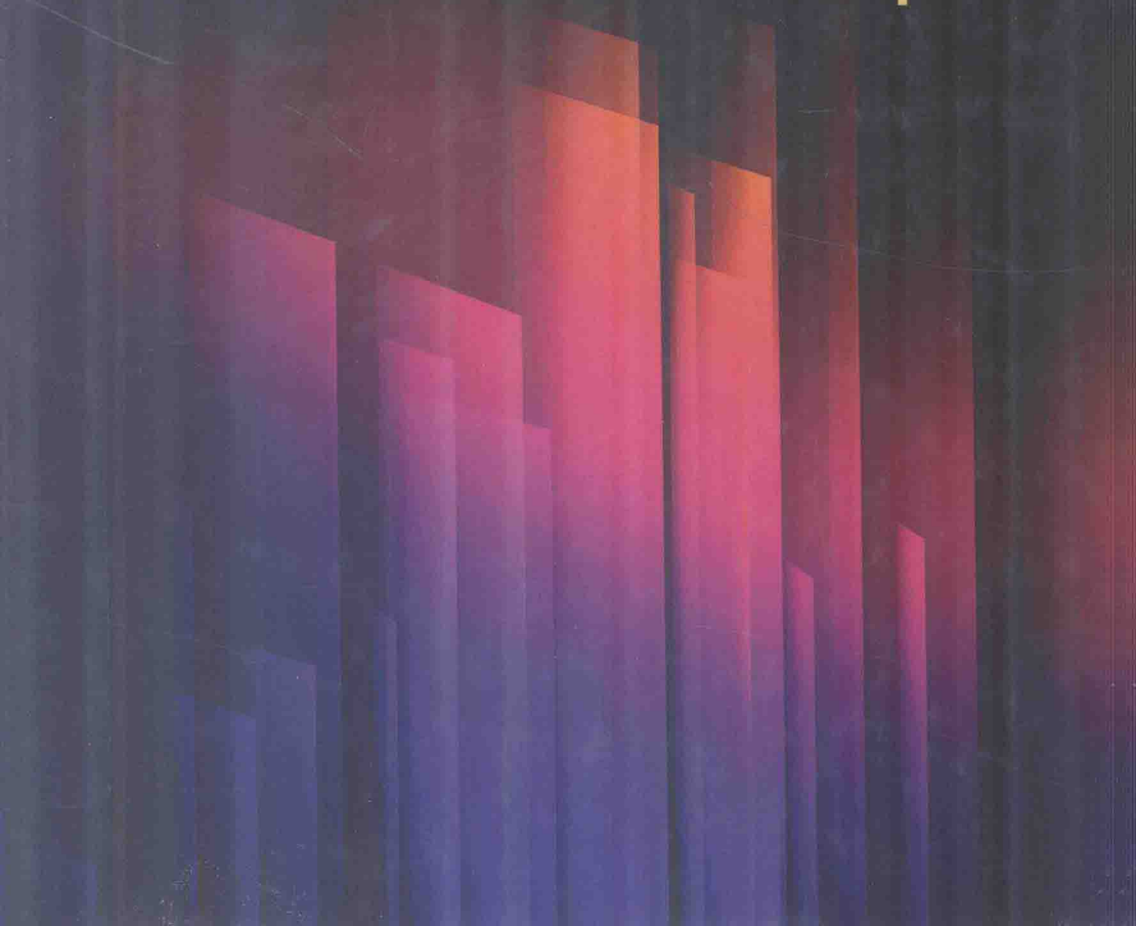
Bayesian Statistics for the Social Sciences

David Kaplan



Bayesian Statistics for the Social Sciences

David Kaplan

Series Editor's Note by Todd D. Little

THE GUILFORD PRESS
New York London

A Division of Guilford Publications, Inc.
72 Spring Street, New York, NY 10012
www.guilford.com

Printed in the United States of America

This book is printed on acid-free paper.

Last digit is print number: 9 8 7 6 5 4 3 2 1

Library of Congress Cataloging-in-Publication Data

Kaplan, David, 1955–
Bayesian statistics for the social sciences / David Kaplan.
pages cm. — (Methodology in the social sciences)
Includes bibliographical references and index.
ISBN 978-1-4625-1651-3 (hardback)
1. Social sciences—Statistical methods. 2. Bayesian statistical decision theory. I. Title.
HA29.K344 2014
519.5'42—dc23 2014017208

Bayesian Statistics for the Social Sciences

Methodology in the Social Sciences

David A. Kenny, Founding Editor
Todd D. Little, Series Editor
www.guilford.com/MSS

This series provides applied researchers and students with analysis and research design books that emphasize the use of methods to answer research questions. Rather than emphasizing statistical theory, each volume in the series illustrates when a technique should (and should not) be used and how the output from available software programs should (and should not) be interpreted. Common pitfalls as well as areas of further development are clearly articulated.

RECENT VOLUMES

APPLIED MISSING DATA ANALYSIS
Craig K. Enders

PRINCIPLES AND PRACTICE OF STRUCTURAL EQUATION MODELING, THIRD EDITION
Rex B. Kline

APPLIED META-ANALYSIS FOR SOCIAL SCIENCE RESEARCH
Noel A. Card

DATA ANALYSIS WITH Mplus
Christian Geiser

INTENSIVE LONGITUDINAL METHODS: AN INTRODUCTION TO DIARY AND EXPERIENCE SAMPLING RESEARCH
Niall Bolger and Jean-Philippe Laurenceau

DOING STATISTICAL MEDIATION AND MODERATION
Paul E. Jose

LONGITUDINAL STRUCTURAL EQUATION MODELING
Todd D. Little

INTRODUCTION TO MEDIATION, MODERATION, AND CONDITIONAL PROCESS ANALYSIS: A REGRESSION-BASED APPROACH
Andrew F. Hayes

BAYESIAN STATISTICS FOR THE SOCIAL SCIENCES
David Kaplan

To Sam, who taught me that babies are Bayesians

Series Editor's Note

I have known David Kaplan for a number of years now. We are both members of the Society of Multivariate Experimental Psychology and the American Psychological Association; we also served together on an Institute of Education Sciences standing review panel and overlapped as Associate Editors for *Multivariate Behavioral Research.* He and one of his students also contributed a terrific chapter to one of the handbooks that I edited. When I see him at the annual conferences of these societies or at panel meetings, he is regularly engaged in deep intellectual discussions with others in attendance because they seek him out for his guidance and input on their own research. I have benefited from his scholarly acumen in this manner a number of times. Given my admiration and respect for him, when David first mentioned that he would like to contribute a book to the Guilford series Methodology in the Social Sciences, I was elated. When he mentioned that the book would be about Bayesian procedures, I was even happier. When he mentioned that he would use the R software platform for all of his working examples, I reached the peak of elation.

David Kaplan is in a very elite class of scholar. He is a methodological innovator who is guiding and changing the way that researchers conduct their research and analyze their data. He is also a distinguished educational researcher whose work shapes educational policy and practice. I see David's book as a reflection of his sophistication as both a researcher and a statistician; it shows depth of understanding that even dedicated quantitative specialists may not have and, in my view, it will have an enduring impact on research practice. David's research profile and research skills are renowned internationally and his reputation is globally recognized. His profile as a prominent power player in the field brings instant credibility. As a result, when David says Bayesian is the way to go, researchers listen. Now his book brings his voice to you in an engaging and highly serviceable manner.

Why is the Bayesian approach to statistics seeing a resurgence across the social and behavioral sciences? (It's an approach that has been around for some time.) One reason for the delay in adopting Bayes is technological. Bayesian estimation can be computer intensive and until about a score of years ago the computational demands limited its widespread application. Another reason is that the social and behavioral sciences needed an accessible translation of Bayes for these fields so that we could understand not only the benefits of Bayes but also how to apply a Bayesian approach. David is clear and practical in his presentation and shares with us his experiences and helpful and pragmatic recommendations. I think the Bayesian perspective will see a spirited adoption now that David has penned this indispensable resource. In many ways, the zeitgeist for Bayes is favorable—because researchers are asking and attempting to answer complex questions.

The blind empiricism of frequentist thinking is giving way to a modeling perspective. Complex theoretical models abound in social science research. Such models are most informative if the modeler has a strong theory and good data. Some analysts will argue that data should not get in the way of good theory and others will argue to never let theory get in the way of good data. Neither position, however, will yield useful and generalizable findings. An informed dialogue with data—the heart of the Bayesian perspective—is essential for good data models. The theory that drives the modeling endeavor derives from the modeler's intuitions, born from experience and informed by the extant literature. These intuitions are coupled with an informed understanding of the utilized design and the acquired measures. These critical ingredients are carefully mixed to specify a Bayesian statistical model grounded in prior knowledge and insights. The model is then estimated against the data and the conversation thereby begins. As an iterative process, finalizing a statistical model is a process of commensuration. Model modifications must balance improvement in model fit and estimation precision with the verisimilitude of any model changes. Model modifications are statements of theory that emerge when the modeler has carefully balanced errors of the type I and II variety. The modifications must be reconciled in the context of the larger model and the broader literature. Blind allegiance to theory or unequivocal adherence to data—hallmarks of traditionally trained researchers—will not deliver useful knowledge. Optimal gains in knowledge can only occur when we engage in an informed dialogue with data.

TODD D. LITTLE
Atlanta's Hartsfield–Jackson International Airport

Preface

Bayesian statistics has long been overlooked in the quantitative methods training of social scientists. Typically, the only introduction that a student might have to Bayesian ideas is a brief overview of Bayes' theorem while studying probability in an introductory statistics class. This is not surprising. First, until recently, it was not feasible to conduct statistical modeling from a Bayesian perspective owing to its complexity and the lack of available software. Second, Bayesian statistics represents a powerful alternative to frequentist (classical) statistics, and is, therefore, controversial.[1]

Recently, however, there has been a renaissance in the developments and application of Bayesian statistical methods, due mostly to the availability of powerful statistical software tools. As a result, scores of books have been written over the last 10 years and at a variety of levels that lead the reader through Bayesian theory and computation. The goal of this book is to introduce practicing social scientists to the Bayesian perspective via methodologies commonly used in the social sciences. Thus, this book is written for social scientists who are well trained in statistical modeling within the frequentist paradigm and are interested in exploring commonly used methodologies from the Bayesian perspective. In addition, this book is written for graduate students in the social sciences who wish to have an accessible entrée into Bayesian statistics.

[1] I use the term *frequentist* to describe the paradigm of statistics commonly used today; it represents the counterpart to the Bayesian paradigm of statistics. Historically, however, Bayesian statistics predates frequentist statistics by about 150 years.

I do not presume that the reader has had any exposure to Bayesian statistics, and I develop the arguments in favor of the Bayesian approach from first principles. That said, it is assumed that the reader does have a good background in applied statistics—including a good command of frequentist hypothesis testing, as well as a background in applied statistical modeling, preferably including courses in regression analysis, analysis of variance, and at least some exposure to multilevel modeling and structural equation modeling. Some background in elementary calculus is useful, but I make every attempt to explain equations thoroughly. Derivations are relegated to appendices unless needed to reinforce conceptual understanding.

The book is organized into three parts. Part I covers the foundations of Bayesian statistics. Chapter 1 provides an introduction to frequentist versus Bayesian probability and introduces Bayes' theorem. Chapter 2 introduces the elements of Bayesian statistical inference, including the concept of exchangeability, likelihood, prior and posterior distributions, and the Bayesian central limit theorem. Chapter 3 introduces some important distributions used in the social sciences. For each distribution, I indicate how it is used in practice and provide its conjugate prior distribution for use in Bayesian analysis. Chapter 4 introduces the method of Markov chain Monte Carlo estimation.

Part II introduces Bayesian model building and Bayesian general and generalized linear modeling. Chapter 5 discusses the topic of Bayesian hypothesis testing and model building, highlighting the fundamental differences between the Bayesian approach and the frequentist approach. Chapter 6 discusses Bayesian linear regression analysis and introduces extensions to Bayesian generalized linear modeling—particularly focusing on Bayesian logistic regression. Chapter 7 ends Part II with a discussion of missing data problems from a Bayesian context.

Part III extends Bayesian statistics to advanced, but popular, methodologies in the social sciences. Chapter 8 discusses Bayesian approaches for multilevel modeling. Chapter 9 examines Bayesian models for continuous and categorical latent variables—focusing on confirmatory factor analysis, structural equation modeling, growth curve modeling, and growth mixture modeling.

As the outline of the book suggests, my aim is to cover Bayesian approaches to the main methodologies that are currently used in the social sciences. Moreover, to support the pedagogical features of this book, the data and software code for each example are available on the book's website (*http://bise.wceruw.org/publications.html*). However, I don't believe that a book on Bayesian statistical inference would be complete without a discussion of the philosophical issues that underlie Bayesian methodologies. Therefore, Chapter 10 covers the main philosophical underpinnings of Bayesian statistical inference, and in particular the theory of subjective probability as a framework for addressing problems of uncertainty and the growth of knowledge in science. I draw on the seminal writings of de Finetti, Lindley, Savage, Jeffreys, Jaynes, Berger, and Bernardo, and I contrast the framework of subjective Bayesian statistics with that of objective Bayesian statistics.

DATA AND SOFTWARE

The examples provided throughout this book utilize data from large-scale educational surveys. Particular focus is on analyses using data from the Program for International Student Assessment sponsored by the Organisation for Economic Co-operation and Development. For analysis of longitudinal data in Chapter 9, I utilize data from the Longitudinal Study of American Youth (Miller, Hoffer, Sucher, Brown, & Nelson, 1992).

This book relies entirely on statistical programs available via the R statistical programming environment (R Development Core Team, 2012). My reason for focusing on R is twofold. First, R is the lingua franca of statistical computing, and thus it is important to demonstrate its power for statistical computing in a Bayesian context. Second, R is an open-source software program, and, as such, the software code is available for inspection, modification, and new dissemination. However, it is not the purpose of this book to introduce the reader to the R programming environment, nor will I provide new R programs specific to the examples in this book. Rather, I draw on a set of programs for Bayesian analysis already available in the Comprehensive R Archive Network (CRAN; R Development Core Team, 2012), which I have found to be particularly flexible and useful. Specifically, many of the examples in this book will rely on programs within the R package "MCMCpack" (A. D. Martin, Quinn, & Park, 2010). Bayesian computation diagnostics make use of the R package "coda" (Plummer, Best, Cowles, & Vines, 2006), and Bayesian model averaging makes use of the R package "BMA" (Raftery, Hoeting, Volinsky, Painter, & Yeung, 2009). Some examples in this book utilize the R interface with the "JAGS" software program (Plummer, 2003) referred to as "rjags" (Plummer, 2011), which closely resembles "WinBUGS" (Spiegelhalter, Thomas, Best, & Lunn, 2003). In this way, the reader does not have to master the R language and can easily adapt the example programs to fit individual needs.

The programs that I have chosen in this book are not exhaustive of the scores of programs available on the CRAN and should not be taken as a specific endorsement of these programs over any others. A perusal of the Bayesian section of the CRAN Task View page (*http://cran.open-source-solution.org/web/views/Bayesian.html*) provides the reader with a detailed list of Bayesian programs available on the CRAN. Finally, all R programs used in this book are made available in chapter appendices and also on the book's website.

Acknowledgments

I am indebted to Jianshen Chen and Soojin Park, my two outstanding doctoral students, for assisting in the data analyses and graphics that appear in this book. Without question, it would have been difficult to complete this book in a reasonable period of time without their assistance. I also wish to especially thank Fabrizzio Sanchez for coming on late to this project but still providing expert R programming support.

I would like to thank my colleagues Daniel Bolt and Jee-Seon Kim, along with the reviewers who were initially anonymous: Feifei Ye, Department of Psychology in Education, University of Pittsburgh; Jay Myung, Department of Psychology, Ohio State University; Jim Albert, Department of Mathematics and Statistics, Bowling Green State University; John J. McArdle, Department of Psychology, University of Southern California; and David Rindskopf, Department of Educational Psychology, City University of New York. All of these scholars' comments have greatly improved the quality and accessibility of the book. Of course, any errors of commission or omission are strictly my responsibility.

I am forever grateful to C. Deborah Laughton, my editor at The Guilford Press, who has been a constant source of encouragement, support, and friendship throughout the development of this book and well beyond. Her professionalism is without peer, and one could not ask for a better editor.

The development of this book was supported in part by a Kellett Mid-Career Award from the University of Wisconsin–Madison and by the Institute of Education Sciences, U.S. Department of Education, through Grant No. R305D110001 to the University of Wisconsin–Madison. The opinions expressed are my own and do not necessarily represent the views of the Institute or the U.S. Department of Education.

Finally, as always, this book would not be possible without the love and support of my family—Allison, Rebekah, Hannah, my son-in-law, Joshua, and my grandson, Sam, to whom this book is dedicated.

Contents