353–371

October 5: Bayesian Inference: Theory

- Required readings:

  1. BDA chapters 2-3

- Additional readings:

1. McElreath chapters 1-3
2. Andrew D. Martin. 2008. "The Oxford Handbook of Political Methodology". Chap. Bayesian Analysis, ed. by Janet M. Box-Steffensmeier, Henry E. Brady, and David Collier, 494–510. Oxford University Press
3. Jeff Gill and Lee D. Walker. 2005. "Elicited Priors for Bayesian Model Specifications in Political Science Research". *Journal of Politics* 67 (3): 841–872

October 12: Bayesian Inference: Applications

- Required readings:
  1. BDA chapters 14-15
- Additional readings:
  1. David K. Park, Andrew Gelman, and Joseph Bafumi. 2004. "Bayesian Multilevel Estimation with Poststratification: State-Level Estimates from National Polls". *Political Analysis* 12:375–385
  2. Macartan Humphreys and Alan Jacobs. 2015. "Mixing methods: A Bayesian approach". *American Political Science Review* 109 (4): 653–673
  3. Andrew Gelman. 2014. "How Bayesian Analysis Cracked the Red-State, Blue-State Problem". *Statistical Science* 29 (1): 26–35

October 19: In-Class Midterm

Note: There will be no section on October 21.

October 26: Bayesian Inference: Computation

- Required readings:
  1. McElreath chapter 8
  2. BDA chapters 11-12
- Additional readings:
  1. Simon Jackman. 2000. "Estimation and Inference Are Missing Data Problems: Unifying Social Science Statistics via Bayesian Simulation". *Political Analysis* 8 (4): 307–332

November 2: Latent Variables

Note: Chris Fariss will join us virtually to guest lecture for part of this class meeting.

- Required readings:
  1. Christopher J. Fariss, Kevin Reuning, and Michael R. Kenwick. 2020. "SAGE Handbook of Research Methods is Political Science and International Relations". Chap. Measurement Models, ed. by Luigi Curini and Robert J. Franzese Jr., 353–370. SAGE London

2. Christopher J. Fariss. 2014. "Respect for human rights has improved over time: Modeling the changing standard of accountability". *American Political Science Review* 108 (2): 297–318

- Additional readings:

  1. Simon Jackman. 2008. "The Oxford Handbook of Political Methodology". Chap. Measurement, ed. by Janet M. Box-Steffensmeier, Henry E. Brady, and David Collier, 120–150. Oxford University Press
  2. Joseph Bafumi et al. 2005. "Practical Issues in Implementing and Understanding Bayesian Ideal Point Estimation". *Political Analysis* 13 (2): 171–187
  3. Pablo Barberá. 2015. "Birds of the Same Feather Tweet Together. Bayesian Ideal Point Estimation Using Twitter Data." *Political Analysis* 23 (1): 76–91

November 9: Mixture Models

- Required readings:

  1. BDA chapter 22
  2. Kosuke Imai and Dustin Tingley. 2011. "A Statistical Method for Empirical Testing of Competing Theories". *American Journal of Political Science* 56 (1): 218–236

- Additional readings:

  1. John Ahlquist and Christian Breuning. 2012. "Model-based Clustering and Typologies in the Social Sciences". *Political Analysis* 20:92–112
  2. Benjamin E Bagozzi and Bumba Mukherjee. 2012. "A Mixture Model for Middle Category Inflation in Ordered Survey Responses". *Political Analysis* 20:369–386
  3. Nils B. Weidmann. 2011. "Violence "from above" or "from below"? The Role of Ethnicity in Bosnia's Civil War". *Journal of Politics* 73 (4): 1178–1190

November 16: High-Dimensional Data

- Required readings:

  1. ISL 5-6

- Additional readings:

  1. Alexandre Belloni et al. 2012. "Sparse Models and Methods for Optimal Instruments with an Application to Eminent Domain". *Econometrica* 80 (6): 2369–2429
  2. Nicholas Beauchamp. 2016. "Predicting and Interpolating State-Level Polls Using Twitter Textual Data". *American Journal of Political Science* 61 (2): 490–503
  3. Connor Huff and Josh D. Kertzer. 2017. "How the Public Defines Terrorism". *American Journal of Political Science* 62 (1): 55–71

November 23: Structural vs. Reduced Form Models

Note: This class session will (provisionally) be held virtually. We will focus on structural and reduced form interpretation of conjoint surveys for preference elicitation. There will be no section on November 25.

- Required readings:
  1. Matthew Wiswall and Basit Zafar. 2017. "Preference for the Workplace, Investment in Human Capital, and Gender". *Quarterly Journal of Economics* 133 (1): 457–507
- Additional readings:
  1. Jens Hainmueller, Daniel J. Hopkins, and Teppei Yamamoto. 2017. "Causal Inference in Conjoint Analysis: Understanding Multidimensional Choices via Stated Preference Experiments". *Political Analysis* 22 (1): 1–30
  2. loewenetal2012. 2010. "Testing the power of arguments in referendums: A Bradley-Terry approach". *Electoral Studies* 31:212–221
  3. Scott F. Abramson, Korhan Kocak, and Asya Magazinnik. 2020. "What Do We Learn about Voter Preferences from Conjoint Experiments?" Working paper, University of Rochester `https://tinyurl.com/au8yt75z`

November 30: Semi- and Non-Parametric Regression

- Required readings:
  1. BDA chapter 20
  2. ISL chapter 7
- Additional readings:
  1. Jens Hainmueller and Chad Hazlett. 2014. "Kernel Regularized Least Squares: Reducing Misspecification Bias with a Flexible and Interpretable Machine Learning Approach". *Political Analysis* 22 (2): 143–168
  2. Nathaniel Beck and Simon Jackman. 1998. "Beyond linearity by default: Generalized additive models". *American Journal of Political Science* 42 (2): 596–627

December 7: Tree-Based Models

- Required readings:
  1. ISL chapter 8
  2. Jennifer Hill. 2011. "Bayesian Nonparametric Modeling for Causal Inference". *Journal of Computational and Graphical Statistics* 20:217–240
- Additional readings:
  1. David Muchlinski et al. 2017. "Comparing Random Forest with Logistic Regression for Predicting Class-Imbalanced Civil War Onset Data". *Political Analysis* 24 (1): 87–103

2. Zachary M. Jones and Yonatan Lupu. 2018. "Is Therre More Violence in the Middle?" *American Journal of Political Science* 62 (3): 652–667
3. Sayash Kapoor and Arvind Narayanan. 2021. "(Ir)Reproducible Machine Learning: A Case Study". Working paper, Princeton University. `https://reproducible.cs.princeton.edu/irreproducibility-paper.pdf`
4. James Bisbee. 2019. "BARP: Improving Mister P Using Bayesian Additive Regression Trrees". *American Political Science Review* 113 (4): 1060–1065