1 Course Description

This course introduces PhD students to the theory and practice of likelihood inference for statistical models, as applied to the social sciences. Models include those designed for binary, nominal, ordinal, count, and continuous random variables. We will discuss theory, estimation, interpretation, and presentation of results. Time and student interest permitting, we will also introduce models for spatially and temporally correlated data; survival and event history models; missing data imputation; and model validation and selection.

Motivation

Many, if not most, social science research problems do not easily conform to the standard linear regression model. The likelihood framework provides a powerful set of tools through which to study many social phenomenon while also subsuming the standard linear regression model. Likelihood-based models are now readily incorporated into most statistical packages, so estimation is often trivial. Nevertheless, interpretation and evaluation of likelihood-based models requires care and nuance. The principles of likelihood theory and maximum likelihood estimation are foundational for continued methodological study.

This course serves as an introduction to maximum likelihood methods with an applied social science focus. We will discuss the selection of appropriate statistical methods for a variety of typical research questions in the social sciences while providing the theoretical and computational foundations for maximum likelihood estimation. Our goal is to empower you to develop your own statistical models, tailored to fit your research problems.

2 Expectations and Student Responsibilities

I assume that you have the following:
1. recent exposure to basic linear regression and elementary probability theory.

2. a working knowledge of multivariate calculus, especially differentiation and optimization (finding maxima and minima and knowing when you have one or the other).

3. a working knowledge of basic linear (matrix) algebra.

4. some experience with statistical computing in \( \mathcal{R} \).

If any of the preceding does not describe you, please come and see me immediately.

I expect you to attend all lectures, to have done the relevant readings, and to complete all assignments in a timely fashion. I also expect you to be self-directed in choosing a topic for your final project, though I am happy to provide feedback on paper topics and suggestions for finding data.

**Course Deliverables**

- Quasi-weekly Homework
- Replication and extension of an article of interest that focuses on non-OLS techniques
  - Presentation of a poster on **December 10**, detailing your project and your findings
  - 10-15 page writeup of the “Methods” and “Results” sections of a paper on your project. No theory, no introduction, no conclusion, no summary. Due at **noon, December 17**.
- Comment and feedback for colleague on his/her project.
- Establish online archive for sharing your data and work with others.
- Midterm exam to check your comprehension of the maximum likelihood principles.

**Grading**

Your grade will be composed of the following components:

- 1% Office hours visit
- 1% data archive set up
- 30% homework assignments. Grading will be assigned based on three categories (unacceptable, acceptable, exceptional). Each is worth four points, you may drop or not turn in one assignment.
- 20% Midterm
- 15% partner comments
• 33% Poster & final write up

There will be no incompletes given in the this class. I will not accept late papers.

Class meetings

We meet weekly. Each week will have a text chapter and a selection of actual applied research articles. I expect that you will come to class meetings prepared, having already read the required texts and developed questions.

A note on reading articles for this class: Each week we will read a selection of research articles applying the tools we are studying that week. In reading these articles will ignore the introduction, literature review, theoretical model, and conclusion. We will focus exclusively on the data, measurement, and modeling sections, including the discussion of results and findings. Come to class prepared to discuss modeling and presentation of results. For this class we don’t care about theory.

I will begin each class meeting by reviewing lingering issues from prior meetings and problem sets. We will then turn toward your questions and issues with the text and readings. I aim to do relatively little lecturing in this course.

Note that the main text is currently being reviewed by an academic press. You are in a great position to help improve it. Please share any and all feedback you have!

Lab sessions

Weekly lab sessions will be conducted by the teaching assistant. They are designed to introduce you to the nuts-and-bolts of working with the data and models in R as well as providing a common space and time for you to collaborate on homework and projects. Lab sessions will introduce the basic computational tools needed to complete the homework assignments as well as answer lingering questions you may have from the standard course meeting. Lab sessions will be organized with the TA.

Homework assignments

There will be periodic (~9) homework assignments during the class designed to get you to work with real data and produce and interpret results. All homework MUST be typewritten and submitted electronically in PDF format via the Learn@UW course site. You should also include your (documented) R code as a text file appendix. You can work in groups on the homework but all writeups, results, and coding must be individual work.

Homework is due at noon on Tuesday each week it is assigned. This will give me a chance to read it in time to cover any items that are necessary before we move on to subsequent topics. You may drop one homework, without penalty.

poster & paper

The final course project has two components. The most familiar is a final paper. The paper is meant to be a draft of an empirical article to be submitted to a social science journal,
using some type of model that is based on maximum likelihood methods. This may be a replication or extension of a published piece, but cannot simply be a recapitulation of it. You have to add something new to the research. This draft should follow the style of an article in the *American Political Science Review*. Be advised that I want *only* the data analysis section of the paper that describes the (statistical) model, the data, the results, and your conclusions. I do now want the introduction, literature review, theory, or motivation.

You will be asked to make progress reports on this research over the course of the semester, so you should come to class each week prepared to describe your project and where you stand. Additionally, everyone in class will be assigned a partner. The partner will function as a co-author, and will provide written comments on one draft of the paper, due to both me and your partner on **Friday, November 21 at 5pm**. You should have your comments back to your partner by 5pm the following Monday.

I’ll be more precise about what I expect in terms of presentation of statistical results, but in general I hate tables (unless they have data in them) and creative graphics are always a welcome contribution. The paper should be no longer than 20 pages double spaced, and ideally it should be a bit shorter. Do not include your references in this total, but do include your tables and graphics. A good summary of how to construct the paper can be found here: [http://gking.harvard.edu/files/paperspub.pdf](http://gking.harvard.edu/files/paperspub.pdf).

The second portion of your final project is the production of a poster summarizing your research and conclusions. Posters are the most common form of communication at academic conferences in the natural sciences, becoming increasingly important in the social sciences. The last class will be dedicated to poster presentations. Political science faculty and other grad students will be invited to come and see the results of your work. Your poster should conform to the criteria for a PolMeth poster session, as given here: [http://polmeth.wustl.edu/conferences/methods2011/polmeth.princeton.edu/Poster_Sessions.html](http://polmeth.wustl.edu/conferences/methods2011/polmeth.princeton.edu/Poster_Sessions.html). The Department of Political Science has generously offered to help defray the cost of poster printing; more on this later.

### 3 Office Hours

I insist that you come talk to me in office hours at least once prior to week 3 of the semester. By this date you will be required to get my consent for your choice of final project. You will be asked to turn in a 1 page summary of your proposed final project, as well as set up an online data archive.

The vast majority of your learning will occur as you struggle with the problem sets and your project. For this reason I will be holding more than the usual number of official office hours for this course. You are strongly encouraged to come and talk to me about any questions or problems you might be having. If you feel lost or that you are falling behind, come see me sooner rather than later so that we can improve your comprehension and the content of labs and lectures. In general, it is a good idea to bring a laptop or a usb key with your code on it when you come to OH.

If you have concerns with lectures, the format of the course, or any other issue PLEASE
bring them to my attention.

4 Email & website policy

I will be using email regularly to communicate with the class. You should feel free to contact me via email. I will do my best to respond to email queries within 12 hours during the week and 24 hours on weekends. Please do NOT expect me to be at your beck and call on Sundays.

If you have comments, questions, or concerns about any aspect of the course that you are uncomfortable bringing to me, I have provided a link on the course website to send me anonymous email feedback.

This course will have a dedicated website. Course notes, assignments, and solutions will be posted regularly on the site. I will also post links to other useful sites (e.g., CRAN, TUG). If you have comments or suggestions on how to make this more useful, please let me know.

5 How to Succeed in this Course

This is a course in learning how to fit and interpret a variety of statistical models using maximum likelihood estimation. Each approach will have at least three examples presented, one in class, a different one in lab, and one as graded homework. Most all of these will be real, not toy, examples that I have taken from contemporary social science research.

You should not expect to understand everything the first time through. It is unlikely that you will be able to solve the assigned problems in their first pass. Thus, tenacity is crucial. You will get negative feedback. The point is to learn from it, not to stop at it. It is also unlikely you will understand all the concepts and details the first time through but I promise you that as you work these models repeatedly in your own work and as you revisit the notes and readings over time you will come to have a greater appreciation for the material covered in this course.

There is not a great deal of reading for this course, but I have assigned readings in two forms. First, there is text material that should help you understand the techniques. Second, there are real research examples of each technique, so that you will see how scholars use these approaches in practice. Carefully read all of the text readings before class each week. It is important to read the relevant methods and analysis sections of the examples before class, as I will be discussing them in some detail and will refer to them in my handouts and notes. It probably wouldn’t hurt to read them again after each class, though that is entirely up to you.

Often times, your course colleagues will have an insight or explanation that far exceeds the instructor’s, or even your own. So collaboration is one way to take advantage of these insights among your fellow travelers in this course. It is also a way to share your own. You should feel encouraged to collaborate in this class. However, when you get to it, sit down and do your own homework, as a synthetic testament to how much you learned.
6 Required materials

Books
The core texts for the class are:


Recommended reference and alternative texts

*Statistical theory*


*Computation in R*


Math


7 Computation

\( \mathcal{R} \) Software

The analysis we undertake in this course can be accomplished in a variety of computing platforms. Feel free to use whatever computational tools will best suit your own work.

That said, \( \mathcal{R} \) will be the primary statistical package used by the instructor. All worked examples in the lecture notes will have \( \mathcal{R} \) code associated with them. Expertise in other programs like Gauss, Matlab, or STATA will be useful, but all assignments and solutions will be given for implementation in \( \mathcal{R} \). If you are capable of programming in lower level languages like C, C++, or Java that’s great, but I am not a programmer and will not be able to offer much help there.

\( \mathcal{R} \) is a free, open source statistical computing environment. You can download and install the most recent version of from CRAN at [http://cran.r-project.org/](http://cran.r-project.org/). \( \mathcal{R} \) runs on most operating systems, including Windows, Linux, Unix and the Apple OS/X (and earlier) operating systems. While \( \mathcal{R} \) is at the cutting edge of statistical software, it takes time to master. Lab sessions for the course will emphasize the nuts-and-bolts of statistical computation in \( \mathcal{R} \).

\LaTeX

\LaTeX is a typesetting system that is very commonly used by technical writers (i.e., those making extensive use of equations, numbers, tables, and graphics). Most in political science have settled on the \LaTeX “dialect” of \TeX. \TeX is freely available via the \TeX Users Group at [http://www.tug.org/](http://www.tug.org/) or via the Comprehensive \TeX Archive Network (CTAN), primary repository for \TeX-related software: [http://www.tug.org/ctan.html](http://www.tug.org/ctan.html). There is a good, free implementation of \LaTeX for Windows via MikTex: [http://www.miktex.org/](http://www.miktex.org/). If you are working in the Mac world, you have a few choices, several of which come pre-installed with OSX.

The learning curve for \LaTeX is fairly steep but I strongly encourage you to use this course to learn and practice writing in \LaTeX. While I do not require that you adopt \TeX for your own purposes, there are numerous benefits to doing so, among them easier formatting of your doctoral dissertation and job market signaling. I will be available to help you with some \LaTeX related problems. While there are numerous books and online references for \TeX the reality is that most find it useful to start with a file template someone else created that already works replace their stuff with your stuff, tweak it a bit, and declare victory.
You are ultimately free to use whatever text processor you wish; if you are using WYSIWYG oriented word processors such as Microsoft Office, StarOffice, WordPerfect or other similar products for this task, you will need to learn about (and possibly purchase) MathType (http://www.dessci.com/en/products/mathtype/) in order to adequately display quantitative materials.

Text editors

Text editors are stand alone programs meant to aid in writing, retaining, and (sometimes) compiling computer code. Regardless of the statistical package you choose to use, you must keep a record of the code you ask the computer to execute. Working in \LaTeX also requires you to enter both your text and the appropriate \LaTeX commands into a text editor. You should therefore find a text editor that you are comfortable working in. While the options are myriad, WinEDT is my preferred text editor in a Windows environment, working very nicely with the MikTeX. If you are interested in a freeware option, you can get copies of the EMACS editor (www.xemacs.org), which links to R directly via the Emacs-Speaks-Statistics (ESS) package. Aquamacs is the commonly used free text editor in for Mac OSX. OSX also come pre-loaded with TeXShop, an implementation of \LaTeX and text editor. TextWrangler is also a popular option. Sublime is an excellent text editor for Mac OSX but does have a cost.

8 Course Outline

Week 1: Introduction to the class; assignments; projects; \texttt{R}; \LaTeX; Good computing practice.

Readings


“How Much Should We Trust Economics?” http://www.npr.org/blogs/money/2013/04/19/177999020/episode-357-how-much-should-we-trust-economics

Week 2: Looking at your data

Readings

Kastellac & Leoni ”Using graphs instead of tables in political science”


Shih et al. 2012. “Getting Ahead in the Communist Party: Explaining the Advancement of Central Committee Members in China.” APSR
**Week 3: Introduction to likelihood inference**

Readings

Ward & Ahlquist Ch. 1
Mebane & Sekhon. 2002. “Coordination and Policy Moderation at Midterm.” *APSR.*

**Week 4: Binary data**

Readings

Ward & Ahlquist ch. 2
Choose at least one of:
- Burden et al. 2014. “Election Laws, Mobilization, and Turnout” *AJPS.*

**Week 5: Challenges & Imputation.**

Replication project must be approved by today!

Readings

Ward & Ahlquist ch. 3.
Choose at least one of:

**Week 6: Interpretation, Cross-validation, Presentation of Results**

Readings

Ward & Ahlquist ch. 4
Choose at least two of the following:
- Goldstone et al. 2010. “A Global Model for Forecasting Political Instability.” *AJPS.*
- Lyall et al. 2013. “Explaining Support for Combatants during Wartime.” *APSR.*
- Metternich et al. 2013. “Antigovernment Networks in Civil Conflicts.” *AJPS.*
Week 7: Ordinal data; contrasts

Readings

Ward & Ahlquist ch 5
Treier & Jackman. 2008. ”Democracy as Latent Variable.” AJPS

Week 8: Nominal data

Readings

Ward & Ahlquist ch. 6
Choose at least one of:


Week 9: Counts

Readings

Ward & Ahlquist ch. 7
Choose at least one of:

- Malesky & Schuler. 2010. “Nodding or Needling.” APSR.

Week 10: MIDTERM EXAM

Week 11: Multiequation & mixture models.

Readings

Ward & Ahlquist ch. 8
Choose at least one of:

Week 12: Non-IID data I: temporal dependence, ”robust” SE.

Readings

Haber & Menaldo. 2011. “Do Natural Resources Fuel Authoritarianism.” *APSR*

Week 13: Non-IID data II: Event history models

Readings

Ward & Ahlquist ch. 9
Choose at least one of:

- Svolik. 2008. “Authoritarian Reversals and Democratic Consolidation.” *APSR*

Week 14: Non-IID data III: Spatial dependence.

Readings

Gleditch & Ward.
Additional reading TBD.

Week 15: Poster Session.

Caveats

The syllabus is intended to provide an overview over the course. You cannot claim any rights from it. In particular, scheduling and dates may change. While the syllabus should be a reliable guide for the course, official announcements are always those made in class by the instructor.

Americans with Disabilities Act

Students with disabilities needing academic accommodation should: (1) register with and provide documentation to the Student Disability Resource Center; (2) bring a letter to the instructor indicating the need for accommodation and what type. This should be done
during the first two weeks of class. For more information about services available to students with disabilities, contact McBurney Center 702 W. Johnson St.; 608-263-2741; mcburney@studentlife.wisc.edu; http://www.mcburney.wisc.edu/