Statistics for Social and Behavioral Sciences

Advisors:

S.E. Fienberg
W.J. van der Linden

For other titles published in this series, go to
http://www.springer.com/series/3463

Jean-Paul Fox

Bayesian Item Response Modeling

Theory and Applications

Springer

Jean-Paul Fox
Department of Research Methodology,
Measurement, and Data Analysis
Faculty of Behavioral Sciences
University of Twente
7500 AE Enschede
The Netherlands

ISBN 978-1-4419-0741-7 e-ISBN 978-1-4419-0742-4
DOI 10.1007/978-1-4419-0742-4
Springer New York Dordrecht Heidelberg London

Library of Congress Control Number: 2010927930

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

To Jasmijn and Kae

Preface

The modeling of item response data is governed by item response theory, also referred to as modern test theory. The field of inquiry of item response theory has become very large and shows the enormous progress that has been made. The mainstream literature is focused on frequentist statistical methods for estimating model parameters and evaluating model fit. However, the Bayesian methodology has shown great potential, particularly for making further improvements in the statistical modeling process.

The Bayesian approach has two important features that make it attractive for modeling item response data. First, it enables the possibility of incorporating nondata information beyond the observed responses into the analysis. The Bayesian methodology is also very clear about how additional information can be used. Second, the Bayesian approach comes with powerful simulation-based estimation methods. These methods make it possible to handle all kinds of priors and data-generating models.

One of my motives for writing this book is to give an introduction to the Bayesian methodology for modeling and analyzing item response data. A Bayesian counterpart is presented to the many popular item response theory books (e.g., Baker and Kim 2004; De Boeck and Wilson, 2004; Hambleton and Swaminathan, 1985; van der Linden and Hambleton, 1997) that are mainly or completely focused on frequentist methods. The usefulness of the Bayesian methodology is illustrated by discussing and applying a range of Bayesian item response models.

Complex Assessments

The recognition of the complexity of the processes leading to responses on test items stimulated the development of more realistic response models. The flexibility of the Bayesian modeling framework makes it particularly useful for making proper adjustments when the response data violate common model assumptions. Such violations might appear due to flaws in the data collection procedure or the complexity of the sample design.

Individuals' performances are best measured under controlled conditions such that other sources of variation are under control. An experimenter will choose a test for measuring a construct consisting of a set of items that minimizes individual variation and emphasize differences between subjects. Variability between respondents due to factors other than the construct under consideration is not desirable. Measurement models can become quite complex when they are adjusted in such a way that all sources of variation are taken into account. Flaws in experimental setups may engender the need for more complex measurement models. The impact of context effects on assessment results may further complicate the modeling process. Context effects appear when items function differently due to factors like item positioning or other material correlated with the item. Test data that violate assumptions of common item response models require a more flexible model that accounts for the violations.

Other complexities may arise due to the fact that besides response information other kinds of information are known (e.g., individual or group characteristics, response times, item characteristics). Different sampling designs can be used to sample respondents and/or items (e.g., adaptive item sampling, multistage sampling, randomized response sampling, simple random sampling, stratified sampling). Different response formats (e.g., multiple choice, binary, polytomous) and the presence of clusters of items may further stimulate the use of a more complex measurement model.

Besides introducing the Bayesian methodology, my aim has been to write a book to introduce Bayesian modeling of response data from complex assessments. Often information beyond the observed response data is available that can be used. The Bayesian modeling framework will prove to be very flexible, allowing simultaneous estimation of model parameters, computation of complex statistics, and simultaneous hypothesis testing.

In the 1990s, Bayesian inference became feasible with the introduction of Bayesian computational methods such as computer simulation and Monte Carlo techniques. The development of powerful computational simulation techniques induced a tremendous positive change in the applicability of Bayesian methodology. This led to the development of more flexible statistical models for test theory but also different strategies with respect to parameter estimation and hypothesis testing. In this book, the Bayesian way of item response modeling combined with the development of powerful numerical simulation techniques that led to a new research area in modern test theory is outlined.

Outside the Scope of This Book

Designing tests and testing whether tests are suited for the intended purpose are very complex subjects. Various questions need to be answered with respect to the response format of the test, the purpose of the test, and the construction of test materials, among others. The tests developed should also be reliable

and valid; that is, consistently result in scores that reflect the construct level of each respondent and measure what they are supposed to measure. Good tests are discriminating in the sense that they show differences in the construct level of respondents. There are a number of sources where this information is readily available. For classical test theory, see, for example, Gulliksen (1950), and Lord and Novick (1968), and for item response theory, see, for example, Lord and Novick (1968) and Lord (1980). A manual of standards for the construction and use of tests has been prepared by a joint committee of the American Educational Research Association, American Psychological Association and National Council of Measurement in Education (2000).

Overview

Statistical computations are necessary for applying the Bayesian methodology, and some programming skills are needed. That is, some familiarity with a statistical software package like R or S+ is needed to perform Bayesian analysis. On the one hand, this book aims to serve those who just want to apply the models, and they can use the software implemented in R packages and S+ programs (see Section 1.5). On the other hand, others may want to learn via programming and/or implement codes by themselves to extend models or adjust priors. For them, the mathematical details of the estimation procedures are discussed in the book, and the computer codes are provided via a website associated with the book. To understand the material, a basic background in probability and statistics is needed, including some familiarity with matrix algebra at the undergraduate level. The contents as well as the algorithms with their implementations make this book self-contained. Hopefully, it will provide an introduction to the essential features of Bayesian item response modeling as well as a better understanding of more advanced topics. The contents, programs, and codes will hopefully help readers implement their own algorithms and build their own set of tools for Bayesian item response modeling.

The book is organized as follows. In Chapter 1, the typical structure of item response data and the common item response models are discussed. Basic elements of Bayesian response modeling are introduced together with the basic building blocks for making Bayesian statistical inferences. WinBUGS is used to illustrate the Bayesian modeling approach. Chapter 2 presents a hierarchical modeling approach that supports the pooling of information, which becomes important when typically limited information is observed about many individuals. The Bayesian hierarchical modeling approach is outlined, which has tremendous potential with the current developments in statistical computing. Before discussing various sampling-based estimation methods for Bayesian item response models, which will be discussed in Chapter 4, in Chapter 3 a more general introduction is given to sampling-based estimation methods, testing hypotheses, and methods for model selection. Chapter 5 discusses methods for testing hypotheses and for model selection for the Bayesian item response models described in Chapter 4.

In Chapters 6–9, more advanced item response models are discussed for response data from complex assessments, response and response time data, and responses from complex sampling designs. In Chapter 6, respondents are assumed to be nested in groups (e.g., schools, countries). A hierarchical population model for respondents is defined to account for the within- and between-group dependencies. In Chapter 7, models for relaxing common measurement invariance restrictions are discussed. Chapter 8 introduces the multivariate analysis of responses and response times for measuring the speed and accuracy of working. Chapter 9 introduces models for (randomized) response data that are masked before they are observed to invite respondents to answer honestly when asked sensitive questions. Several empirical examples are presented to illustrate the methods and the usefulness of the Bayesian approach.

Acknowledgments

I would like to thank numerous people for their assistance and/or input during the writing of this book. I acknowledge the input from collaborators on earlier research projects that were addressed in the book. The cooperation of Rinke Klein Entink, Cees Glas, Wim van der Linden, Martijn de Jong, and Jan-Benedict Steenkamp has been greatly appreciated. I am indebted to Jim Albert, who provided very helpful comments on earlier drafts. Cheryl Wyrick has kindly provided data for a randomized response application. I thank my colleagues at the Department of Research Methodology, Measurement, and Data Analysis at the University of Twente. My colleagues Rinke Klein Entink and Josine Verhagen read drafts of the book, and their suggestions and comments led to its substantial improvement. I also thank John Kimmel for his confidence and assistance during the preparation of the book. The VIDI grant of the Netherlands Organization for Scientific Research (NWO) supported the writing of this book, for which I am most grateful.

Finally, I thank my wife, Miranda, for her support, encouragement, and patience during the writing of this book.

University of Twente, Enschede *Jean-Paul Fox*
March 2010

Contents