

New York University  
Wilf Family Department of Politics

## Quantitative Methods III (Fall 2020)

Instructor: Professor Arturas Rozenas

Seminar time: Tuesdays 9:00-10:50

Seminar location: Room 217 or <https://nyu.zoom.us/j/93182237865>

Office hours: Tuesdays 12:00-2:00, Room 411 or <https://nyu.zoom.us/j/6109047457>

Email: ar199@nyu.edu

## Overview & Prerequisites

The course covers a range of techniques for data analysis and modeling: maximum likelihood, Bayesian inference, and non-parametric methods (“machine learning”). The class is designed to enable students to read the state of the art political science literature that uses advanced methods and to apply these methods in their own work. The course requires working knowledge of the 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.

## Textbooks & Readings

Chapters from the following two books will be frequently assigned:

- (ISL) G. James, D. Witten, T. Hastie, and R. Tibshirani, *An Introduction to Statistical Learning in R*, Springer, 2013.
- (BDA) Andrew Gelman et al. *Bayesian Data Analysis*. 3rd ed. Boca Raton: CRC Press, 2014

The first one (ISL) is a very applied, accessible, and highly popular introduction to non-parametric methods with nice examples in *R*. The second one (BDA) is the most popular textbook on Bayesian data analysis. It is recommended that you purchase these books. When chapters from other books are assigned, scanned copies will be uploaded to the course website. In addition, it might be useful to have access to and consult the following textbooks:

- Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The elements of statistical learning: Data Mining, Inference, and Prediction*. 2nd ed. Springer series in statistics New York, NY, USA:, 2001. The main textbook on which ISL is based; widely used reference text in the ‘machine learning’ community.
- Yudi Pawitan. *In All Likelihood: Statistical Modeling and Inference Using Likelihood*. Oxford and New York: Oxford University Press, 2013. An accessible introduction to MLE methods, that has a good coverage of theory and practical examples. The book is available online through NYU libraries.
- George Casella and Roger S. Berger. *Statistical Inference*. Duxbury: Thompson Learning, 2002. A standard masters-level textbook on statistical inference, covers mostly prerequisites for this class plus some of the MLE and robust inference material.
- William H Greene. *Econometric Analysis*. 7th ed. Boston, MA: Prentice Hall, 2010. Widely used PhD-level econometrics textbook, covers many topic in this class, but with greater emphasis on panel data and causal inference.
- Christian P. Robert and George Casella. *Monte Carlo Statistical Methods*. 2nd ed. Springer, 2004. A very convenient textbook on simulation methods (technical).

# Course Material

Syllabus, homework assignments, and other course materials will be hosted on the [Dropbox folder](#).

## Software

The class will use R, and it is assumed that you have some prior experience with the language.

## Grading

- In-class participation (20%). You have to read the required material. I may ask questions about the readings. If you don't understand something in the readings, you should be able to explain what you tried to understand.
- Homework assignments (20%). You will be asked to submit 10 homework problem sets throughout the semester; basically, every week, starting with the third week of the semester.
- Replication (20%): At the end of the semester, you will have to submit a short report replicating results of a political science paper that uses the methods covered in this class. You do not need to replicate all of the results but choose the ones that are most interesting/challenging. To make this worthwhile, pick a paper/result that utilizes more challenging methods, rather than simple ones.
- In-class midterm test (20%).
- Take home final test (20%).

## Schedule

### 1. (Sept 8) Introduction

Course overview, estimation and inference, classical and Bayesian inference, precision-interpretability trade-off, bias-variance trade-off, supervised vs unsupervised learning.

Required readings:

- BDA-1, ISL-2

Further readings:

- Michael D Ward, Brian D Greenhill, and Kristin M Bakke. "The perils of policy by p-value: Predicting civil conflicts". In: *Journal of Peace Research* 47.4 (2010), pp. 363–375

### 2. (Sep 15) MLE: theory

Analytical and numerical optimization, score function, Fisher's information, asymptotic properties of the MLE's, likelihood inference, GLM

Required readings:

- Pawitan, Chs. 2, 4.

Further readings:

- William H Greene. *Econometric Analysis*. 7th ed. Boston, MA: Prentice Hall, 2010, Ch. 14.

### 3. (Sep 22) MLE: Applications (I)

Binary data models, estimation, interpretation, separability problem.

Required readings:

- ISL-4

Further readings:

- Christopher Zorn. “A solution to separation in binary response models”. In: *Political Analysis* 13.2 (2005), pp. 157–170
- C. Signorino and K. Yilmaz. “Strategic Misspecification in Regression Models”. In: *American Journal of Political Science* 47.3 (2003), p. 551
- Michael J Hanmer and Kerem Ozan Kalkan. “Behind the curve: Clarifying the best approach to calculating predicted probabilities and marginal effects from limited dependent variable models”. In: *American Journal of Political Science* 57.1 (2013), pp. 263–277

### 4. (Sept 29) MLE: Applications (II)

Ordinal, count, and multinomial data, over-dispersion, structural zeros, application to probabilistic record linkage.

Required readings:

- ISL-4
- William D Berry, Jacqueline HR DeMeritt, and Justin Esarey. “Testing for interaction in binary logit and probit models: is a product term essential?” In: *American Journal of Political Science* 54.1 (2010), pp. 248–266

Further readings:

- Justin Esarey and Andrew Pierce. “Assessing Fit Quality and Testing for Misspecification in Binary-Dependent Variable Models”. In: *Political Analysis* 20.4 (2012), pp. 480–500
- William Greene. “Testing hypotheses about interaction terms in nonlinear models”. In: *Economics Letters* 107.2 (2010), pp. 291–296

### 5. (Oct 6) Bayesian Inference: Theory

Principles of Bayesian inference and terminology, conjugacy, constructing uninformative and weakly informative priors, analytical posterior computation, interpretation, shrinkage, credible intervals

Required readings:

- BDA-2/3

Further readings:

- Jeff Gill and Lee D Walker. “Elicited priors for Bayesian model specifications in political science research”. In: *The Journal of Politics* 67.3 (2005), pp. 841–872
- Andrew Gelman et al. “Objections to Bayesian statistics”. In: *Bayesian Analysis* 3.3 (2008), pp. 445–449
- Martin, Andrew (2008) ”Bayesian Analysis.” In *The Oxford Handbook of Political Methodology*, eds. Janet M. Box-Steffensmeier, Henry E. Brady, and David Collier

## 6. (Oct 13) Bayesian Inference: Applications

Linear Bayesian regression, hierarchical models, multi-level regression with post-stratification, Bayesian GLM's, using qualitative information to improve causal inference.

### Required readings:

- BDA-14/15

### Further readings:

- Macartan Humphreys and Alan M Jacobs. "Mixing methods: A Bayesian approach". In: *American Political Science Review* 109.4 (2015), pp. 653–673
- Gelman, Andrew (2014) "How Bayesian Analysis Cracked the Red-State, Blue-State Problem." *Statistical Science* 29(1): 26-35.
- Park, David K., Andrew Gelman, and Joseph Bafumi. (2004) "Bayesian Multilevel Estimation with Poststratification: State-Level Estimates from National Polls." *Political Analysis* 12:375-385.

## 7. (Oct 20) Review / Midterm

## 8. (Oct 27) Bayesian Inference: Computation

Introduction to Monte Carlo integration, Metropolis-Hastings and Gibbs sampler with examples in R and `rstan`.

### Required readings:

- BDA-11/12

### Further readings:

- Simon Jackman. "Estimation and inference are missing data problems: Unifying social science statistics via Bayesian simulation". In: *Political Analysis* 8.4 (2000), pp. 307–332.

## 9. (Nov 3) Latent Variables

Latent classification, ideal point estimation, principled measure aggregation.

### Required readings:

- Jackman, Simon (2008) "Measurement." In *The Oxford Handbook of Political Methodology*, eds. Janet M. Box-Steffensmeier, Henry E. Brady, and David Collier

### Further readings:

- Christopher J Fariss. "Respect for human rights has improved over time: Modeling the changing standard of accountability". In: *American Political Science Review* 108.2 (2014), pp. 297–318
- Shawn Treier and Simon Jackman. "Democracy as a Latent Variable". In: *American Journal of Political Science* 52.1 (2008), pp. 201–17
- Royce Carroll et al. "The Structure of Utility in Spatial Models of Voting". In: *American Journal of Political Science* 57.4 (2013), pp. 1008–1028
- Mark R Beissinger. "The Semblance of Democratic Revolution: Coalitions in Ukraine's Orange Revolution". In: *American Political Science Review* 107.03 (2013), pp. 574–592
- Joseph Bafumi et al. "Practical Issues in Implementing and Understanding Bayesian Ideal Point Estimation". In: *Political Analysis* 13.2 (2005), pp. 171–187

## 10. (Nov 10) Mixture Models

Density estimation using mixtures, EM algorithm, finite gaussian mixtures for regression analysis, mixture models for random effects, parametric flexibility and robustness, split-population models, profile regression, component selection

### Required readings:

- BDA-22
- Kosuke Imai and Dustin Tingley. “A statistical method for empirical testing of competing theories”. In: *American Journal of Political Science* 56.1 (2012), pp. 218–236.

### Further readings:

- John S Ahlquist and Christian Breunig. “Model-based clustering and typologies in the social sciences”. In: *Political Analysis* 20.1 (2012), pp. 92–112
- Benjamin E Bagozzi and Bumba Mukherjee. “A mixture model for middle category inflation in ordered survey responses”. In: *Political Analysis* 20.3 (2012), pp. 369–386
- Jong Hee Park. “Changepoint analysis of binary and ordinal probit models: An application to bank rate policy under the interwar gold standard”. In: *Political Analysis* 19.2 (2011), pp. 188–204
- Nils B Weidmann. “Violence “from above” or “from below”? The Role of Ethnicity in Bosnia’s Civil War”. In: *The Journal of Politics* 73.04 (2011), pp. 1178–1190

## 11. (Nov 17) High-Dimensional Data

Bias-variance trade-off, resampling methods (cross-validation and bootstrapping), regularization, ridge/lasso regression.

### Required readings:

- ISL-5/6

### Further readings:

- Alexandre Belloni et al. “Sparse models and methods for optimal instruments with an application to eminent domain”. In: *Econometrica* 80.6 (2012), pp. 2369–2429
- Nicholas Beauchamp. “Predicting and interpolating state-level polls using Twitter textual data”. In: *American Journal of Political Science* 61.2 (2017), pp. 490–503
- Connor Huff and Joshua D Kertzer. “How the public defines terrorism”. In: *American Journal of Political Science* 62.1 (2018), pp. 55–71
- Bruce A Desmarais, Raymond J La Raja, and Michael S Kowal. “The fates of challengers in us house elections: The role of extended party networks in supporting candidates and shaping electoral outcomes”. In: *American Journal of Political Science* 59.1 (2015), pp. 194–211

## 12. (Nov 24) Structural topic model of language (no recitation)

Structural topic models encompass a great deal of the material covered so far in the class (Bayesian inference, mixture models, high-dimensional data)

### Required readings:

- Margaret E Roberts et al. “Structural topic models for open-ended survey responses”. In: *American Journal of Political Science* 58.4 (2014), pp. 1064–1082

## 13. (Dec 3) Semi- and Non-parametric Regression

Generalized additive models, spline basis functions, penalized spline regression, cross-validation, dimensionality problem, kernel regularization methods, Bayesian Gaussian process regression.

Required readings:

- BDA-20, ISL-7

Further readings:

- Jens Hainmueller and Chad Hazlett. “Kernel Regularized Least Squares: Reducing Misspecification Bias with a Flexible and Interpretable Machine Learning Approach”. In: *Political Analysis* 22.2 (2014), pp. 143–168.
- Simon Wood. *Generalized additive models: an introduction with R*. Boca Raton: CRC press, 2006.
- Nathaniel Beck and Simon Jackman. “Beyond linearity by default: Generalized additive models”. In: *American Journal of Political Science* (1998), pp. 596–627.

## 14. (Dec 10) Tree-Based Methods

Trees for classification and regression, bagging, random forests, applications for causal inference and prediction.

Required readings:

- ISL-8
- Jennifer L Hill. “Bayesian nonparametric modeling for causal inference”. In: *Journal of Computational and Graphical Statistics* 20.1 (2011), pp. 217–240

Further readings:

- David S Siroky et al. “Navigating random forests and related advances in algorithmic modeling”. In: *Statistics Surveys* 3 (2009), pp. 147–163
- David Muchlinski et al. “Comparing random forest with logistic regression for predicting class-imbalanced civil war onset data”. In: *Political Analysis* 24.1 (2016), pp. 87–103
- Zachary M Jones and Yonatan Lupu. “Is There More Violence in the Middle?” In: *American Journal of Political Science* 62.3 (2018), pp. 652–667