POLS/CSSS 510: Maximum Likelihood Methods for the Social Sciences

Christopher Adolph · Associate Professor · Political Science and CSSS

University of Washington
Fall Quarter 2017

Class Meets	Office
TTh 4:30–5:50 PM	Gowen Hall 145
Savery Hall 264	cadolph@uw.edu
Section Meets	Teaching Assistant
F 4:00–5:20 PM	Daniel Yoo
Savery Hall 117	dhyoo@uw.edu

Overview and Class Goals. Most social science data do not fit the assumptions of the linear regression model taught in introductory statistics courses. For example, social science data often consist of discrete categorizations or counts of events, rather than continuous outcomes. Observations may be correlated across periods, as in time series, or clustered into correlated groups, violating the linear regression assumption of independence. In this course, students will learn how use the method of maximum likelihood to derive statistical models that suit the particular behavior of their social science data and how to clearly communicate the substantive import of their findings to a broad audience. In the process, students will gain familiarity with basic statistical programming in R, a free and increasingly popular language. Topics of special interest to students will be covered as time permits.

Prerequisites. No specific courses are required; however, students should have a solid grounding in linear regression, as provided, for example, by courses in Political Science (POLS 501 and POLS 503), Sociology (SOC 504 and SOC 506), or Statistics (STAT/CSSS 504). Familiarity with or a willingness to quickly learn basic probability and matrix algebra is also required.

Course Requirements. Course evaluation will be based on problem sets (approximately six over the term), a student poster presentation in class, and a research paper (due Tuesday, 13 December 2015 at 3 PM in my mailbox in Gowen Hall). Students are **required to collaborate in groups of two or three** on the poster presentation and paper, which should apply methods studied in the course (or with instructor approval, related methods of similar sophistication) to a student-chosen topic. Papers developing new methodological tools are also acceptable. Each paper-writing group must propose a paper topic as part of the first assigned homework, so students should plan to form research partnerships quickly. Additional paper requirements and guidelines can be found on the course website.

Office Hours. Chris Adolph: Thursdays, 3:00 – 4:15 PM and by appointment in Gowen 145. Daniel Yoo: Tuesdays, 3:00 – 4:20 PM and by appointment in Smith 220.

Course Website. Consult `http://faculty.washington.edu/cadolph/mle` for problem sets, notes, and announcements.

Course textbooks

Required (sharing encouraged)

Andrew Gelman and Jennifer Hill. 2007. *Data analysis using regression and multilevel/hierarchical models*. University of Cambridge Press. ISBN-10: 052168689X. ISBN-13: 978-0521686891.

Excellent general introduction to linear regression and hierarchical modeling; good for code and intuition, especially on simulation methods.

Gary King. 1989. *Unifying political methodology*. University of Michigan Press. ISBN-10: 0472085549. ISBN-13: 978-0472085545.

Classic introduction to maximum likelihood for social scientists. Emphasis on intuition and basic deriviations.

J. Scott Long. 1997. *Regression models for categorical and limited dependent variables*. Sage Publications. ISBN-10: 0803973748. ISBN-13: 978-0803973749.

Clearest step-by-step introduction to the logic of logistic regression and related models.

Alain F. Zuur, Elena N. Ieno, and Erik H.W.G. Meesters. 2009. *A beginner's guide to R*. Springer. ISBN-10: 0387938362. ISBN-13: 978-0387938363.

Straightforward first course in R *programming. Assumes you know (or are learning) introductory statistics already.*

Optional books for further study

A. Colin Cameron and Pravin K. Trivedi. 2013. *Regression analysis of count data*. 2nd. Ed. Cambridge University Press.

Norman Matloff. 2011. *The Art of R Programming: A Tour of Statistical Software Design*. No Starch Press.

Richard McElreath. 2016. *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*. Chapman & Hall/CRC Press.

Will H. Moore and David A. Siegel. 2013. *A Mathematics Course for Political & Social Research*. Princeton University Press.

W. N. Venables and B. D. Ripley. 2010. *Modern applied statistics with S*. Fourth edition. Springer.

Required and optional articles

R. Michael Alvarez and Jonathan Nagler. 1998. "When Politics and Models Collide: Estimating Models of Multiparty Elections." *American Journal of Political Science*. 42(1) 55–96.

William D. Berry, Jacqueline H. R. DeMeritt, and Justin Esarey. 2010. "Testing for Interaction in Binary Logit and Probit Models: Is a Product Term Essential?" *American Journal of Political Science*. 54(1) 243–266.

Kosuke Imai and David A. van Dyk. 2005. "A Bayesian Analysis of the Multinomial Probit Model Using Marginal Data Augmentation." *Journal of Econometrics*. 124: 311–334.

King, Gary, Michael Tomz, and Jason Wittenberg. 2000. "Making the Most of Statistical Analyses: Interpretation and Presentation" *American Journal of Political Science* 44(2): 341–355.

Raftery, Adrian, 1995. "Bayesian Model Selection in Social Research (with Discussion)." *Sociological Methodology*. 111-196.

Brian Greenhill, Michael D. Ward, and Audrey Sacks. 2011. "The Separation Plot: A New Visual Method for Evaluating the Fit of Binary Models." *American Journal of Political Science*. 55(4): 990–1002.

Course outline

This outline of topics is a guideline and may be altered to meet course needs. In particular, the pace of the course may vary to make sure we are moving as fast as possible conditional on everyone understanding the material. Students should come to class having read the material for the next topic to be covered. Optional material is marked "(opt.)".

Part I: Fundamentals

28 September – 5 October: Introduction to the course, probability, and R

Readings: Handout on probability

King Ch. 3 (opt.)

Math review (opt.): `csss.washington.edu/academics/math-camp/lectures`

Zuur Ch 1–6 (opt. for R beginners)

Handout on matrix algebra

10–17 October: Overview of maximum likelihood estimation

Readings: Long Ch. 1, 2 (less detailed)

King Ch. 1, 2, 4 (more detailed)

PROBLEM SET 1 + PROJECT PROPOSAL DUE THURSDAY 12 OCTOBER IN CLASS

Part II: Modeling Choice

19–24 October: Estimating & Interpreting models of binary data

Readings: Long 3.1–3.9

King 5.1–5.3

Gelman & Hill 5.1–5.3, 6.4, 7.1–7.4

King, Tomz, and Wittenberg

Berry, DeMeritt, and Esarey (opt.)

PROBLEM SET 2 DUE TUESDAY 24 OCTOBER IN CLASS

26–31 October: Fitting & selecting models of binary data

Readings: Long 4
Gelman & Hill 5.4–5.8, 8.1–8.3
Greenhill, Ward, and Sacks (opt.)
Raftery (opt.)

2–7 November: Modeling ordered categorical data

Readings: Long 5.1–5.7
King 5.4

PROBLEM SET 3 DUE TUESDAY 7 NOVEMBER IN CLASS

9 November: Non-ordered data: Multinomial logit & Multinomal probit

Readings: Long 6
Gelman & Hill 6.5
Alvarez and Nagler
Imai and van Dyk

Part III: Modeling Counts

14–21 November: Poisson models of counts / Overdispersed & zero-inflated counts

Readings: Long 8.1–8.7
King 5.5–5.9
Gelman & Hill 6.2

PROBLEM SET 4 DUE THURSDAY 16 NOVEMBER IN CLASS

NO CLASS THURSDAY 24 NOVEMBER – THANKSGIVING

Part IV: Advanced Topics

28–30 November: Missing data and multiple imputation

PROBLEM SET 5 DUE TUESDAY 28 NOVEMBER IN CLASS

Part V: Student Poster Presentations

30 November – 7 December: Student poster presentations

Students will prepare and present a poster on their research projects in progress; this yields valuable feedback prior to final write-ups. Requirements and suggestions for poster construction will be discussed in class. Presentation dates will be assigned to minimize discrepancy with student preferences. Early presentations are expected to be less complete and evaluated accordingly. Students unable to present on certain days due to schedule conflicts should inform the instructor.

PROBLEM SET 6 DUE THURSDAY 7 DECEMBER IN CLASS

FINAL PAPER DUE TUESDAY 12 DECEMBER AT 3 PM
BOTH IN MY GOWEN MAILBOX AND BY EMAIL