Maximum Likelihood for Social Science

This volume provides a practical introduction to the method of maximum likelihood as used in social science research. Michael D. Ward and John S. Ahlquist focus on applied computation in $\mathcal{R}$ and use real social science data from actual, published research. Unique among books at this level, it develops simulation-based tools for model evaluation and selection alongside statistical inference. The book covers standard models for categorical data, as well as counts, duration data, and strategies for dealing with data missingness. By working through examples, math, and code the authors build an understanding about the contexts in which maximum likelihood methods are useful and develop skills in translating mathematical statements into executable computer code. Readers will not only be taught to use likelihood-based tools and generate meaningful interpretations, but they will also acquire a solid foundation for continued study of more advanced statistical techniques.

MICHAEL D. WARD is Professor Emeritus at Duke University. He has taught at Northwestern University, the University of Colorado, and the University of Washington. He worked as a principal research scientist at the WZB Berlin Social Science Center and held a municipal chair at the University of Pierre Mendès France (Grenoble II). His work began with a study of the links between global and national inequalities, continued with seminal articles on the conflict processes in the Cold War, and more recently turned to analyses of networks of conflict and cooperation in the contemporary era. At Duke, he established an innovative research lab of graduate and undergraduate students focusing on conflict prediction. One of the first political scientists to focus on the role of prediction in scholarly and policy work, he continues these efforts in his company, Predictive Heuristics, a data analytics firm that provides risk analysis for commercial and institutional clients.

JOHN S. AHLQUIST is Associate Professor of Political Economy at UC San Diego's School of Global Policy and Strategy and a 2017–18 Fellow at Stanford's Center for Advanced Study in the Behavioral Sciences. He previously held faculty positions at the University of Wisconsin, Madison, and Florida State University. His work has focused on the political structure and actions of labor unions, as well as the politics of redistribution and social insurance in a globalized economy. His methodological interests have spanned statistical models for network data, machine learning and cluster analysis, and the analysis of survey list experiments. He is author of more than twenty journal articles appearing in a variety of outlets, including the *American Journal of Political Science, American Political Science Review, Journal of Politics*, and *Political Analysis*. His most recent book (with Margaret Levi; 2013) is *In the Interest of Others*. He is a past winner of a variety of prizes, including the Mancur Olson Award, the Michael Wallerstein Award, and the APSA Labor Project Best Book Award. Ahlquist holds a PhD from the University of Washington and B.A. from UC Berkeley.

Analytical Methods for Social Research

Analytical Methods for Social Research presents texts on empirical and formal methods for the social sciences. Volumes in the series address both the theoretical underpinnings of analytical techniques as well as their application in social research. Some series volumes are broad in scope, cutting across a number of disciplines. Others focus mainly on methodological applications within specific fields such as political science, sociology, demography, and public health. The series serves a mix of students and researchers in the social sciences and statistics.

Series Editors

Other Titles in the Series

Time Series Analysis for the Social Sciences, by Janet M. Box-Steffensmeier, John R. Freeman, Jon C.W. Pevehouse and Matthew Perry Hitt
Event History Modeling: A Guide for Social Scientists, by Janet M. Box-Steffensmeier and Bradford S. Jones
Ecological Inference: New Methodological Strategies, edited by Gary King, Ori Rosen, and Martin A. Tanner
Spatial Models of Parliamentary Voting, by Keith T. Poole
Essential Mathematics for Political and Social Research, by Jeff Gill
Political Game Theory: An Introduction, by Nolan McCarty and Adam Meirowitz
Data Analysis Using Regression and Multilevel/Hierarchical Models, by Andrew Gelman and Jennifer Hill
Counterfactuals and Causal Inference, by Stephen L. Morgan and Christopher Winship

Maximum Likelihood for Social Science

Strategies for Analysis

MICHAEL D. WARD
Duke University

JOHN S. AHLQUIST
University of California, San Diego

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE
UNIVERSITY PRESS

University Printing House, Cambridge CB2 8BS, United Kingdom

One Liberty Plaza, 20th Floor, New York, NY 10006, USA

477 Williamstown Road, Port Melbourne, VIC 3207, Australia

314–321, 3rd Floor, Plot 3, Splendor Forum, Jasola District Centre, New Delhi – 110025, India

79 Anson Road, #06–04/06, Singapore 079906

Cambridge University Press is part of the University of Cambridge.

It furthers the University's mission by disseminating knowledge in the pursuit of education, learning, and research at the highest international levels of excellence.

www.cambridge.org
Information on this title: www.cambridge.org/9781107185821
DOI: 10.1017/9781316888544

First published 2018

A catalogue record for this publication is available from the British Library.

Library of Congress Cataloging-in-Publication Data
Names: Ward, Michael Don, 1948– author. | Ahlquist, John S., author.
Title: Maximum likelihood for social science : strategies for analysis/
Michael D. Ward, John S. Ahlquist.
Description: 1 Edition. | New York : Cambridge University Press, 2018. |
Series: Analytical methods for social research
Identifiers: LCCN 2018010101 | ISBN 9781107185821 (hardback) |
ISBN 9781316636824 (paperback)
Subjects: LCSH: Social sciences–Research. | BISAC: POLITICAL SCIENCE / General.
Classification: LCC H62 .W277 2018 | DDC 300.72–dc23
LC record available at https://lccn.loc.gov/2018010101

ISBN 978-1-107-18582-1 Hardback
ISBN 978-1-316-63682-4 Paperback

Cambridge University Press has no responsibility for the persistence or accuracy of URLs for external or third-party internet websites referred to in this publication and does not guarantee that any content on such websites is, or will remain, accurate or appropriate.

Cambridge University Press
978-1-107-18582-1 — Maximum Likelihood for Social Science
Michael D. Ward , John S. Ahlquist
Frontmatter
More Information

Contents

PART II MODEL EVALUATION AND INTERPRETATION

PART III THE GENERALIZED LINEAR MODEL

Cambridge University Press
978-1-107-18582-1 — Maximum Likelihood for Social Science
Michael D. Ward , John S. Ahlquist
Frontmatter
More Information

Cambridge University Press
978-1-107-18582-1 — Maximum Likelihood for Social Science
Michael D. Ward , John S. Ahlquist
Frontmatter
More Information

PART V A LOOK AHEAD

Cambridge University Press
978-1-107-18582-1 — Maximum Likelihood for Social Science
Michael D. Ward , John S. Ahlquist
Frontmatter
More Information

Figures

Cambridge University Press
978-1-107-18582-1 — Maximum Likelihood for Social Science
Michael D. Ward , John S. Ahlquist
Frontmatter
More Information

Tables

Preface

This project began many years ago at the University of Washington's Center for Statistics and the Social Sciences (CSSS). There two ambitious graduate students, John S. Ahlquist and Christian Breunig (now at the University of Konstanz), asked Michael D. Ward if he would supervise their training in maximum likelihood methods so that they could be better prepared for taking more advanced CSSS courses as well as those in the statistics and biostatistics departments. Ward gave them a stack of materials and asked them to start by preparing a lecture on ordinal regression models. Ward subsequently developed a class on maximum likelihood methods, which he has taught at the University of Washington (where it is still taught by Christopher Adolph) and, more recently, at Duke University. Ahlquist has gone on to teach a similar course at Florida State, the University of Wisconsin, and UC San Diego.

The point of the course was singular, and this book has a simple goal: to introduce social scientists to the maximum likelihood principle in a practical way. This praxis includes (a) being able to recognize where maximum likelihood methods are useful, (b) being able to interpret results from such analyses, and (c) being able to implement these methods both in terms of creating the likelihood and in terms of specifying it in a computational language that permits empirical analysis to be undertaken using the developed model.

The text is aimed at advanced PhD students in the social sciences, especially political science and sociology. We assume familiarity with basic probability concepts, the application of multivariate calculus to optimization problems, and the basics of matrix algebra.

OUR APPROACH

We take a resolutely applied perspective here, emphasizing core concepts, computation, and model evaluation and interpretation. While we include a chapter that introduces some of the important theoretical results and their derivations, we spend relatively little space discussing formal statistical properties. We made this decision for three reasons. First, there are several ways to motivate the likelihood framework. We find that a focus on a method's "desirable properties" in a frequentist setting to be a less persuasive reason to study maximum likelihood estimators (MLE). Instead we prefer to emphasize the powerful conceptual jump that likelihood-based reasoning represents in the study of statistics, one that enables us to move to a Bayesian setting relatively easily. Second, the statistical theory underlying the likelihood framework is well understood; it has been for decades. The requisite theorems and proofs are already collected in other excellent volumes, so we allocate only a single chapter to recapitulating them here. Rather, we seek to provide something that is missing: an applied text emphasizing modern applications of maximum likelihood in the social sciences. Third, and perhaps most important, we find that students learn more and have a more rewarding experience when the acquisition of new technical tools is directly bound to the substantive applications motivating their study.

Many books and even whole graduate training programs start with so-called Ordinary Least Squares (OLS). There is a certain logic to that. OLS is easy to teach, implement, and utilize while introducing a variety of important statistical concepts. OLS was particularly attractive in a world before powerful modern computers fit in our pockets. But OLS can be viewed as a special case of a more general class of models. Practically speaking, a limited range of social science data fit into this special case. Data in the social sciences tend to be lumpier, often categorical. Nominal, truncated, and bounded variables emerge not just from observational datasets but in researcher-controlled experiments as well (e.g., treatment selection and survival times). Indeed, the vast majority of social science data comes in forms that are profitably analyzed without resort to the special case of OLS. While OLS is a pedagogical benchmark, you will have to look hard for recent, state-of-the-art empirical articles that analyze observational data based on this approach. After reading this book and working through the examples, student should be able to fit, choose, and interpret many of the statistical models that appear in published research. These models are designed for binary, categorical, ordered, and count data that are neither continuous nor distributed normally.

WHAT FOLLOWS

We have pruned this book down from versions that appeared earlier online. We wanted the main text to focus entirely on the method and application of maximum likelihood principles. The text is divided into four parts.

Part I (Chapters 1–4) introduces the concept of likelihood and how it fits with both classical and Bayesian statistics. We discuss OLS only in passing, highlighting how, under certain assumptions, it can be thought of as a maximum likelihood estimator. We identify and derive the major theoretical results and then show how to apply them in the context of binary response variables. Chapter 4 provides a discussion of how MLE is implemented computationally, with a particular emphasis on the $\mathcal{R}$ computational environment.

Part II (Chapters 5 and 6) is the core of this volume. Its two chapters cover model selection and interpretation. Unique among texts at this level, we emphasize that model selection must occur prior to any inference about estimated parameters or other quantities. We argue explicitly for a wide application of an out-of-sample predictive heuristic in this area, something that is seeing increased attention with the machine learning revolution. In Chapter 6, we discuss how we might use the models we fit, and we focus on the power of modern computation to present nuanced and detailed interpretation of our statistical findings. We de-emphasize mechanical hypothesis testing against arbitrary null values, instead focusing on estimating meaningful quantities of interest and discussing our uncertainty around these estimates. In both chapters we include reflections on the mechanics and aesthetics of constructing tables and displays for effective communication, as well as thoughts on improving research transparency and credibility. While the material covered in this section is in no way unique to the study of maximum likelihood, we view this section as critical to continued progress in studying both maximum likelihood and more advanced statistical and computational topics.

Part III (Chapters 7–10) covers the Generalized Linear Model (GLM). Chapter 7 is short, introducing the basic structure of the GLM and some terminology and concepts. Chapters 8–10 present models for categorical variables, both ordered and nominal, as well as integer counts. Unlike some other texts for categorical data, these chapters are designed to be approached in a particular order and all rely on concepts and computational tools developed in Parts I and II.

In Part IV (Chapters 11 and 12) of the book we introduce more advanced topics. In Chapter 11 we begin the process of relaxing the standard assumption of conditional independence by presenting an introduction to duration models. This chapter is somewhat idiosyncratic, glossing over many of the details and complications one might expect in a full-fledged treatment of survival analysis, not to mention time series. Instead, we focus on how we can develop models for data that are inherently connected in time using likelihood tools and principles. Chapter 12 takes on the ubiquitous problem of missing data. We view this subject as woefully understudied in graduate training, while also presenting the pedagogical opportunity to discuss model construction and computation from a different perspective. We have also found that many of the most successful student replication projects came from critical interrogation of the earlier scholars' treatments of missing data.

Covering all the material in this book in a 15-week semester with beginning graduate students is certainly a challenge; doing so in a 10-week academic quarter is even more demanding. In a quarter-length course or with first-year students we have found that Chapters 2, 4, and 7 are better left as reference, instead emphasizing intuition, computation, and examples. When the temporal budget constraint binds, we typically allow student interest to determine whether we focus on duration models or missing data.

SPECIAL FEATURES

This volume contains several special features and sections that deserve further elaboration.

Real Examples from Published Research

Each chapter contains at least one example drawn from actual published social science research. These examples use real data drawn from scholars' data repositories to illustrate the models, highlight the modeling assumptions involved, and present detailed interpretations. All these datasets are archived in the online repository accompanying this volume.

$\mathcal{R}$ Code

This is an applied, computational text. We are particularly interested in helping students transform mathematical statements into executable computer code. $\mathcal{R}$ has become the dominant language in statistical computing because it is object-oriented, based on vectors; still has the best statistical graphics; and is open-source, meaning it is free to students and has a large network of contributors submitting new libraries almost daily. The newest statistical tools generally appear in $\mathcal{R}$ first.

We include code directly in the text in offset and clearly marked boxes. We include our own comments in the code chunks so students can see annotation clarifying computational steps. We also include $\mathcal{R}$ output and warnings in various places to aid in interpreting actual $\mathcal{R}$ output as well as trouble-shooting. All analysis and graphics are generated in $\mathcal{R}$. The online repository contains the $\mathcal{R}$ code needed to reproduce all tables and graphics.

"In case you were wondering ..."

Throughout the text there are special boxes labeled "In case you were wondering ...." The purpose of the boxes is to provide basic information about important mathematical tools and statistical distributions. These are things not easily defined in the main text and likely already familiar to some readers while appearing de novo to others. Our goal is to provide this information at the point of need while setting it off from the main text and marking it as "supplemental."

We do not refer to the boxes directly in the main text, unlike equations, tables, figures, and code chunks. The title of the boxes reflects their function and status; they present supplemental information for the curious.

"Further Reading"

Each chapter ends with a "further reading" section. These sections all follow a similar format, with subheadings for "applications," "past work," "advanced study," and "software notes," depending on the context these have for different topics.

The "applications" section highlights two to four studies using the tools discussed in that chapter and published in major social science journals in the last four years. These studies are meant to be examples of the types of papers students might consider when choosing replication projects.

The "past work" section is designed to provide pointers to the major contributors to the development and popularization of these tools in the social sciences. The "advanced study" section collects references to more advanced texts and articles where interested students can look for more detail on the math or computational algorithms. We consulted many of these texts in writing this book.

In the "software notes" sections we collect references to the major $\mathcal{R}$ libraries that we found useful in preparing the book or in conducting analysis ourselves. Since $\mathcal{R}$ is open-source, these references will surely become stale. We nevertheless thought it beneficial to collect references to $\mathcal{R}$ packages in a single place in each chapter.

NOTATION GLOSSARY

In our experience students often find mathematical notation a particularly frustrating barrier. To mitigate that problem we have included a notation "glossary" at the beginning of the book

ONLINE RESOURCES

The online repository, maxlikebook.com, accompanying this volume contains

- all datatsets used in this volume,
- $\mathcal{R}$ code for producing all tables and graphics,
- suggested problem sets and partial solutions, and
- some of our teaching slides.

We expect that repository content will evolve as we continue to teach this material and receive feedback from other instructors and students.

Acknowledgments

We would like to thank colleagues and former students worldwide, including Kathryn Alexander, Andrew Ballard, Sanuel Bagg, Jeanette Birnbaum, Anders Bjorn, Sarah Bouchat, Xun Cao, Chris Carrington, Mariana Carvalho Barbosa, Hannah Chapman, Cindy Cheng, Hsiao Chi, Andrew Cockrell, Amelia Cronan, Chris de Sante, Yao-yao Dai, Jared Daugherty, Matt Dickenson, Nhat-Dang Do, Zhang Dong, Cassy Dorff, Michael Duda, Mark Dudley, Josh Eastin, Idil Edes, Brian Engelsma, Brad Epperly, Amy Finnegan, Laura Frankel, Marcela García-Castañon, Asaph Glosser, Brian Greenhill, Ana Guzman, Andrew Heiss, John Holbein, Molly Hogan, Hsiao-Chi Hsu, Ashley Jochim, Arzu Kibris, Jin-Young Kim, Dimitrii Kofanov, Tobias Konitzer, Sabino Kornrich, Sophie Lee, Tae-dong Lee, Rosanna Shuk-Yin Lee, Brian Ho-Yin Lee, Ning Leng, Josh Lerner, Yuting Li, Nimah Mazaheri, Shahryar Minhas, Zoe Nemerever, Alexandra Oprea, Mathias Orlowski, Victoria Paniagua, Francisco Pedraza, Ryan Powers, Barry Pump, Ben Radford, Carlisle Rainey, Eric Schmidt, Kristan Seibel, Jeffry Smith, David Sparks, Won Steinbach, Tanja Srebotnjak, Rebecca Szper, Aaron Tsang, Ashley Thirkill, Jason Thomas, Linda Tran, Peter Vining, Samantha Vortherms, Austin Wang, Zach Warner, Simon Weschle, Undes Wen, Michelle Wolfe, Xinyan Xuan, McKenzie Young, Steve Zech, and countless others for β-testing this material in an unusual format. They were old-school: reading text, asking questions, finding mistakes, asking questions, doing homework, asking questions, etc.

Several colleagues gave us valuable feedback on the manuscript and the content of the courses on which it was based. We would like to thank Christian Breunig, Seth Hill, Susan Holmes, Will Moore, Molly Roberts, Alex Tahk, and Yiquing Xu. We are grateful to Michael Alvarez and Neal Beck for their advice, diligence, and high standards. Ethan J. Davis provided superb feedback as well.

In very different ways both authors received institutional support from the Center for Statistics and the Social Sciences (CSSS) at the University of

Washington, under the leadership of Adrian Raftery and subsequently by Thomas Richardson. This book would not have happened with the intellectual environment CSSS provided. Ahlquist benefited from the institutional support of the University of Wisconsin–Madison and from the School of Global Policy and Strategy as well as the department of political science at UC San Diego. Final editing of the book ocurred while Ahlquist was a fellow at the Center for Advanced Studies in the Behavioral Sciences at Stanford University. Joshua Penney was extremely helpful in the production process. We are grateful for his patience and skill in navigating the final process.

Notes on Notation

We generally follow notational standards common in applied statistics. But to a student, notation can often prove a barrier. This notation "glossary" is meant to ease the transition to reading notation-heavy material and provide a place to look up unfamiliar symbols. The underlying assumption is that students have already been introduced to basic probability, calculus, and linear algebra concepts.

Random variables and sets are denoted using script capitals. Thus, for example, $X = \{\ldots, -2, 0, 2, \ldots\}$ denotes the set of even integers. $Y \sim f_N(0, 1)$ states that Y is random variable that is distributed according to a Gaussian normal distribution with mean of 0 and variance of 1. We will denote the set of admissible values for X (its support) as $\mathcal{X}$.

Both upper- and lowercase letters can represent functions. When both upper- and lowercase versions of the same letter are used, the uppercase function typically represents the integral of the lowercase function, e.g., $G(x) = \int_{-\infty}^{x} g(u)du$.

To conserve notation we will use $f_s(\cdot;\theta)$ to represent the probability distribution and mass functions commonly used in building Generalized Linear Models. θ denotes generic parameters, possibly vector-valued. The subscript will denote the specific distribution:

- f_B is the Bernoulli distribution
- f_b is the binomial distribution
- f_β is the Beta distribution
- f_c is the categorical distribution
- f_e is the exponential distribution
- f_{EV_1} is the type-I extreme value distribution
- f_Γ is the Gamma distribution
- f_{GEV} is the generalized extreme value distribution
- f_L is the logistic distribution

- f_{lL} is the log-logistic distribution
- f_m is the multinomial distribution
- $f_{\mathcal{N}}$ is the Gaussian (Normal) distribution
- f_{Nb} is the negative binomial
- f_P is the Poisson distribution
- F_W is the Weibull distribution

To conform with conventional terminology and notation in $\mathcal{R}$, we refer to one-dimensional vectors as *scalars*. Scalars and observed realizations of random variables are denoted using lowercase math script. $\Pr(Y_i \leq y_i)$ denotes the probability that some random variable, Y_i, takes a value no greater than some realized level, y_i.

Matrices are denoted using bolded capital letters; $\mathbf{X}_{n\times k}$ is the matrix with n rows and k columns. The symbol $\intercal$ denotes matrix or vector transposition, as in $\mathbf{X}^\intercal$. Vectors are represented with bolded lowercase letters, e.g., $\mathbf{x}_i$. In our notation we implicitly treat all vectors as *column* vectors unless otherwise stated. For example, $\mathbf{x}_i$ is a column vector even though it may represent a row in the $\mathbf{X}_{n\times k}$ matrix. "Barred" items denote the sample mean e.g., $\bar{\mathbf{y}}$.

Lowercase Greek letters are typically reserved for parameters of models and statistical distributions. These parameters could be either scalars or vectors. Vectors will be expressed in bold font. Where more specificity is needed we will subscript.

"Hatted" objects denote fitted or estimated quantities; when used in the context of an MLE then hatted objects are the MLE. For example, β might be a regression parameter and $\hat{\beta}$ is the estimated value of that parameter.

Common functions, operators, and objects:

- $\propto$ means "is proportional to"
- $\dot{\sim}$ means "approximately distributed as"
- $\xrightarrow{d}$ means "convergence in distribution."
- $\xrightarrow{p}$ means "convergence in probability," what some texts denote plim.
- $\mathbb{1}(\cdot)$ is the indicator function that returns a 1 if true and a 0 otherwise.
- $\text{cov}(\cdot,\cdot)$ is the covariance function
- $\det(\cdot)$ is the determinant of a square matrix
- $\text{E}[\cdot]$ is the expectation operator
- $\exp(\cdot)$ is the exponential function
- $\Gamma(\cdot)$ is the Gamma function
- $\mathbf{I}_n$ is the $n \times n$ identity matrix
- $\mathcal{I}(\cdot)$ is the expected Fisher information
- $I(\cdot)$ is the observed Fisher information
- *iid* means "independently and identically distributed."
- $\Lambda(\cdot)$ is the logistic cumulative distribution function
- log is the logarithm. If no base is given, then it denotes the natural logarithm (base e)

- ∇ is the gradient vector of some function.
- $\Phi(\cdot)$ is the standard Normal cumulative distribution function
- $\phi(\cdot)$ is the standard Normal density function
- $\Pr(\cdot)$ denotes probability
- $\text{var}(\cdot)$ is the variance function