PSCI 505
Likelihood + Topics

Fall 2023
Wed/Thurs 2pm - 3:15, H329

Prof Curt Signorino
303 Harkness Hall
Office Hours: Tues 2-4pm
curt.signorino@rochester.edu

TA: Thomas Chamberlain
305 Harkness Hall
Office hours: TBD
tchambe6@ur.rochester.edu

COURSE DESCRIPTION: This course builds upon the analytical and applied foundations of PSC 404 and 405, taking the classical linear model as its point of departure. In this course, students will learn methods to analyze a variety of data – e.g., binary data, event counts, durations, censored data, truncated data, or data from self-selection. Much of the course will focus on two ``likelihood"-based methods of estimation and inference: maximum likelihood estimation and Bayesian estimation. Students will learn the statistical theory underpinning these statistical techniques, as well as how to apply them to data in political science and in international relations. Although students will learn ``canned" commands for implementing many of these methods, a major goal of the course is to teach students how to estimate models for which there is no canned command in R. Students will be required to program their own statistical routines in R, jags, and stan. Time permitting, we will also survey a number of techniques from the machine learning literature, including penalized (or shrinkage) estimators, random forests, and neural networks.

PREREQUISITES: PSC 404, 405, or the equivalent. Calculus. Matrix algebra.

COURSE REQUIREMENTS: Course grades will be based on a series of homeworks (50%), a course paper (45%), and participation (5%). The HWs will consist of a mix of analytical problems, programming, and data analysis. For HWs, students are encouraged to work in groups of any size, so long as that size is no greater than two. HWs must be submitted as a pdf via Blackboard by lecture the day it is due. Late HWs are penalized one point per day late. Late HWs submitted within the first 6hr period after the class in which they are due will receive a half-point penalty. Those submitted between 6-24 hours after the due date will receive a full point off. Each additional 24 hour period is another point off.

Coauthored final papers are not allowed. Papers submitted for another course this semester are acceptable (so long as the other instructor agrees as well). However, the research, statistical analysis, and writing must have been conducted *this semester*. Replication papers are strongly encouraged. We will discuss replication papers later in the semester.

COURSE CREDITS: This is a four-credit course, consisting of in-class lecture (3 credit hours), a section (1 credit hour), and out-of-class student time spent on reading, homeworks, and the final paper.

READINGS: Students are responsible for keeping up with the reading each week. I post my lecture notes and will provide links or copies of articles from time to time. In addition, students should read the appropriate chapters in the following:

Required

- **Yudi Pawitan. 2013. *In All Likelihood*.**
- **Jeff Gill. *Bayesian Methods: A Social and Behavioral Sciences Approach*.**
- **Braun & Murdoch. *A First Course in Statistical Programming with R*.***
- **Hastie, Tibshirani, Friedman. *The Elements of Statistical Learning* (*ESL*)***

Useful

- Gary King. 1998. *Unifying Political Methodology*.
- A. Colin Cameron & Pavin K. Trivedi. *Microeconometrics*.*
- William H. Greene. 1997. *Econometric Analysis*.*
- Patrick Burns. 2011. *The R Inferno*.*
- The star lab introduction to R. http://www.sas.rochester.edu/psc/thestarlab/resources.php

* PDF available on course blackboard page under Course Materials >> Texts.

COURSE OUTLINE:

1. **R Programming and Simulation Methods, pt 1**
   Conditional statements, loops, vectorization, Monte Carlo analysis

   - Braun & Murdoch. *A First Course in Statistical Programming with R*.
   - Burns, Patrick. 2011. *The R Inferno*.

   Section: R programming

LIKELIHOOD PART 1: MAXIMUM LIKELIHOOD ESTIMATION

2. **Theory and Implementation**
   (Log)Likelihood, information matrix, covariance matrix, LR test, maxLik

   - Pawitan, 2013. *In All Likelihood*. Chapters 1-3, 6.1.
   - King, Gary. 1998. *Unifying Political Methodology*. Chapters 1-4.
   - Zivot, E. 2009. "Maximum Likelihood Estimation." Notes.

   Section: maxLik(), calculating se's using covariance matrix, LR test "by hand", lrtest()

3. **Binary Data and Count Data Models**
   Logistic regression, latent variable models, negative binomial regression

   - Pawitan, 2013. *In All Likelihood*. Chapters 4.1-4.8. (Bernoulli, Binomial, Poisson)
   - Pawitan, 2013. *In All Likelihood*. Chapter 6.2-6.3. (Logistic & Poisson Regression)
   - King, Gary. 1998. *Unifying Political Methodology*. Chapters 5.6-5.10.

   Section: Scaling, computational issues, fitted values "by hand" and using predict()

4. **Confidence Intervals for E(y|X)**
Delta method, parametric bootstrap (CLARIFY), nonparametric bootstrap

- Pawitan, 2013. *In All Likelihood*. Chapter 5.
- King, Gary. 1991. "Calculating Standard Errors of Predicted Values based on Nonlinear Functional Forms." *The Political Methodologist* 4(2).
- Efron, Bradley and Gail Gong. 1983. "A Leisurely Look at the Bootstrap, the Jackknife, and Cross-Validation." *The American Statistician*. 37(1):36-48.

Section: Parametric and nonparametric bootstrap

5. **Censoring, Truncation, and Self-Selection**
Tobit model, Heckman selection model

- Sigelman, Lee and Langche Zeng. 1999. "Analyzing Censored and Sample-Selected Data with Tobit and Heckit Models." *Political Analysis* 8. Read pages 167-177.
- Amemiya, Takeshi. 1984. "Tobit Models: A Survey." *Journal of Econometrics* 24: 3-60.
- Heckman, James J. 1979. "Sample Selection Bias as a Specification Error." *Econometrica* 47: 153-162.

LIKELIHOOD PART 2: BAYESIAN INFERENCE

6. **The Bayesian Posterior and Prior**
- Gill, Jeff. *Bayesian Methods*. Ch 1-5.
- Efron, Bradley. 1986. "Why Isn't Everyone Bayesian." *The American Statistician*.

7. **Simulation Methods, pt 2: Markov Chain Monte Carlo Simulation**
- Gill, Jeff. *Bayesian Methods*. Ch 9 & 11.
- Elements of Statistical Learning, Ch 8.1-8.6.
- Casella, George and Edward I. George. 1992. "Explaining the Gibbs Sampler." *The American Statistician*.
- Jackman, Simon. 2000. "Estimation and Inference via Bayesian Simulation." *AJPS*.
- Cowles, Mary Kathryn and Bradley P. Carlin. 1996. "Markov Chain Monte Carlo Convergence Diagnostics." *JASA*.

Section: Jags, Stan

8. **Bayesian Hierarchical Models**
- Gill, Jeff. *Bayesian Methods*. Ch 10.
- Jackman, Simon. 2008. Ch 7.1-7.2. *Bayesian Analysis for the Social Sciences.*
- Jackman Simon. 2008. Ch 7.3-7.5. *Bayesian Analysis for the Social Sciences.*
- Park, David K., Andrew Gelman, and Joseph Bafumi. 2004. "Bayesian Multilevel Estimation with Poststratification." *Political Analysis*.
- Shor, Boris, Joseph Bafumi, Luke Keele, and David Park. 2007. "Bayesian Multilevel Modeling Approach to Time-Series Cross-Sectional Data." *Political Analysis*.

SURVIVAL ANALYSIS

9. Parametric Models

- Pawitan, 2013. *In All Likelihood*. Chapters 4.9, 11.5-11.6.
- Box-Steffensmeier, Janet and Bradford S. Jones. 2004. *Event History Modeling: A Guide for Social Scientists*. Chapters 2 – 8.

10. Cox Proportional Hazard Models

- Pawitan, 2013. *In All Likelihood*. Chapter 11.7.
- Box-Steffensmeier, Janet and Bradford S. Jones. 2004. *Event History Modeling: A Guide for Social Scientists*. Chapter .
- Box-Steffensmeier, Janet M. and Christopher J. W. Zorn. 2001. "Duration Models and Proportional Hazards in Political Science." *American Journal of Political Science* 45: 972-988.

Section: survreg()

11. Grouped Binary Duration Data

- Beck, Nathaniel, Jonathan N. Katz, and Richard Tucker. 1998. "Taking Time Seriously: Time-Series-Cross-Section Analysis with a Binary Dependent Variable." *American Journal of Political Science* 42: 1260-1288.
- Carter, David B. and Curtis S. Signorino. 2009. "Back to the Future: Modeling Time Dependence in Binary Data." *Political Analysis*. 18(3):271-292.

Section: splines

MACHINE LEARNING

12. Penalized Estimators: LASSO & Ridge Regression

- ESL, 3-4.

Section: brglm()

13. Random Forests

- ESL, 9, 10, & 15.

14. Neural Networks

- ESL, 11.

15. Text as Data

- Spirling & Rodriguez papers
- TBD

IMPORTANT DATES

Fall Break	*10/16-10/17*
Topic and Data OK'd	*10/26*
Thanksgiving Break	*11/22 - 11/24*
Rough Draft Due	*11/22*
Comments Returned	*11/29*
Class Presentations	*12/13*
Final Paper Due	*5pm, Wed 12/20*

ACADEMIC HONESTY

Students are expected to abide by the College's policy on academic honesty. Please review the policy at https://www.rochester.edu/college/honesty/graduates.html
In short, don't plagiarize; don't cheat. When in doubt, please ask the professor for guidance.

Please review the course academic honesty page on Blackboard.

For this course in particular:

- For the purposes of this course, an AI site or app is considered the same as another person.
- You may use AI sites and apps to learn about concepts in this course – much like you would use Wikipedia, other online sites, or discuss a topic with a classmate. However, you may not use AI sites or apps to answer HW problems or to write code for you that is then used in HW submissions or in your final paper.
- You are not allowed to consult material from previous years that other PSCI 505 students have kept or archived.
- All R/jags/stan code should be written by you. I will provide you with many examples throughout the semester. Sometimes it will make sense to use the same variable names and techniques. However, you should never simply copy and paste my R code for use in your homework or final paper.
- If you have never written a research paper, please consult with your Director of Graduate Studies for information on plagiarism and academic honesty.
- You should not post or share course material (e.g., lectures and homeworks) online without first obtaining the permission of the instructor.

Updated: 8/31/23