Theoretical and Methodological Foundations

Empirical Implications of Theoretical Models Summer Institute
Monday, June 15, through Thursday, June 18, 2009
Washington University in St. Louis

Randall Calvert Andrew D. Martin

Course Description

Although most participants in the Summer Institute know the basics of rational choice theory and of statistical analysis, it is useful to review and supplement some basic techniques from a standpoint that will prepare participants for the advanced seminars, and for the general project of uniting game theoretic and statistical tools. The Foundations seminar presents important varieties of rational choice models, specifically spatial voting models and non-cooperative game theory, in a form that emphasizes the techniques by which these models can be used to generate testable implications through rigorous analysis of outcome and equilibrium correspondences. It reviews statistical techniques particularly widely suited to such applications. More specifically, we examine:

- deriving testable predictions from game theoretic models
- sources of statistical variation in game theoretic models
- the implications of the theoretical model for statistical properties
- statistical techniques suitable for testing those implications, such as probability model specifications and Bayesian statistics
- common useful inferences and hypothesis tests based on game-theoretic models

Game theoretic models lend themselves naturally to statistical analysis. Such analysis takes two broad approaches. In general, the source of testable hypotheses is the correspondence between a game's *parameters* (such as payoff parameters and chance-move probabilities) and the characteristics of the game's equilibrium *outcomes*. One approach to statistical testing of theoretical implications, then, is to derive robust predictions from this parameter-outcome correspondence, and to test them using robust statistical techniques. By "robust," we mean predictions and techniques likely to be insensitive to small inaccuracies in the modeling assumptions. For example, an appropriately formulated game can generate regular properties of *monotone comparative statics* amenable to testing with standard regression models. Rebecca Morton has referred to this theoretical approach as using a "partial data generating process."

A second approach to statistical testing is to adopt justifiable assumptions about the statistical processes generating the game's parameters among numerous cases in which the game is played; and then to use the game-theoretic model to derive the statistical properties of the resulting predicted outcomes. For example, *quantal response models* start from specific assumptions about the population distribution of true payoffs, noise, or mistaken choices, and derive the resulting distribution of choice outcomes in a game to specify appropriate statistical tests. Morton refers to such a model as a "complete data generating process." In general, the statistical properties that result from analyzing games in this fashion may or may not lend themselves nicely to analysis using standard techniques, such as ordinary least squares or other common regression-based models. In that case, the techniques of maximum likelihood, Bayesian statistics, and their associated computational techniques are often useful, since they place relatively few constraints on the distributional forms upon which statistical analyses are based.

Logistics

The course will meet in Seigle Hall, room 109, from 9:00-11:00 AM (Morning Session) and 2:00-4:00 PM (Afternoon Session) all four days. We will pose short homework assignments for the extended lunch hour, and longer assignments for the evenings. In the first three days of the course there will be an additional short lab session in the evening (times TBA) that will cover the R Language for Statistical Computing and implementation of much of the material discussed in the main sessions. R will be used extensively in this course. These lab sessions will meet in the Social Science Computing Facility in the basement of Seigle Hall.

Professors Calvert and Martin will be available during the Foundations seminar week for out-of-class discussion, and may be available at various times during the rest of the EITM program. Their office locations and email addresses are:

- Randy Calvert: `calvert@wustl.edu`, Seigle 328
- Andrew Martin: `admartin@wustl.edu`, Seigle 212A

Background Readings

To make the best use of our brief time in the Foundations course, we will assume all participants are familiar with certain basics of statistics and game theory. The background readings listed below are designed as a brush-up for any who may need it, and should definitely be examined prior to participants' arrival in St. Louis.

Moreover, students should read in advance the material covered in each unit of the course schedule. Time during the course can thus be spent re-reading important materials, and implementing some of the techniques developed in the class. "Further readings," however, can be postponed indefinitely.

For both background and covered readings, we have provided guidance to what is important. We will draw from these readings throughout the sessions. In the course schedule below, we note the particularly salient readings for the session.

Background Readings: Mathematical Statistics

DeGroot, Morris H., and Mark J. Schervish. 2002. *Probability and Statistics, Third Edition.* Reading, MA: Addison-Wesley.

This is an excellent introduction to mathematical statistics. It covers the basics of probability and distribution theory, statistical inference (from both frequentist and Bayesian perspectives), and simulation. Students should focus on Chapters 1-4 (probability and distribution theory), Chapter 6 (on Bayesian inference and maximum likelihood), Chapters 7 and 8 (on confidence intervals and hypothesis testing), and Chapter 11 (on simulation). Except for Section 11.4, the sections noted as optional can be skimmed.

Background Readings: Noncooperative Game Theory

Osborne, Martin J. 2003. *An Introduction to Game Theory.* Oxford, UK: Oxford University Press.

or alternatively

Gibbons, Robert. 1992. *Game Theory for Applied Economists.* Princeton: Princeton University Press.

Morrow, James D. 1994. *Game Theory for Political Scientists.* Princeton: Princeton University Press.

The course assumes that students are already acquainted with the basic mechanics of game theory. The Osborne text is an excellent introduction, not too advanced but reaching lots of important ideas and with tons of extremely useful illustrations and exercises across several disciplines.

Morrow's book is well known among political scientists, and gives solid coverage to the needed topics, along with political science applications. The Gibbons text is quite efficient, readable although fast-moving, while definitely introductory; the main downside is that its illustrations are confined entirely to economics topics. Morrow and Osborne both offer useful mathematical appendices, but as usual, this serves better as a refresher than as a math course.

The following list specifies a minimal set of references from Osborne, as well as the relevant sections of Morrow, for each topic. We will assume seminar participants to be substantially familiar with the first four topics, and to have some basic acquaintance with the last two.

- preference, utility, expected utility
  - Osborne 1.2 gives scant coverage of these basics.
  - Morrow Chapter 2, esp. pp. 16-34, a very nice chapter.
- specifying a game
  - Osborne 2.1-2.5

  - Morrow Chapter 3, esp. pp. 51-67
- dominance, best replies, and equilibrium; mixed strategies
  - Osborne 2.6-2.10, 4.1-4.5, 4.11; in general, the section on Nash equilibrium (Chapters 2-4) is perhaps the strongest feature of this text.
  - Morrow Chapter 4, esp. pp. 77-98
- candidate competition
  - Osborne 3.3
  - Morrow Chapter 4, esp. pp. 104-111
- extensive games, backward induction, subgame perfection
  - Osborne 5
  - Morrow Chapter 5, esp. pp. 121-132
- Imperfect and incomplete information
  - Osborne 10.1-10.5; a nice, clear treatment of the topic. Also, 9.1-9.3 gives an innovative, but somewhat more demanding, treatment of simultaneous-move Bayesian games.
  - Morrow Chapter 6

Background Readings: Spatial Voting Models

Keith Krehbiel. 1988. "Spatial Models of Legislative Choice." *Legislative Studies Quarterly* vol. 13, issue 3 (Aug., 1988), pp. 259-319.

> We will also assume that seminar participants are familiar with the basics of the one-dimensional spatial voting model, including the median voter theorem. If you would like a brush-up, the above survey article is as serviceable as anything; it sets out the basic model and its variations (this is needed background), and thoroughly explores the application of such models to legislative politics up to its publication date (read at your option). The Osborne game theory text, above, also offers game-theoretic analyses of equilibrium in spatial models of voting and candidate competition; see sections 3.3 and 8.6.

R Programming Language

Venables, W. N., D. M. Smith, and the R Development Core Team. 2005. *An Introduction to R.* http://www.r-project.org.

> We will be using the R Langauge for Statistical Computing throughout the Institute. R is an extremely flexible environment which can be used flexibly for many types of analyses, including simulation, model fitting, and graphing. R is also open-source free software, and is available for Windows, MacOS, and various versions of unix, and can be downloaded from the web. R is not a menu-driven canned statistical package, but rather an environment that is easily used to perform all kinds of analysis. Of course many commonly used tools have been implemented in R, but the real utility of the

language is its ability to easily implement non-standard models and simulations by hand. All students should work through this document to familiarize themselves with R before arriving in St. Louis. Three lab sessions will be devoted to further training in the software.

Course Schedule

Monday Morning

- EITM defined
- Random Variables and Distributions
- Joint and Conditional Distributions
- Bayes Theorem
- Change of Variables
- Probability Models
- Problem Set for discussion on Tuesday

Readings: DeGroot and Schervish (2000), Chapter 3, all; Chapter 5, Sections 1, 2, 4, 6, 9, and 12. Skim Chapters 1, 2, and 4.

Monday Afternoon

Equilibrium in strategic games

- Games in strategic form; best response and Nash equilibrium
- Mixed strategies
- Parameterized games; equilibrium and outcome correspondences
- Sources of statistical variation in games
- The "data-generating process" and empirical analysis
- Running examples:
  - the prisoner's dilemma
  - the battle of the sexes
  - candidate competition
  - Problem for discussion on Tuesday: "a synergistic relationship"

Readings: Review background reading selections from Osborne (2003), Chapters 2-4; or Morrow (1994), Chapters 3, 4.

For further reading on data generating processes: Rebecca Morton. 1999. *Methods and Models: A Guide to the Empirical Analysis of Formal Models in Political Science.* New York: Cambridge. Chapter 4.

Monday Evening: R Lab Session I

- Basic data structures and interface
- Loading data
- Graphics and visualization
- Linear regression and other common models

Tuesday Morning

- The logic of maximum likelihood
- Statistical inference and maximum likelihood
- Maximum likelihood for a battle of the sexes game
- Lunchtime Problem Set

Readings: DeGroot and Schervish (2000), Chapter 6, Sections 1, 5, and 6; Chapter 10, Sections 1-3. Skim remainder of Chapter 6, and Chapters 7 and 8.

Tuesday Afternoon

Backward induction in sequential games

- Review of exercise on the PD preference estimation problem
- Review of exercise on the "synergistic relationship" problem
- Sequential games, backward induction, and perfect Bayesian equilibrium
  - An "entry" or "challenge" game
  - A sequential "synergistic relationship"
  - A committee gatekeeping game
- generalizing backward induction
  - moves by Chance

  - simultaneous moves
  - information sets
  - asymmetric information

Readings: Review background reading selections on sequential games and backward induction: Osborne (2003), Chapters 5-7; or Morrow (1994), Chapter 5.

Tuesday Evening: R Lab Session II

- Maximum likelihood with the `optim()` command
- Simulation in R using standard distributions

Readings: DeGroot and Schervish (2000), Chapter 11. Raftery (1995).

Wednesday Morning

- Introduction to Bayesian statistics
- Model adequacy and non-nested model testing

Readings:

- DeGroot and Schervish (2000), remainder of Chapter 6
- Andrew D. Martin. N.D. "Bayesian Analysis." Prepared for *The Oxford Handbook of Political Methodology.*
- Edward Greenberg. 2008. *Introduction to Bayesian Econometrics.* Cambridge University Press, Chapters 2-4.
- Raftery, Adrian. 1995. "Bayesian model selection in social research (with Discussion)." In *Sociological Methodology 1995*, (Peter V. Marsden, ed.), Cambridge, MA: Blackwells, pp. 111-196.
- Robert E. Kass and Adrian E. Raftery. 1995. "Bayes factors." *Journal of the American Statistical Association.* 90: 773-795.

Wednesday Afternoon

The entry game as a complete data generating process

- Sources of uncertainty and statistical variation
  - Errors in players' actions

  - Private information about payoffs
  - Errors in measurement by the analyst
- Statistical models

Readings:

- Review background reading selections on sequential games, backward induction, and incomplete information: Osborne (2003), Chapter 10; or Morrow (1994), Chapter 6.
- Curtis S. Signorino. 2003. "Structure and Uncertainty in Discrete Choice Models." *Political Analysis* 11(4): pp. 316-344.

Wednesday Evening: R Lab Session III

- Bayesian statistics in R

Thursday Morning

- More Bayesian Statistics
- Monte Carlo simulation
- The Gibbs sampler
- Metropolis-Hastings algorithm

Readings:

- Greenberg (2008), Chapters 4, 5, 7, and 8.
- Casella and George. 1992. "Explaining the Gibbs Sampler." *The American Statistician.* 46: 167-174.
- Chib and Greenberg. 1995. "Understanding the Metropolis-Hastings Algorithm." *The American Statistician.* 49: 327-335.

Thursday Afternoon

Comparative statics in a partial data generating process

- What is comparative statics?
- Three approaches to comparative statics analysis
  - Explicit solutions

  - General functional forms and implicit differentiation
  - Supermodularity and monotone comparative statics

Readings

- **on comparative statics:** Charles M. Cameron and Rebecca Morton. 2002. "Formal Theory Meets Data" in *Political Science: The State of the Discipline* (third edition). New York: Norton.
- **on implicit differentiation:** Carl P. Simon and Lawrence Blume. 1994. *Mathematics for Economists*. New York: Norton. Chapter 15, "Implicit Functions and Their Derivatives."
- **on monotone comparative statics:** Scott Ashworth and Ethan Bueno de Mesquita. 2005. "Monotone Comparative Statics for Models of Politics." *American Journal of Political Science* 50(1):214-231.