350C/150C: Advanced topics in statistical models

Jonathan Wand   Gustavo Robles   Dorothy Kronick

Summary

Topics this quarter:

- mathematical and statistical models of choice;
- properties of ML (and other) estimators;
- simulation and resampling methods;
- estimation of unknown functions;
- shape constrained inference.

Contact information

email: `wand(at)stanford.edu`
Office: Encina Hall West, Rm 307
Office Hours: TBA and by appointment

Email: `grobles(at)stanford.edu`
Office: Encina Hall West, Room 311
Office hours: Th 2–4 and by appointment

Email: `dkronick(at)stanford.edu`
Office: Encina Hall West, Room 431
Office hours: F 1–3 and by appointment

Times and Locations

Lectures: T/Th 10-11:50 in Room GSB Patterson P101
Section I: F 9:30-11:20 in Room ZZZ
Computation Section: TBA

Prerequisites

Probability and statistics (Polisci 350A or equivalent)
Linear models (Polisci 350B or equivalent)
Experience with R is assumed.

Requirements/Evaluation

1. Weekly problem sets (30)

   Problem sets are due to be posted to website (or by email) by time/date. This will generally be 9pm on a given day.

2. Final exam (20)

3. Co-authored replication paper (30)

   The paper is meant as a way for you to apply the course material to a replication and possibly critique of existing published work. You are not restricted to what is covered in this course, but some connection to issues raised should be part of what you do. The goal is for you to produce a publishable paper.

   For this project you need to work in pairs. The ideal is that you share all parts of project. Use your judgement, but here are some guidelines for sharing coding and data:

   Deliverables:

   - Monte Carlo analysis (and de facto unit test): *code and logs*
     In short, this will be a simulation based analysis of what you will run on actual data. You will demonstrate that you can a) simulate the "Data generating process", b) apply your code to these simulations, and c) summarize the findings.
   - Replication data, code, and logs
     Data should have timestamps preserved where possible You need to show data processing to extent possible.
   - Paper

   For more suggestions and logic of tackling a replication paper, see

   - `http://projects.iq.harvard.edu/gov2001/book/replication-paper`
   - `http://gking.harvard.edu/papers/`

4. Review (and replication) of Replication paper (10)

   You will be assigned to replicate and review a fellow student's paper, and give comments. The goal is to take the role of a journal reviewer and give *constructive* comments.

   A good review will identify key contributions, and discussion of merits.

   A good review will also identify points which could be improved, with recommendations of *how* it could be different. Identification of a flaw without recommendation are less valuable.

5. Participation (10)

   Active participation in

   - on-line discussion
   - in class
   - in section

   are all important components of your overall education. If you don't understand, ask. If you have an answer or a conjecture, share.

   - Sections begin in first week
   - You are expected to have worked through the assigned readings prior to lecture, and begun a discussion and posted questions on Piazza.

Requirements/Evaluation – continued

A core part of this course is the computation and application of statistical methods. As such, part of being successful in this course is based on aspects which are too often dismissed, to the detriment of science. You will be evaluated on:

- programming (clarity, quality)
- data and project management
- replicability

1. For all computation projects, you must provide well documented R code and logs in ascii format in a well organized manner that enables replicated by another.

   Work must be submitted in a structured manner, e.g., for a problem set:

   ```
     <yourname>/ps1/R/ps1_q1.R
     <yourname>/ps1/R/ps1_q1.Rout
     <yourname>/ps1/R/ps1_q2a.R
     <yourname>/ps1/R/ps1_q2a.Rout
     <yourname>/ps1/R/ps1_q2b.R
     <yourname>/ps1/R/ps1_q2b.Rout
   [...]
     <yourname>/ps1/tex/ps1.tex
     <yourname>/ps1/tex/ps1.pdf
   ```

2. R: All code must be able to be run by anyone else without any editing. As such

   - all code must be OS independent
   - all code must be directory location independent
   - all log files must be run in BATCH mode of R

   You can develop code interactively, but for submission only results that are produced by BATCH will be considered. We will help you with this.

3. LaTeX: Except where explicitly noted, all written work must be submitted in a pdf file, produced by LaTeX.

4. There shall not be cut-and-paste or retyping of results from R log files into tex file, except where explicitly allowed.

5. Using Sweave or similar integrated replication/documentation machinery is also good (and in some cases the ideal rather than separate R and LaTeX code), but not required.

Recommended Texts

- There are no assigned texts, except Train (2003).
- For each topic, we offer multiple citations to different reference materials when they are available. These will vary in rigor and abstraction. Our suggestion is that you consult multiple texts to see the differences, and to find the text that works best for you.
- The following list is not exhaustive. At this point, you should begin reading broadly, referencing texts from previous courses you have taken and finding new books on your own. Think of this as an initial guide.

General econometrics

(pick one or two of the following):

- (CT) Cameron, A Colin and Pravin K Trivedi. 2005. *Microeconometrics: methods and applications*. Cambridge university press

  * A friendly/chatty introduction to topics, few proofs, lots of topics

- (W) Wooldridge, Jeffrey M. 2010. *Econometric Analysis Cross Section and Panel Data*. MIT press, 2nd edition edition. MIT Press

  * Still friendly/chatty, a useful alternative

- (DM) Davidson, Russell and James G MacKinnon. 2004. *Econometric theory and methods*. Oxford University Press New York

  * More abstract and general treatment

  Note: while the 2004 book covers more topics, for a reference book on general topics, you may find their earlier text superior:

  Davidson, Russell and James MacKinnon. 1993. *Estimation and Inference in Econometrics*. NY: Oxford

- Hayashi, Fumio. 2000. Econometrics. 2000. *Princeton University Press*

  * More abstract and general treatment

- Amemyia, Takshi. 1985. *Advanced Econometrics*. Cambridge, MA: Harvard University Press

  * Yet more abstract and general treatment

A good summary of the methods we will cover in this class can be found in Cameron and Trivedi. Woolridge and Davidson and MacKinnon provide more intuition for some topics. Finally, Amemiya is a good reference for technical and advanced topics. Additionally, there will be a couple of journal articles to deepen the understanding of the material.

Topics

- **(T) Train, Kenneth. 2003. *Discrete Choice Methods with Simulation*. Cambridge University Press

  * This is highly recommended.

- (Y) Yatchew, Adonis. 2003. *Semiparametric Regression for the Applied Econometrician*. Cambridge: Cambridge
- (PU) Pagan, Adrian and Aman Ullah. 1999. *Nonparametric Econometrics*. Cambridge: Cambridge
- (M) Manski, Charles F. 2008. *Identification for prediction and decision*. Harvard University Press

R

Textbooks on statistical analysis using R (most are bad, these are okay):

- Fox, John. 2002. *An R and Splus Companion to Applied Regression*. Thousand Oaks: Sage
- Fox, J., and Weisberg, S. 2011. *An R companion to applied regression.* Second edition. Los Angeles: SAGE.

Prob-stat

You should have this from ealier courses, but some standard texts are

- Casella, George and Roger L. Berger. 2002. *Statistical Inference*. Belmont, CA: Duxbury Press
- DeGroot, M.H. and M.J. Schervish. 2002. *Probability and Statistics*. Addison Wesley, Boston, MA, 3rd edition
- Resnick, Sidney I. 1992. *Adventures in stochastic processes*. Birkhauser

Students with Documented Disabilities

Students who may need an academic accommodation based on the impact of a disability must initiate the request with the Student Disability Resource Center (SDRC) located within the Office of Accessible Education (OAE). SDRC staff will evaluate the request with required documentation, recommend reasonable accommodations, and prepare an Accommodation Letter for faculty dated in the current quarter in which the request is being made. Students should contact the SDRC as soon as possible since timely notice is needed to coordinate accommodations. The OAE is located at 563 Salvatierra Walk (phone: 723-1066, 723-1067 TTY).

Course outline and readings

Introduction

This course is split into parts:

I Choice models and ML

II Estimating unknown functions

III Inference among models

IV Additional topics

Part I: Stochastic models, estimation, and inference

I.1. Models of choice

We start by examining how to translate a formal theory of choice to a statistical model. We will review the foundational work from mathematical psychology on the properties of choice behavior and qualitative judgments. We will begin with Luce's axiomatic approach and Thurstone's measurement approach. Though in appearance quite different, these models share a number of close connections, particularly for pairs of choices. Yellot derives the circumstances under which they are equivalent and the circumstances under which they are unique.

The statistical models we will derive are commonly referred to as the logit and the probit models. These will provide the working example while examining the theory of Maximum Likelihood.

Modules

1. Motivation: published research with choice models

   * Bartels, Larry M. 1996. Uninformed Votes: Information Effects in Presidential Elections. *American Journal of Political Science*, 40(1):194–230.
   URL http://www.jstor.org/stable/2111700

   * Londregan, John, Henry Bienen, and Nicolas van de Walle. 1995. Ethnicity and Leadership Succession in Africa. *International Studies Quarterly*, 39:1–25

2. Review: distributions (bernoulli, binomial, normal, extreme value, ...), and functions of distributions (differences of normals, evt)

   * Casella and Berger, Chapters 3, 4.1–4.3

   * Train, 3.10

3. Deriving choice models from analytical theories (theory of comparative choice; axioms of choice)

   * Train, Chapters 2.1–2.4, 3.1

   * Thurstone, L. L. 1927. A Law of Comparative Judgement. *Pychological Review*, 34:272–86

   * Chapter 1 of Luce, R. Duncan. 1959. *Individual choice behavior: a theoretical analysis*. NY: Wiley

4. Identification

   * Train, Chapter 2.5, 3.2

Section

- Coding standards, project management, replicability and R BATCH

  * Fox Chapter 3

- Review of probability theory (distributions, convergence concepts)

  * Wooldridge, Chapters 3.1–3.3

  * Casella and Berger, Chapter 5.5

  * Hayashi, Chapter 2.1

Supplementary Readings

- Luce, R. Duncan. 1994. Thurstone and Sensory Scaling: Then and Now. *Psychological Review*, 101(2):271–277
- McFadden, Daniel. 1974. Conditional logit analysis of qualitative choice behavior. In P Zarembka, editor, *Frontiers of Econometrics*, New York: Acadmic Press. pages 105–42
- McFadden, Danel. 1981. Structural Discrete Probability Models Derived from Theories of Choice. In Charles F. Manski and Daniel L. McFadden, editors, *Structural Analysis of Discrete Data and Econometric Applications*, Cambidge, MA: MIT Press, chapter 5. pages 198–272
- Yellot, John I. Jr. 1977. The Relationship between Luce's Axiom, Thurstone's Theory of Comparative Judgment, and the Double Exponential Distribution. *Journal of Mathematical Psychology*, 15:109–144
- Fearon, James D. and David D. Laitin. 2003. Ethnicity, Insurgency, and Civil War. *American Political Science Review*, 97(1):75–90
- Bartels, Larry M. 1986. Issue Voting Under Uncertainty: An Empirical Test. *American Journal of Political Science*, 30(4):709–728.
  URL http://www.jstor.org/stable/2111269
- Amemiya, Chapter 3

In 350a, ML theory was introduced. In this lecture we will derive ML estimators for dichotomous choice models. Our focus in this lecture will be on the mathematical properties of the choice models.

We will parameterize a model for dichotomous (i.e., binary) choices as a function of coefficients and individual specific variables.

We will consider methods for estimating the coefficients in the Bernoulli choice model. Emphasis will be placed on derivations and understanding features of the likelihood functions for the models we have seen. Introductory discussion of optimization methods will also be covered. We will return to the topic of maximum likelihood estimation and consider it more rigorously in later weeks.

Modules

1. Review: Definitions and basic theorems related to likelihood and ML

   * Casella and Berger 6.3, 7.2.2

2. Review: estimation of univariate bernoulli by ML

3. Likelihood for dichotomous choice, and generalization

   index/link/aggregator functions

   * Train, 3.7

   * Cameron and Trivedi, 14.1–14.5

   * Wooldridge, 17.1

4. Application

   * Londregan, John, Henry Bienen, and Nicolas van de Walle. 1995. Ethnicity and Leadership Succession in Africa. *International Studies Quarterly*, 39:1–25

Section

- Interpretation of non-linear models, marginal effects.

  * Cameron and Trivedi, Section 14.3.2

- Optimization

  * Gill, P., W. Murray, and M. Wright. 1981. *Practical Optimization.* Harcourt Brace and Company, London, Chapters 4.1, 4.3–4.6.

- Implementation in R. Tables, graphics, and optimization methods.

Supplementary Readings

- Long, 3.1–3.4, 3.7–3.8
- McFadden, Daniel. 1973. Conditional logit analysis of qualitative choice behavior. In P Zarembka, editor, *Frontiers of Econometrics*, New York: Acadmic Press. pages 105–42
- Davidson and McKinnon, 15.1–15.3, 15.5
- Train, 8
- Long, 3.5–3.6
- (opt) Pratt, John W. 1981. Concavity of the Log Likelihood. *Journal of the American Statistical Association*, 76(373):103–6
- Luenberger, D. 1973. *Introduction to Linear and Nonlinear Programming.* Addison Wesley, Massachusetts, Chapter 9, 189-213.

I.3. Generalized choice and other models

We will derive generalizations for multiple choice categories, including nested and ordered categories.

Modules

1. Multiple choices: categorical

   * Train 3.3

   * Cameron and Trivedi, chapter 15.

   * Woolridge, chapter 15.9-15.10.

   * Amemiya, chapter 9.

2. Multiple choices: nested models

   * Sanders, Mitchell S. 1998. Unified models of turnout and vote choice for two-candidate and three candidate elections. In Walter R. Mebane Jr., editor, *Political Analysis*, Ann Arbor: University of Michigan Press, volume 7. pages 89–116

3. Truncation and censoring, Tobit models.

Section

- Ordered models.

  * Krehbiel, Keith and Douglas Rivers. 1988. The Analysis of Committee Power: An Application to Senate Voting on the Minimum Wage. *American Journal of Political Science*, 32(4):1151–1174. URL http://www.jstor.org/stable/2111204

- Models of Count Data.

  * Cameron and Trivedi, chapters 20.1 - 20.4

- Duration models, wait times.

  * Cameron and Trivedi, chapter 17.

Supplementary Readings

- Amemiya, chapter 10.
- Train Chap 3 and 4
- Long Chap 5 (and 6)
- (GEV): Mebane, Jr., Walter R. 2000. Coordination, Moderation, and Institutional Balancing in American Presidential and House Elections. *The American Political Science Review*, 94(1):37–57. URL http://www.jstor.org/stable/2586379
- Lancaster, T. 1979. *Econometric Methods for the Analysis of Duration Data.* Econometrica, Vol. 47.

I.3b. Generalized choice II

Additional models and issues in choice models

Modules

1. IIA, spatial models of voting

   * Alvarez, R. Michael and Jonathan Nagler. 1998. When Politics and Models Collide: Estimating Models of Multiparty Elections. *American Journal of Political Science*, 42(1):55–96.
   URL http://www.jstor.org/stable/2991747

   * Otte, Clemens. 2013. Safe and Interpretable Machine Learning: A Methodological Review. In Christian Moewes and Andreas Nürnberger, editors, *Computational Intelligence in Intelligent Data Analysis*, Springer, volume 445 of *Studies in Computational Intelligence*. pages 111–122.
   URL http://dx.doi.org/10.1007/978-3-642-32378-2_8

2. Endogeneity and confoundedness

   * Freedman, David A. and Jasjeet S. Sekhon. 2010. Endogeneity in Probit Response Models. *Political Analysis*, 18(2):138–150.
   URL http://pan.oxfordjournals.org/content/18/2/138.abstractN2-Welookatconventionalmethodsforremovi
   includingthelinearmodelandtheprobitmodel.ItisknownthattheusualHeckmantwo-stepprocedureshouldnotbeus
   fromatheoreticalperspective,itisunsatisfactory,andlikelihoodmethodsaresuperior.However,
   seriousnumericalproblemsoccurwhenstandardsoftwarepackagestrytomaximizethebiprobitlikelihoodfunctio
   evenifthenumberofcovariatesissmall.Wedrawconclusionsforstatisticalpractice.Finally,weprovethecondit
   Theconditionsforidentificationaredelicate;webelievetheseresultsarenew.

   * Imbens, Guido W. and Joshua D. Angrist. 1994. Identification and Estimation of Local Average Treatment Effects. *Econometrica*, 62(2):467–475.
   URL http://www.jstor.org/stable/2951620

   * Heckman, James. 1997. Instrumental Variables: A Study of Implicit Behavioral Assumptions Used in Making Program Evaluations. *The Journal of Human Resources*, 32(3):441–462.
   URL http://www.jstor.org/stable/146178

Supplementary Readings

- Schofield, Normal, Andrew D. Martin, Kevin M. Quinn, and Andrew B. Whitford. 1998. Multiparty Electoral Competition in the Netherlands and Germany: A Model Based on Multinomial Probit. *Public Choice*, 97(3):257–293.
  URL http://www.jstor.org/stable/30024432
- Goldberger, A.S. 1991. *A course in econometrics*. Harvard Univ Pr, pp. 337–346
- Angrist, Joshua D., Guido W. Imbens, and Donald B. Rubin. 1996. Identification of Causal Effects Using Instrumental Variables. *Journal of the American Statistical Association*, 91(434):444–455.
  URL http://www.jstor.org/stable/2291629
- AP 4

I.4. Properties of ML estimators

Now that we understand some models, the issue is how to estimate the probabilities conditional on some observed variables, compare competing models, and draw inference. To this end we will spend the week on the theory of maximum likelihood estimation.

Modules

1. Consistency

   * Woolridge, chapter 13.4.

2. Identification

3. Fisher information and hessian

   * Woolridge, chapter 13.3.

   * Cameron and Trivedi, chapter 5.6.

4. Information equality theorems

Section

- Asymptotic normality of MLE

  * Woolridge, chapter 13.5.

- Estimation of Asymptotic Variance

  * Woolridge, chapter 13.5 - 13.6.

I.5. ML & inference

We will cover estimates of the covariance matrix and inference. Classical test statistics will be reviewed.

Modules

1. Hypothesis testing (Wald test, likelihood ratio, Lagrange multiplier)
2. Criterion (AIC, BIC)
3. Overfitting
4. Bootstrap

   * Wooldridge, chapter 12.8.2.

   * Cameron and Trivedi, chapters 11.1 - 11.4.

   * Hall, P. 1994. *Handbook of econometrics.* Vol 4. Chapter 39.

Section

- Review of hypothesis testing
- LM test and Three classical tests of equality restrictions

Supplementary Readings

- Train, Chap 8
- Long, Chap 4
- Davidson and McKinnon Chapter 8, 13
- Vuong, Q. 1989. Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses. *Econometrica*, 57:307–33
- Engle, R. 1984. *Wald, Likelihood Ratio and Lagrange Multiplier Tests in Econometrics*, in Griliches and Intrilligator (eds), Handbook of Econometrics, Vol III, Elsevier, North Holland.

Part II: Estimating unknown functions

II.1. Sequence of means, ordered tables

Modules

1. Preliminary: multiple comparisons

   * Yosef Hochberg , Ajit C. Tamhane. 2009. Multiple Comparison Procedures. Wiley

2. Tests of sequence of means

   * Hollander, Myles and Douglas A Wolfe. 1999. *Nonparametric statistical methods*. John Wiley New York, 2nd edition

   * Bartholomew, DJ. 1959. A test of homogeneity for ordered alternatives. *Biometrika*, 46(1-2):36

Section

- stochastic dominance
- finite sample properties of estimators (MSE, L1, PRE)
- critical of complex tests
- designing a Monte Carlo Study

  * Wooldridge, chapter 12.8.1.

  * Cameron and Trivedi, chapter 13.5.

Supplementary Readings

- Barlow, R. E., D. J. Bartholomew, J. M Bremer, and H.D. Brunk. 1972. *Statistical Inference Under Order Restrictions*. NY: Wiley
- Robertson, Tim, F. T. Wright, and R. L. Dykstra. 1988. *Order Restricted Statistical Inference*. NY: Wiley

II.2. Flexible and Nonparametric functions
Fitting flexible functions, Parametric, semi-parametric and nonparametric methods. Generalized additive models.

Modules

1. Polynomials
2. Kernel estimation

   * Cameron and Trivedi, chapters 9.1 - 9.6

   * Hastie, Trevor J., Robert J. Tibshirani, and Jerome Friedman. 2001. *The Elementsof Statistical Learning: Data Mining, Inference, and Prediction*. NY: Springer

3. Local Polynomial, Loess
4. Basis fucntions and Splines
5. Lasso

   Hastie, Trevor J., Robert J. Tibshirani, and Jerome Friedman. 2001. *The Elementsof Statistical Learning: Data Mining, Inference, and Prediction*. NY: Springer

Section

- Implementation in R.

Supplementary Readings

- Efromovich, S. 1999. *Nonparametric Curve Estimation.* Springer Verlag, New York.
- Fan, J. and I. Gijbels. 1996. Local Polynomial Modelling and Its Applications Chapman and Hall, London.
- Pagan, Adrian and Aman Ullah. 1999. *Nonparametric Econometrics*. Cambridge: Cambridge. Chapter 3.
- Hardle, W. 1990. *Applied Nonparametric Regression.* Cambridge University Press, Cambridge.
- Hastie, Trevor J., Robert J. Tibshirani, and Jerome Friedman. 2001. *The Elementsof Statistical Learning: Data Mining, Inference, and Prediction*. NY: Springer Chapter 5-6.
- Silverman, B. 1984. *Density Estimation for Statistics and Data Analysis.* Chapman and Hall.
- Wand, M., and M. Jones. 1995. *Kernel Smoothing.* Chapman and Hall.

II.3. Shapes constrained inference

Modules

1. algorithms for shape constrain estimation
2. Inference with shape constraints

   * Wolak, F.A. 1989. Testing inequality constraints in linear econometric models. *Journal of econometrics*, 41(2):205–235

   * Wand, Jonathan. 2010. More than a Science of Averages: Testing Theories Based on the Shapes of Relationships

Section

- Degrees of freedom in non-paramtric estimators
- backfitting

Supplementary Readings

- Silvapulle, Mervyn J. and Pranap P. Sen. 2005. *Constrained Statistical Inference*. New Jersey: Wiley
- Geyer, Charles J. 1991. Constrained Maximum Likelihood Exemplified by Isotonic Convex Logistic Regression. *Journal of the American Statistical Association*, 86(415):717–24
- Bachetti, Peter. 1989. Additive Isotonic Models. *Journal of the American Statistical Association*, 84(405):289–94
- Meyer, Mary and Michael Woodroofe. 2000. On the Degrees of Freedom in Shape-restricted Regression. *Annals of Statistics*, 28(4):1083–1104

Part III: Inference among models

III.I. Picking a model

Modules

1. Model selection, model validation and model confidence sets

   * Hansen, P.R., A. Lunde, and J.M. Nason. 2011. The model confidence set. *Econometrica*, 79(2):453–497

   * White, Halbert. 2000. A Reality Check for Data Snooping. *Econometrica*, 68(5):1097–1126. URL http://www.jstor.org/stable/2999444

2. Cross-validation

3. Mixture models

   * Geoffrey McLachlan David Peel. 2000. Finite Mixture Models. Wiley

Section

- Bayes; EM Algorithms
- Implementation in R