POL 574: Quantitative Analysis IV

Spring 2009

Kosuke Imai

1 Contact Information

Office: Corwin Hall 041
Office Phone: 258–6601
Email: kimai@Princeton.Edu
URL: http://imai.princeton.edu

2 Logistics

Fridays 2 – 4:50pm in Corwin 127

3 Course Description

The main goal of this course is to help students to write a publishable paper that uses advanced statistical methods. At the beginning of the semester, I will give brief introductory lectures on causal inference and applied Bayesian statistics to cover the fundamentals. Thereafter the materials covered will focus on the statistical methods appropriate for the projects selected by students.

4 Prerequisites

There are three prerequisites for this course.

1. Mathematics covered in POL 502
2. Probability and statistics covered in POL 571, POL 572, and POL 573
3. Statistical computing and programming covered in the statistical software workshop held at the end of January

If you are unsure about whether you have a sufficient background, please talk to me as soon as possible.

5 Course Requirements

The final grades are based on the following two items:

1. Participation and presentations (40%)
2. Final paper (60%) due on the Dean's Date

Students are required to read the assigned papers and monographs prior to weekly meetings where we discuss the readings in detail to gain a better understanding of each topic. The student doing research related to the topic will be lead discussant but others are also expected to do the readings and participate in the discussions. At the beginning and end of the semester, each student gives a brief presentation reporting the proposal and progress of their project. I will advise students on next steps beyond the course towards the goal of submitting their written work to a peer-review journal for publication.

6 Possible Topics and Suggested Readings

Here are possible course topics (along with suggested readings) based on the emails I have received prior to the semester. Of course, depending on the empirical projects chosen by students, we can add new topics. They are listed in the alphabetical order. Many of the suggested readings are not very easy, and students are not expected to understand everything at their first reading. The goal of the course is to help understand the main materials presented in these papers without paying too much attention to mathematical details.

General Topics

1. Longitudinal data analysis

   Parts II and III of Garrett M. Fitzmaurice, Nan M. Laird, and James H. Ware. *Applied Longitudinal Analysis*. Wiley, 2004.

2. Missing data

   Part II of Roderick J. A. Little and Donald B. Rubin. *Statistical Analysis with Missing Data*. 2nd ed. Wiley, 2002.

3. Multilevel modeling

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

More Specific Topics

1. Difference-in-differences models

   Marianne Bertrand, Esther Duflo, and Sendhil Mullainathan. How much should we trust differences-in-differences estimates? *Quarterly Journal of Economics*, 119(1):249–275, 2004.

   Alberto Abadie. Semiparametric difference-in-differences estimators. *Review of Economic Studies*, 72:1–19, 2005.

   Susan Athey and Guido W. Imbens. Identification and inference in nonlinear difference-in-differences models. *Econometrica*, 74(2):431–497, March 2006.

2. Bayesian causal inference

Keisuke Hirano, Guido W. Imbens, Donald B. Rubin, and Xiao-Hua Zhou. Assessing the effect of an influenza vaccine in an encouragement design. *Biostatistics*, 1(1):69–88, 2000.

John Barnard, Constantine E. Frangakis, Jennifer L. Hill, and Donald B. Rubin. Principal stratification approach to broken randomized experiments: A case study of school choice vouchers in New York (with discussion). *Journal of the American Statistical Association*, 98(462):299–311, 2003.

Hui Jin and Donald B. Rubin. Principal stratification for causal inference with extended partial compliance. *Journal of the American Statistical Association*, 103(481):101–111, March 2008.

3. Causal inference using longitudinal data

James M. Robins, Miguel Ángel Hernán, and Babette Brumback. Marginal structural models and causal inference in epidemiology. *Epidemiology*, 11(5):550–560, September 2000.

Miguel A. Hernán, Babette Brumback, and James M. Robins. Marginal structural models to estimate the joint causal effect of nonrandomized treatments. *Journal of the American Statistical Association*, 96(454):440–448, June 2001.

Marshall M. Joffe, Thomas R. Ten Have, Harold I Feldman, and Stephen E. Kimmel. Model selection, confounder control, and marginal structural models: Review and new applications. *American Statistician*, 58(4):272–279, November 2004.

4. Competing risks models

Martin Crowder. *Classical Competing Risks*. Chapman & Hall, London, 2001.

John D. Kalbfleisch and Ross L. Prentice. *The Statistical Analysis of Failure Time Data*. Wiley & Sons, Hoboken, New Jersey, 2nd edition, 2002.

5. Doubly robust inference for missing data and causal inference

James R. Carpenter, Michael G. Kenward, and Stijn Vansteelandt. A comparison of multiple imputation and doubly robust estimation for analyses with missing data. *Journal of the Royal Statistical Society, Series A*, 169(3):571–584, 2006.

Joseph D.Y. Kang and Joseph L. Schafer. Demystifying double robustness: A comparison of alternative strategies for estimating a population mean from incomplete data (with discussions). *Statistical Science*, 22(4):523–539, November 2007.

Zhiqiang Tan. Comment: Understanding OR, PS, and DR. *Statistical Science*, 22(4):560–568, 2007.

6. Interference between units in causal inference

Michael E. Sobel. What do randomized studies of housing mobility demonstrate?: Causal inference in the face of interference. *Journal of the American Statistical Association*, 101 (476):1398–1407, 2006.

Paul R. Rosenbaum. Interference beteween units in randomized experiments. *Journal of the American Statistical Association*, 102(477):191–200, 2007.

Michael G. Hudgens and Elizabeth Halloran. Toward causal inference with interference. *Journal of the American Statistical Association*, 103(482):832–842, June 2008.

7. Most recent advances in matching methods

Alberto Abadie, Alexis Diamond, and Jens Hainmueller. Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program. Technical report, Harvard University, Cambridge, MA, 2007.

Paul R. Rosenbaum, Richard N. Ross, and Jeffrey H. Silber. Minimum distance matched sampling with fine balance in an observational study of treatment for ovarian cancer. *Journal of the American Statistical Association*, 102(477):75–83, March 2007.

Stefano M. Iacus, Gary King, and Giuseppe Porro. Matching for causal inference without balance checking. *Working Paper*, 2008.

8. Necessary and sufficient causes

James Robins and Sander Greenland. The probability of causation under a stochastic model for individual risk. *Biometrics*, 45:1125—1138, 1989.

Jin Tian and Judea Pearl. Probabilities of causation: Bounds and identification. *Annals of Mathematics and Artificial Intelligence*, 28:287–313, 2000.

9. Regression discontinuity design

Jinyong Hahn, Petra Todd, and Wilbert Van der Klaauw. Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica*, 69(1):201–209, January 2001.

Guido W. Imbens and Thomas Lemieux. Regression discontinuity designs: A guide to practice. *Jounal of Econometrics*, 142(2):615–635, February 2008.

David S. Lee. Randomized experiments from non-random selection in U.S. house elections. *Journal of Econometrics*, 142(2):675–697, February 2008.

10. Social network analysis

Peter D. Hoff, Adrian E. Raftery, and Mark S. Handcock. Latent space approaches to social network analysis. *Journal of the American Statistical Association*, 97(460):1090–1097, December 2002.

Peter Hoff and Michael Ward. Modeling dependencies in international relations networks. *Political Analysis*, 12(2):160–175, Spring 2004.

Mark S. Handcock, Adrian E. Raftery, and Jeremy M. Tantrum. Model-based clustering for social networks. *Journal of the Royal Statistical Society, Series A (with discussions)*, 170 (2):301–354, 2007.

11. Statistics and game theory

Thomas Palfrey and Richard McKelvey. Quantal response equilibria in extensive form games. *Experimental Economics*, 1:9–41, 1998.

Curtis Signorino and Kuzey Yilmaz. Strategic misspecification in regression models. *American Journal of Political Science*, 47(3):551–566, 2003.

Mihammet Bas, Curtis Signorino, and Robert Walker. Statistical backwards induction: A simple method for estimating recursive strategic models. *Political Analysis*, 16:21–40, 2008.