Methodology of Educational Measurement and Assessment

Matthias von Davier
Young-Sun Lee *Editors*

Handbook of Diagnostic Classification Models

Models and Model Extensions, Applications, Software Packages

Springer

Methodology of Educational Measurement and Assessment

This book series collates key contributions to a fast-developing field of education research. It is an international forum for theoretical and empirical studies exploring new and existing methods of collecting, analyzing, and reporting data from educational measurements and assessments. Covering a high-profile topic from multiple viewpoints, it aims to foster a broader understanding of fresh developments as innovative software tools and new concepts such as competency models and skills diagnosis continue to gain traction in educational institutions around the world. Methodology of Educational Measurement and Assessment offers readers reliable critical evaluations, reviews and comparisons of existing methodologies alongside authoritative analysis and commentary on new and emerging approaches. It will showcase empirical research on applications, examine issues such as reliability, validity, and comparability, and help keep readers up to speed on developments in statistical modeling approaches. The fully peer-reviewed publications in the series cover measurement and assessment at all levels of education and feature work by academics and education professionals from around the world. Providing an authoritative central clearing-house for research in a core sector in education, the series forms a major contribution to the international literature.

More information about this series at http://www.springer.com/series/13206

Matthias von Davier • Young-Sun Lee
Editors

Handbook of Diagnostic Classification Models

Models and Model Extensions, Applications, Software Packages

Springer

Editors
Matthias von Davier
National Board of Medical Examiners (NBME)
Philadelphia, PA, USA

Young-Sun Lee
Teachers College
Columbia University
New York, NY, USA

ISSN 2367-170X ISSN 2367-1718 (electronic)
Methodology of Educational Measurement and Assessment
ISBN 978-3-030-05583-7 ISBN 978-3-030-05584-4 (eBook)
https://doi.org/10.1007/978-3-030-05584-4

This Springer imprint is published by the registered company Springer Nature Switzerland AG.
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

The *Handbook of Diagnostic Classification Models* represents a collection of chapters reviewing diagnostic models, their applications, and descriptions of software tool, written by leading experts in the field. This volume covers most (one can never claim completeness) of the current major modeling families and approaches as well as provides a resource that can be used for self-study, teaching, or research that applies or extends the materials included in the book.

While virtually any project of this type takes longer than expected, and many will be tempted to remind us that Murphy's law strikes almost surely, we were amazed by the willingness of all contributors to put in the hours to finish their chapters and to review other chapters and, finally, to revise their contributions in order to help putting together a coherent volume. We hope that this process, together with some occasional assistance from the editors and the publisher, helped to compile a multi-authored work together that covers most aspects of doing research around diagnostic modeling.

We also want to remind readers as well as ourselves of colleagues who passed away and who leave a void in the research community. We lost Kikumi Tatsuoka, of whom one can truthfully say that her rule space approach is one of the major roots, maybe even the most important one, of this field. In her long career, she shaped many aspects of diagnostic modeling, and we should recall that, among these, the Q-matrix is one of the central building blocks present in the vast majority of these methods. The rule space method is described along with other early approaches in Chap. 1.

We furthermore would like to remember Lou DiBello, who made important contributions to the field, notably in his modified rule space work, and his work on the unified model together with colleagues. The work around extensions of the unified model is described in Chap. 3. We also want to remind readers of Wen-Chung Wang who just recently passed away. Wen-Chung and his coauthors worked on many topics around diagnostic models and other psychometric approaches. His work around DIF methods for use with diagnostic modeling approaches is found in Chap. 18. We hope that the friends we lost would have liked this volume.

Ending on a more positive note: working in a dynamic field that produces new knowledge every day, we are aware that the handbook is one stepping stone on the long path to fully understanding the potential of these powerful modeling approaches. We are expecting to see books that extend the material we have put together here; moreover, we expect to see this handbook be replaced or superseded by a new edition in a couple of years. If we are lucky, we may be involved in putting together some of the chapters of these future collections describing what will then be the state of the art in diagnostic modeling.

Philadelphia, PA, USA — Matthias von Davier
New York, NY, USA — Young-Sun Lee

Contents

Contributors

Stephan Abele Institute of Vocational Education and Vocational Didactics, Technische Universität Dresden, Dresden, Germany

Russell G. Almond Department of Educational Psychology and Learning Systems, Florida State University, Tallahassee, FL, USA

Daniel M. Bolt Department of Educational Psychology, University of Wisconsin – Madison, Madison, WI, USA

Li Cai University of California, Los Angeles, CA, USA

Vector Psychometric Group, LLC, Chapel Hill, NC, USA

Hua-Hua Chang Department of Educational Studies, Purdue University, IN, USA

Yunxiao Chen London School of Economics and Political Science, London, UK

Ying Cheng Department of Psychology, University of Notre Dame, Notre Dame, IN, USA

Chia-Yi Chiu Department of Educational Psychology, Rutgers, The State University of New Jersey, New Brunswick, NJ, USA

Pravin Chopade ACTNext ACT Inc., Iowa City, IA, USA

Steven Andrew Culpepper Department of Statistics, University of Illinois at Urbana-Champaign, Champaign, IL, USA

Jimmy de la Torre Division of Learning, Development and Diversity, University of Hong Kong, Hong Kong, China

Lawrence T. DeCarlo Department of Human Development, Teachers College, Columbia University, New York, NY, USA

Benjamin Deonovic ACTNext ACT Inc., Iowa City, IA, USA

Lou DiBello Department of Statistics, University of Illinois at Urbana-Champaign, Urbana, IL, USA

Jeff Douglas Department of Statistics, University of Illinois at Urbana-Champaign, Champaign, IL, USA

Susan E. Embretson School of Psychology, Georgia Institute of Technology, Atlanta, GA, USA

Meghan Fager National University, Precision Institute, La Jolla, CA, USA

Ann Cathrice George Federal Institute for Educational Research, Innovation and Development of the Austrian School System, Salzburg, Austria

Zhuangzhuang Han Department of Human Development, Teachers College, Columbia University, New York, NY, USA

Robert Henson Educational Research Methodology (ERM) Department, The University of North Carolina at Greensboro, Greensboro, NC, USA

Carrie R. Houts Vector Psychometric Group, LLC, Chapel Hill, NC, USA

Hong Jiao Department of Human Development and Quantitative Methodology, University of Maryland, College Park, MD, USA

Matthew S. Johnson Educational Testing Service, Princeton, NJ, USA

Hyeon-Ah Kang Department of Educational Psychology, University of Texas, Austin, TX, USA

Lale Khorramdel National Board of Medical Examiners (NBME), Philadelphia, PA, USA

Hans-Friedrich Köhn Department of Psychology, University of Illinois at Urbana-Champaign, Champaign, IL, USA

Young-Sun Lee Teachers College, Columbia University, New York, NY, USA

Xiaomin Li The University of Hong Kong, Pokfulam, Hong Kong

Xiaoou Li School of Statistics, University of Minnesota, Minneapolis, MN, USA

Dandan Liao American Institutes for Research, Washington, DC, USA

Jingchen Liu Department of Statistics, Columbia University, New York, NY, USA

Xiang Liu Department of Human Development, Teachers College, Columbia University, New York, NY, USA

Diego A. Luna-Bazaldua School of Psychology, National Autonomous University of Mexico, Mexico City, Mexico

Wenchao Ma The University of Alabama, Tuscaloosa, AL, USA

Nathan D. Minchen Pearson, Bronx, NY, USA

Jesse Pace University of Kansas, Lawrence, KS, USA

Yoon Soo Park Department of Medical Education, College of Medicine, University of Illinois at Chicago, Chicago, IL, USA

Xue-Lan Qiu The University of Hong Kong, Pokfulam, Hong Kong

Alexander Robitzsch Department of Educational Measurement, IPN Leibniz Institute for Science and Mathematics Education, Kiel, Germany

Centre for International Student Assessment, Munich, Germany

Gina Rossi Research Group Personality and Psychopathology, Vrije Universiteit Brussel, Brussels, Belgium

Benjamin Shear Research and Evaluation Methodology, University of Colorado, Boulder, CO, USA

Hyo Jeong Shin Educational Testing Service, Princeton, NJ, USA

Klaas Sijtsma Department of Methodology and Statistics, TSB, Tilburg University, Tilburg, The Netherlands

Sandip Sinharay Educational Testing Service, Princeton, NJ, USA

William Stout Department of Statistics, University of Illinois at Urbana-Champaign, Urbana, IL, USA

Jonathan L. Templin Educational Measurement and Statistics Program, University of Iowa, Iowa City, IA, USA

L. Andries van der Ark Research Institute of Child Development and Education, University of Amsterdam, Amsterdam, The Netherlands

Alina A. von Davier ACTNext ACT Inc., Iowa City, IA, USA

Matthias von Davier National Board of Medical Examiners (NBME), Philadelphia, PA, USA

Shiyu Wang Department of Educational Psychology, University of Georgia, Athens, GA, USA

Wen-Chung Wang The University of Hong Kong, Pokfulam, Hong Kong

Gongjun Xu Department of Statistics, University of Michigan, Ann Arbor, MI, USA

Xueli Xu Educational Testing Service, Princeton, NJ, USA

Xiaofeng Yu Department of Psychology, University of Notre Dame, Notre Dame, IN, USA

Jiangxi Normal University, Nanchang, China

Michael Yudelson ACTNext ACT Inc., Iowa City, IA, USA

Juan-Diego Zapata-Rivera Educational Testing Service, Princeton, NJ, USA

Peida Zhan Department of Psychology, College of Teacher Education, Zhejiang Normal University, Zhejiang, China

Susu Zhang Department of Statistics, Columbia University, New York, NY, USA

Chapter 1
Introduction: From Latent Classes to Cognitive Diagnostic Models

Matthias von Davier and Young-Sun Lee

Abstract This chapter provides historical and structural context for models and approaches presented in this volume, by presenting an overview of important predecessors of diagnostic classification models which we will refer to as DCM in this volume, or alternatively cognitive diagnostic models (CDMs). The chapter covers general notation and concepts central to latent class analysis, followed by an introduction of mastery models, ranging from deterministic to probabilistic forms. The ensuing sections cover knowledge state and rule space approaches, which can be viewed as deterministic skill-profile models. The chapter closes with a section on the multiple classification latent class model and the deterministic input noisy and (DINA) model.

1.1 Introduction

This chapter provides historical and structural context for models and approaches presented in this volume, by presenting an overview of important predecessors of diagnostic classification models which we will refer to as DCM in this volume, or alternatively cognitive diagnostic models (CDMs). We are attempting to organize the growing field somewhat systematically to help clarify the development and relationships between models. However, given the fact that DCMs have been developed based on at least two, if not three traditions, not all readers may necessarily agree with the order in which we put the early developments. While there is a multitude of approaches that can be considered predecessors of current

M. von Davier (✉)
National Board of Medical Examiners (NBME), Philadelphia, PA, USA
e-mail: mvondavier@nbme.org

Y.-S. Lee
Teachers College, Columbia University, New York, NY, USA
e-mail: sly2003@columbia.edu

M. von Davier, Y.-S. Lee (eds.), *Handbook of Diagnostic Classification Models*, Methodology of Educational Measurement and Assessment,
https://doi.org/10.1007/978-3-030-05584-4_1

approaches to cognitive diagnostic modeling, there are many connections between these seemingly different approaches, while several different lines of development may be later understood as variants of one common more general approach (e.g., von Davier, 2013, 2014). In essence, any attempt to organize the many different approaches that exist today may lead to simplifications, and potentially omissions of related approaches.

The aim of all these approaches, however, can be summarized as the attempt to provide powerful tools to help researchers learn about how observed behaviors, such as responding to test items, can be used to derive information about generalizable behavioral tendencies.

We begin the chapter with a section on general notation and concepts central to latent class analysis, followed by an introduction of mastery models, ranging from deterministic to probabilistic forms. The ensuing sections cover knowledge state and rule space approaches, which can be viewed as deterministic skill-profile models. The chapter closes with a section on the multiple classification latent class model and the deterministic input noisy and (DINA) model.

1.2 Notation, Log-linear Models, and Latent Class Analysis

This section introduces notation used in subsequent chapters. We use the case of binary observed variables as a standard example but note that all definitions can be directly extended to polytomous nominal or ordinal response variables. Let $\boldsymbol{X} = (X_1, \ldots, X_K)$ denote K binary (or polytomous) response variables and let $\boldsymbol{x}_n = (x_{n1}, \ldots, x_{nK})$ denote the observed responses for test takers $n = 1, \ldots, N$. Let G denote a grouping variable with $g_n \in \{1, \ldots, M\}$ for all test takers. In the case of discrete mixture (or latent class) models, g_n is unobserved, while for multiple group models, g_n is completely or partially observed (von Davier & Yamamoto, 2004; von Davier & Carstensen, 2007).

The probability of observing $\boldsymbol{x} = (x_1, \ldots, x_K)$ will be denoted by

$$P(\boldsymbol{X} = \boldsymbol{x}) = P(x_1, \ldots, x_K).$$

Obviously, these probabilities are unknown, while we may have some idea which observed variables have higher or lower probability of exhibiting certain values. For cognitive tasks, we may have some idea about the order of items with respect to the likelihood of successful completion, but typically, there is no complete knowledge about the joint distribution of response variables.

The aim of modeling response data is to provide a hypothesis of how this unknown joint distribution can be constructed in a systematic way, either through associations and interactions between observables, or by means of predictors, or through assumed unobserved variables, or a combination of these.

1.2.1 Log-linear Models

One customary way to model the joint distribution of the responses $x_1, \ldots, x_K$ is using log-linear models (e.g., Haberman, 1979; Hagenaars, 1993). Log-linear models can be used with or without assuming latent variables. Log-linear models describe transformed probabilities, using the natural logarithm. We can write

$$lnP(x_1, \ldots, x_K) = f(x_1, \ldots, x_K),$$

where $f(x_1, \ldots, x_K)$ is a function of the observed responses. One possible assumption is that the log of the response probabilities can be expressed as

$$f(x_1, \ldots, x_K) = \lambda_0 + \sum_{i=1}^{K} \lambda_{1i}x_i + \sum_{\{i \neq j\}} \lambda_{2ij}x_ix_j + \ldots + \sum_{\{i \neq \ldots \neq k\}} \lambda_{Ki\ldots k}\left[\prod_{v=1}^{K} x_v\right].$$

Log-linear models in the context of CDMs have been discussed for example by von Davier and Yamamoto (2004) and Xu and von Davier (2008) for dichotomous and ordinal skill attribute variables. von Davier (2018) showed how certain log-linear models used in the context of CDMs can be considered generalizations of models recently discussed under the term *network psychometrics* (e.g., Marsman et al., 2018; von Davier, 2018). In the example above, all products of any possible subset of observed variables are included, however, it is customary to also consider models that only include terms up to a certain degree D, assuming for higher degrees $E > D$ that $\lambda_{Ei\ldots,i_E} = 0$.

One central issue when estimating log-linear models for large numbers of observables is that a normalization factor is needed. Since, $1 = \sum_{(x_1,\ldots,x_K)} P(x_1, \ldots, x_K) = \sum_{(x_1,\ldots,x_K)} \exp f(x_1, \ldots, x_K)$, it follows that

$$\lambda_0 = log\left[\sum_{(x_1,\ldots,x_K)} \exp\left[\sum_{i=1}^{K} \lambda_{1i}x_i + \sum_{\{i \neq j\}} \lambda_{2ij}x_ix_j + \ldots + \sum_{\{i \neq \ldots \neq k\}} \lambda_{Ki\ldots k}\left[\prod_{v=1}^{K} x_v\right]\right]\right].$$

This normalization factor involves a sum over all possible configurations $(x_1, \ldots, x_K)$. For K binary variables, this is a sum involving 2^K terms, for $K = 30$ items this is a summation over 1,073,741,824 terms. von Davier (2018) describes how noise contrastive estimation (e.g., Guttmann & Hyvärinen, 2010, 2012) can be used for estimation of log-linear and network psychometrics models, as well as extended log-linear models for polytomous and dichotomous CDMs.

Log-linear models can be extended by assuming latent variables (Haberman, 1979; Hagenaars, 1993) so that the distribution of observed response variables depends on an unobserved variable α,

$$f(x_1, \ldots, x_K|\alpha) = \lambda_0(\alpha) + \sum_{i=1}^{K} \lambda_{1i}(\alpha)x_i + \ldots + \sum_{\{i \neq \ldots \neq k\}} \lambda_{Ki\ldots k}(\alpha)\left[\prod_{v=1}^{K} x_v\right]$$

and by definition

$$P\left(x_1,\ldots,x_K|\alpha\right)=\exp\left[\lambda_0\left(\alpha\right)+\sum_{i=1}^{K}\lambda_{1i}\left(\alpha\right)x_i+\ldots+\sum_{\{i\neq\ldots\neq k\}}\lambda_{Ki\ldots k}\left(\alpha\right)\left[\prod_{v=1}^{K}x_v\right]\right].$$

If the latent variable is discrete, it can be written as $\alpha \in \{g_1, \ \ldots \ , g_G\}$, with G sets of each parameter type $\lambda_{dgi_1\ldots,i_d} = \lambda_{di_1\ldots,i_d}(g)$ for $g = g_1, \ \ldots \ , g_G$ *and* $d = 0, \ \ldots \ , K$. With this modification, the model becomes more complex. All parameters may depend on some unobserved quantity α, some grouping variable g, or some combination of both.

1.2.2 Latent Class Analysis

Latent Class Analysis (LCA) can be understood as an approach that assumes the dependence of response probabilities on an unobserved discrete variable, which we denote by c. In this sense, LCA is a direct application of the definition of conditional response probabilities, as introduced above. LCA assumes a latent categorical variable that cannot be directly observed. The LCA model equation follows from a set of three assumptions, some of which match assumptions commonly used in other latent variable models:

1. Class dependent response probabilities: For response variables x_i, LCA assumes class specific response probabilities. While there is no direct constraint that imposes

$$P\left(x_i|c_v\right) \neq P\left(x_i|c_w\right),$$

it is a prerequisite for class separation so that respondents who are members of different classes $c_v \neq c_w$ can be reliably classified given their observed responses.

2. Local independence: A central assumption is local independence given class membership c,

$$P\left(x_1,\ldots,x_K|c\right) = \prod\nolimits_{i=1}^{K} P\left(x_i|c\right).$$

In LCA the class membership variable c is the latent variable that is expected to 'explain' the dependencies between observed responses. Once conditional probabilities are considered, the dependencies between observed variables vanish, under this assumption.

3. Classes are mutually exclusive and exhaustive: For each examinee v there is one, and only one, 'true' latent class membership $c_v \in \{1, \ldots, G\}$. While the latent variable in LCA is nominal, this assumption is analogous to the assumption of a true (but unobserved expected) score in classical test theory (CTT) or a true ability θ in item response theory (IRT).

These three assumptions make the LCA a discrete mixture distribution model, since it follows from this set of assumptions that the marginal probability of a response pattern is given by

$$P(x_1, \ldots, x_K) = \sum_{c=1}^{G} \pi_c P(x_1, \ldots, x_K | c) = \sum_{c=1}^{G} \pi_c \prod_{i=1}^{K} P(x_i | c)$$

with mixing proportions (class sizes) $\pi_c = P(C = c)$. A logarithmic transform following assumption 2 above yields,

$$\ln P(x_1, \ldots, x_K | c) = \sum_{i=1}^{K} \ln P(x_i | c) = \sum_{i=1}^{K} [x_i \ln P(X_i = 1 | c) + (1 - x_i) \ln P(X_i = 0 | c)]$$

and further, using standard rules for the logarithm,

$$\ln P(x_1, \ldots, x_K | c) = \sum_{i=1}^{K} \ln P(X_i = 0 | c) + \sum_{i=1}^{K} x_i \left[\ln \frac{P(X_i = 1 | c)}{P(X_i = 0 | c)} \right].$$

As such, LCA can be understood as a log-linear model without interactions (as local independence is assumed), conditional on a nominal latent variable. This can be seen by setting $\lambda_{1ci} = \left[\ln \frac{P(X_i=1|c)}{P(X_i=0|c)} \right]$ (a term that represents the *log-odds* for item i conditional on class membership c) and $\lambda_{0c} = \sum_{i=1}^{K} \ln P(X_i = 0 | c)$ (an intercept term) and observing that

$$\ln P(x_1, \ldots, x_K | c) = \lambda_{0c} + \sum_{i=1}^{K} x_i \lambda_{1ci}.$$

Note that the log-odds λ_{1ci} and the conditional response probabilities have the following relationship:

$$\frac{\exp(\lambda_{1ci})}{1 + \exp(\lambda_{1ci})} = P(X_i = 1 | c).$$

While the within-class model of LCA is rather restrictive, as independence of all responses is assumed, the LCA is a very flexible model, since the number of classes C is not specified a priori. Any dependence between observed variables can be modeled by increasing the number of classes, however, identifiability may be an issue (e.g., Goodman, 1974; Allman, Matias, & Rhodes, 2009; Xu, this volume). Therefore, this flexibility is also a weakness of the LCA. With the addition of classes to the model the fit between model predictions and observed data will always improve, which may result in a LCA solution that overfits the observed dependencies. In addition, the increase in number of classes leads to a substantial increase in the number of parameters to be estimated. For additional details on applications of LCA, see the volumes by Langeheine and Rost (1988), Rost and Langeheine (1997), and Hagenaars and McCutcheon (2002), as well as the chapter by Dayton and Macready (2006).

Confirmatory approaches to LCA constrain the number of classes and often also impose inequality or equality constraints on class specific response probabilities (e.g., Croon, 1990). Most DCMs covered in this volume can be written as constrained variants of LCA (von Davier, 2009). Some constrained versions of LCA share many interesting similarities with (M-)IRT models (e.g., Haberman, von Davier, & Lee, 2008) and can be used to replace these models.

1.3 Mastery Models

Mastery models assume a skill domain for which we can sort any person into one of two classes: expert versus novice, master versus non-master, or professional versus amateur. This may not be adequate for most domains, even if there is a distinct 'can do' versus 'cannot do'; there are often gradual differences in the 'can do'. In this section, however, we use this notion of mastery and assume all respondents can be classified into two groups without further distinction.

While these types of distinctions may be oversimplifications, can they still be useful categories to describe how test takers respond to a test? If we consider ideas from developmental psychology (e.g., Piaget, 1950; Wilson, 1989), we find that some things in life are thought of as being acquired or learned in terms of qualitative jumps. We may want to entertain the idea of mastery learning for a while and examine where this leads us in terms of how a latent variable model may represent this concept. For example, young children cannot perform or solve task X until they mature and 'get it', after which the same task becomes quite easy for them.

The mastery-state can be represented by a random variable that takes on two values: '1' = mastery and '0' = non-mastery. Formally, we define a latent variable A, with $a_v \in \{0, 1\}$ for all respondents $v = 1, \ldots, n$, and with

$$a_v = 1 \text{ if person } v \text{ masters the skill of interest}$$

and

$$a_v = 0 \text{ if person } v \text{ does not master the skill.}$$

The two mastery levels are expected to differ with respect to the probabilities of success, just as in assumption 1 presented in the section on LCA above. However, in mastery models, there is an order expectation, or even an order restriction in place: it is expected (and potentially specified directly in the model) that for all response variables the probability of success is larger for masters than for non-masters. More formally,

$$P\left(X_i = 1|a = 1\right) = 1 - s_i \geq g_i = P\left(X_i = 1|a = 0\right)$$

may be assumed for all response variables $X_1, \ldots, X_K$. For each item, there are four probabilities to consider, the conditional probabilities of success and failure under mastery and non-mastery. These are often denoted as follows (e.g., Dayton & Macready, 1977):

- Guessing correctly by non-masters: $g_i = P(X_i = 1|\, a = 0)$
- Incorrect response by non-masters: $1 - g_i = P(X_i = 0|\, a = 0)$
- Slipping = unexpected incorrect response by masters: $s_i = P(X_i = 0 \,|\, a = 1)$
- Correct response by masters: $1 - s_i = P(X_i = 1|\, a = 1)$

A variety of constraints on these parameters have been suggested in the literature, some examples are discussed by Macready and Dayton (1977). Nowadays, the term ‘slipping’ is often used instead of ‘unexpected error’ while ‘guessing’ is still in use (Junker & Sijtsma, 2001). Just like LCA, mastery models also assume local independence and that masters and non-masters are mutually exclusive and exhaustive. Based on the equivalency shown in the previous section, a mastery model with two levels can be written either in the form of a 2-class LCA or as a log-linear model with latent variables:

$$P\left(x_1, \ldots, x_K|a\right) = \prod_{i=1}^{K} P\left(X_i = 0|a\right)\left[\frac{P\left(X_i = 1|a\right)}{P\left(X_i = 0|a\right)}\right]^{x_i}$$

and with the definitions above, we have $P\left(X_i=1|a\right) = (1-s_i)^a g_i^{[1-a]}$, and for the complement we have $P\left(X_i = 0|a\right) = s_i^a (1 - g_i)^{[1-a]}$. A logarithmic transformation and insertion of the definitions yields the following:

$$\ln P\left(x_1, \ldots, x_K|a\right) = \sum_{i=1}^{K} \ln s_i^a (1-g_i)^{[1-a]} + \sum_{i=1}^{K} x_i \ln\left[\frac{(1-s_i)}{s_i}\right]^a \left[\frac{g_i}{(1-g_i)}\right]^{[1-a]}.$$

As before, by setting $\sum_{i=1}^{K} \ln s_i^a (1 - g_i)^{[1-a]} = \lambda_{0a}$ and $\ln \left[\frac{(1-s_i)}{s_i}\right]^a \left[\frac{g_i}{(1-g_i)}\right]^{[1-a]} = \lambda_{1ai}$, the equivalency of the mastery model to a log-linear model with a binary latent variable is obtained. Note that λ_{1ai} can be written as

$$a \ln \left[\frac{P\left(X_i = 1|a = 1\right)}{P\left(X_i = 0|a = 1\right)}\right] + [1-a] \ln \left[\frac{P\left(X_i = 1|a = 0\right)}{P\left(X_i = 0|a = 0\right)}\right] = \lambda_{10i} + a\left[\lambda_{11i} - \lambda_{10i}\right],$$

which again contains the log-odds for masters and non-masters, multiplied by the mastery status.

1.4 Located Latent Class or Multi State Mastery Models

The additional model specifications needed to move from LCA, which is characterized by a nominal latent class variable, to located classes are easily introduced. The last section that examined mastery models provides the basis for these developments. For a correct response $x_i = 1$, the term $\lambda_{1ai} = \lambda_{10i} + a[\lambda_{11i} - \lambda_{10i}]$ is part of the sum. This term is linear in the mastery level $a \in \{0, 1\}$ and if $\lambda_{11i} > \lambda_{10i}$ or equivalently, $P(X_i = 1|\, a = 1) > P(X_i = 1|\, a = 0)$, the term λ_{1ai} is monotone increasing over the (in the case of mastery models: two) ordered mastery levels.

With more than two levels of mastery, for example an ordinal variable that represents non-mastery as zero, but allows multiple levels of mastery represented as successive integer, i.e., $a^{'} \in \{0, 1, 2, \ldots M\}$, a model can be defined as

$$\ln P\left(x_1, \ldots, x_K | a'\right) = \lambda_{0a'} + \sum_{i=1}^{K} x_i \lambda_{1a'i}$$

with

$$\lambda_{1a'i} < \lambda_{1a''i} \text{ for all } a' < a''.$$

This ensures that

$$P\left(X_i = 1|a'\right) < P\left(X_i = 1|a''\right) \text{ for all } a' < a''.$$

This produces a monotone increasing sequence of response probabilities over $a^{'} \in \{0, 1, 2, \ldots M\}$. Note, however, that this type of constraints (still) produces a comparably large number of quantities that need to be estimated. However, this model includes equality constraints (e.g., Formann, 1985, 1992) which may be imposed via additional assumptions about how model parameters relate to the ordered levels of mastery. Essentially, each latent class in this model becomes an ordered mastery level, but the distances between classes differ by item i and class level $a' \in \{0, 1, 2, \ldots M\}$. This model requires $(M + 1)K$ parameters one set of K item parameters for each class. As before, probabilities can be derived using the equivalency

$$P\left(X_i = 1|a'\right) = \frac{\exp\left(\lambda_{1a'i}\right)}{1 + \exp\left(\lambda_{1a'i}\right)}.$$

A more parsimonious model can be implemented by imposing the following constraint

$$\lambda_{1a'i} = \beta_i + \gamma_i \theta_{a'}$$

which requires $2K$ item location β_i and slope parameters γ_i and $M + 1$ ordered class specific locations $\theta_{a'} < \theta_{a''}$ for $a' < a'' \in \{0, \ldots, M\}$. With the transformation

$$\beta_i + \gamma_i \theta_{a'} = a\left(\theta - b\right)$$

it can be easily observed that located latent class models define the class specific response probabilities as

$$P\left(X_i = 1|\theta_{a'} = \theta\right) = \frac{\exp\left(a\left(\theta - b\right)\right)}{1 + \exp\left(a\left(\theta - b\right)\right)}$$

which is very similar to IRT (Lord & Novick, 1968), while assuming a discrete latent variable with located latent classes (e.g., Formann, 1992; Haberman et al., 2008).

1.5 Rule Space Methodology and Knowledge Spaces

Rule space (RS; e.g., Tatsuoka, 1983, 1990, 2009) and knowledge spaces (KS; Doignon & Falmagne, 1985, 1998; Albert & Lukas 1999) are independently developed approaches to the question of how the association between performance on heterogeneous tasks and multiple skills can be conceptualized. Much like mastery models, RS and KS assume that a respondent who masters a certain number of skills is on a regular basis capable of solving tasks that require these skills. In contrast to the first generation of mastery models, both RS and KS assume that there are multiple skills to be considered, and that each respondent is characterized by a skill pattern or attribute pattern – or a *knowledge state* – and that every task requires a subset of the skills represented in the skill space of respondents.

Consider an example with two skills, addition and multiplication, ignoring for a moment that there is an additional skill required that tells us in what order these operations have to be executed. If asking examinees to solve tasks of the type

(a) $3 + 4 =?$
(b) $4 * 5 =?$
(c) $3 * 3 + 2 =?$

one could argue that there are four potential groups of test takers. Group 1 does neither master addition nor multiplication and cannot solve any of the task types; Group 2 masters only addition and can solve tasks of type (a) only; Group 3 only masters multiplication (no matter how unlikely that may seem to a math educator) and hence can solve only tasks of type (b); and Group 4 masters both addition and multiplication, and hence can solve tasks of type (a), (b), and (c) on a regular basis.

More formally, for tasks that require a subset of D skills, we can assign to each task $i = 1, \ldots, K$ a vector of skill requirements $\boldsymbol{q}_i = (q_{i1}, \ldots, q_{iD}) \in \{0, 1\}^D$ that indicates which skill (or attribute) is required for that task. The matrix

$$Q = \begin{pmatrix} q_{11} & \cdots & q_{1D} \\ \cdots & \cdots & \cdots \\ q_{K1} & \cdots & q_{KD} \end{pmatrix}$$

is referred to as Q-matrix and represents a hypothesized relationship of how a skill vector (skill state) or attribute pattern $\boldsymbol{a} = (a_1, \ldots, a_D)$ is connected to expected performance on each task. The ideal (the most likely, or expected given a skill pattern) response on item i given can be written as

$$x_i^{[I]}\left(\boldsymbol{q}_i, \boldsymbol{a}\right) = \prod_{d=1}^{D} a_d^{q_{id}} \in \{0, 1\}$$

which equals one if the attribute mastery pattern $\boldsymbol{a}$ matches or exceeds non-zero entries of the skill requirements $\boldsymbol{q}_i$, i.e., if at least all required skills are mastered, and is zero otherwise. The above equation can be applied to all items to construct an ideal response pattern

$$\boldsymbol{x}^{[I]}(\boldsymbol{a}) = \left(\prod_{d=1}^{D} a_d^{q_{1d}}, \ldots, \prod_{d=1}^{D} a_d^{q_{Kd}}\right)$$

for each attribute mastery pattern $\boldsymbol{a}$**.** The observed response pattern $\boldsymbol{x}_v$ produced by respondent v can then be compared to each of these ideal response vectors, and the closest match determined. This can be done in a variety of ways; for example, Tatsuoka (1983, 1985) discussed methods based on distance measures, but also presents classification based on IRT ability estimates and person fit. von Davier, DiBello, and Yamamoto (2008) provide a summary of the IRT and fit based approach. A simple measure of agreement can be defined as

$$sim\left(\boldsymbol{x}_v, \boldsymbol{a}\right) = \frac{\sum_{i=1}^{K} x_{vi} * x_i^{[I]}\left(\boldsymbol{q}_i, \boldsymbol{a}\right)}{\sqrt{\left(\sum_{i=1}^{K} x_{vi}^2\right)\left(\sum_{i=1}^{K}\left[x_i^{[I]}\left(\boldsymbol{q}_i, \boldsymbol{a}\right)\right]^2\right)}}$$

which equals the cosine similarity of the observed and ideal vectors. The cosine similarity is a correlation related measure commonly used in data mining, machine learning and natural language processing (Tan, Steinbach, & Kumar, 2005). Respondents can be assigned to the attribute pattern that produces the largest similarity measure relative to the observed vector $\boldsymbol{x}_v$.

Tatsuoka's RS has demonstrated its utility in many applications over the years. Recently, the method gained new interest under the name 'attribute hierarchy method' (AHM; Leighton, Gierl, & Hunka, 2004). The authors describe the AHM as being an instantiation of rule space that differs from Tatsuoka's (1983, 1985, 1990, 2009) methodology in that it allows attribute hierarchies. Attribute hierarchies limit the permissible attribute space, as some attributes have to be mastered before other can be mastered, by definition of what a hierarchy encompasses. von Davier and Haberman (2014) show how the assumption of hierarchical attributes restricts the number and type of parameters of diagnostic classification and multiple mastery models.

Both RS and KS were initially conceptualized as deterministic classification approaches. Respondents would be classified according to their similarity to ideal response patterns, regardless of the observation that only very few respondents will produce exactly the 'ideal' patterns that can be expected based on the Q-matrix. Attempts to produce a less deterministic version of these approaches have been made, and Schrepp (2005) describes similarities between KS approaches and latent class analysis. The next section describes models that share many of the features of RS and KS approaches, but provide a structured latent attribute space, and a probabilistic approach to define how multiple mastery levels relate to response probabilities in a systematic way, rather than by means of unstructured class profiles as used in LCA.

1.5.1 Multiple Classification Models and Deterministic Input Noisy and (DINA) Models

Latent class models with multiple latent variables (Haberman, 1979; Haertel, 1989) or multiple classification latent class models (MCLCM; Maris, 1999) extend latent class analysis (LCA) in such a way that multiple nominal or ordinal latent variables can be identified simultaneously. This approach retains the defining properties of LCA, local independence given latent class, assumption of an exhaustive and disjunctive latent classification variable, and distinctness of conditional probabilities across classes.

The MCLCM approach can be viewed as a non-parametric precursor to many of the diagnostic models introduced in subsequent chapters. For a MCLCM with two latent variables $c_1 \in \{0, \ldots, C_1\}$, $c_2 \in \{0, \ldots, C_2\}$ denote the joint distribution of these with π_{c_1,c_2} and define

$$P\left(x_1, \ldots, x_K\right) = \sum_{c_1=0}^{C_1} \sum_{c_2=0}^{C_2} \pi_{c_1,c_2} \prod_{i=1}^{K} P\left(x_i | c_1, c_2\right).$$

This is a well-defined LCA that can be rewritten as a single latent[1] variable LCA with 'attribute' $\boldsymbol{a} = \{c_1, c_2\}$ and MNCL $= (C_1 + 1)(C_2 + 1)$ latent classes representing all possible combinations. However, one may introduce additional structure – constraints on the response probabilities – for the two-variable case to specify whether the conditional probabilities may for some items depend on only one or the other component c_1 or c_2. More specifically, one may assume

$$P(x_i | c_1, c_2) = P(x_i | f_{i1}(c_1), f_{i2}(c_2)).$$

As a special case with specific relevance to diagnostic models, we will consider the following form of these constraints in the example

$$f_{id}(c_d) = c_d^{q_{id}}$$

for $d = 1, 2$ and with $q_{i1}, q_{i2} \in \{0, 1\}$.

Basically, if one or the other q_{i_*} is zero, the dependency on that component of the multiple classification LCM variable vanishes from the conditional probability of item response x_i. This is true because

$$c_d^0 = 1$$

for all levels of c_d whenever $q_{i1} = 0$. With this constraint, the conditional probabilities of a response variable may depend on both c_1, c_2 in MNCL levels for some items, on c_1 only in $(C_1 + 1)$ levels for some other items, or on c_2 with $(C_2 + 1)$ levels for a third set of items, or on neither one of them in a fourth group of response variables.

Two additional restrictions lead to the model that is commonly known as the DINA (Deterministic Input, Noisy And) model (Macready & Dayton, 1977; Junker & Sijtsma, 2001). First, all components of the latent skill pattern $\boldsymbol{a}$ are assumed to be binary (and as before, we use a_d for binary attributes, while for nominal classes, we use $c_1, c_2, \ldots$), that is

$$\boldsymbol{a} = (a_1, \ldots, a_D) \in \{0, 1\}^D$$

and for the conditional probabilities we assume

$$P(x_i | a_1, \ldots, a_D) = P\left(x_i | \prod_{d=1}^{D} a_d^{q_{id}}\right).$$

[1]Class variables are represented as integers, but the use of integers do not imply any ordering here; only equivalence classes are used in the context of LCA.

Note that the conditional probability depends on a binary variable $\xi_{\boldsymbol{a}q_i} = \prod_{d=1}^{D} a_d^{q_{id}} \in \{0, 1\}$ which is a function of the skill pattern $\boldsymbol{a}$ and one row of the Q-matrix, a vector that specifies the skill requirements for a specific item. Just as in the section on mastery models, applying this definition leads to the following expressions:

$$P\left(X_i = 1 | \xi_{\boldsymbol{a}q_i} = 1\right) = 1 - s_i$$

and

$$P\left(X_i = 1 | \xi_{\boldsymbol{a}q_i} = 0\right) = g_i.$$

The DINA model is said to be conjunctive because it reduces the respondent-skill by item-attribute comparison to only two levels $\prod_{d=1}^{D} a_d^{q_{id}} = 1$ or $\prod_{d=1}^{D} a_d^{q_{id}} = 0$. With this, we can write

$$P\left(x_i | a_1, \ldots, a_D\right) = \left[(1 - s_i)^{\xi_{\boldsymbol{a}q_i}} {g_i}^{1-\xi_{\boldsymbol{a}q_i}}\right]^{x_i} \left[{s_i}^{\xi_{\boldsymbol{a}q_i}} (1 - g_i)^{1-\xi_{\boldsymbol{a}q_i}}\right]^{1-x_i}.$$

Only those respondents who possess all necessary skills have a "high" probability $1 - s_i$ of solving an item, while respondents who lack at least one of the required skills have a "low" probability g_i —the same "low" probability no matter whether only one or all required skills are not mastered.

Note that the g_i and the s_i denote the item parameters in the DINA model, so that there are two parameters per item in this model. In addition, the skill vectors $a_v = (a_{v1}, \ldots a_{vK})$ are unobserved, so we typically have to assume that the distribution of skills $P\left(A = (a_1, \ldots a_K)\right) = \pi_{(a_1,\ldots a_K)}$ is unknown. Therefore, there are $\|\{0, 1\}^K\| - 1 = 2^K - 1$ independent skill pattern probabilities with $\sum_{(a_1,\ldots a_K)} \pi_{(a_1,\ldots a_K)} = 1.0$ if an unconstrained estimate of the skill distribution is attempted. There may be fewer parameters if a constrained distribution over the skill space (von Davier & Yamamoto, 2004; Xu & von Davier, 2008) is used. For model identification, no constraints are needed on the guessing and slipping parameters (even though it is desirable that $1 - s_i > g_i$ for somewhat sensible results).

While de la Torre (2009) does not make statements about identifiability of the DINA model and the uniqueness of the model parameters, Junker and Sijtsma (2001) discuss (a lack of) empirical identification in the context of their data example used in conjunction with Markov chain Monte Carlo (MCMC) estimation. Haertel (1989) describes identification of latent class skill patterns in the DINA model, and notes that *"it may be impossible to distinguish all these classes empirically using a given set of items. Depending upon the items' skill requirements, latent response patterns for two or more classes may be identical (p.303)."* One of the remedies Haertel (1989) suggests is the combination of two or more latent classes that cannot be distinguished. In subsequent chapters, identifiability of diagnostic models is discussed in more detail (Xu, this volume; Liu & Kang, this volume; DeCarlo, this volume) and von Davier (2014) provides an example of how the (lack of) empirical identifiability of diagnostic models can be checked.

The DINA model is a very restrictive model as it assumes only two parameters per item, and skill attributes only enter the item functions through conjunction function $\xi_{\boldsymbol{a}q_i} = \prod_{d=1}^{D} a_d^{q_{id}}$. This restricts the probability space so that different attribute mastery patterns, in particular those that are not a perfect match of the Q-matrix for an item, are all mapped onto the same low "guessing" probability. There are several issues with the assumption made in the DINA model. Formally, this assumption is equivalent to assuming a log-linear model (see Eq. 1) in which all parameters are set to zero except the one that parameterizes the highest order interaction term. Additionally, from the point of view of most applications of skills, compensation happens: Multiplication can be replaced by repeated addition, a lack of vocabulary when acquiring a new language, or even learning disabilities can be compensated for (and eventually remedied) by higher general intelligence (e.g., Reis, McGuire, & Neu, 2000), etc. In total darkness, hearing can be used to, admittedly poorly, compensate for lack of vision. For diagnostic models and compensatory and non-compensatory MIRT models, it was found that real data examples are often fit better (in terms of item fit, or overall goodness of fit assessed with information criteria or similar) with additive/compensatory models rather than conjunctive models (de la Torre & Minchen, this volume; von Davier, 2013).

In addition, it was found that the DINA model may be affected by model identification issues. DeCarlo (2011) and Fang, Liu, and Ying (2017) show that the DINA model is not identified unless there are what some may call 'pure' items in the Q-matrix, that is, items that only measure a single attribute. DeCarlo (2011) shows that the DINA model with the Q-matrix provided for the Fraction Subtraction data (Tatsuoka, 1985) is not able to identify all attribute patterns. Fang, Liu, and Ying (2017) provide more general results on the requirements for the Q-matrix. Xu (this volume) and Liu and Kang (this volume) provide further results and more recent examples.

1.6 Summary

The notation and models introduced in this chapter form the basis for many of the subsequent chapters. Most, if not all DCMs can be written as constrained latent class models or alternatively, log-linear models with discrete latent variables.

This introduction does not provide an in-depth coverage of how to evaluate the different approaches. However, all models presented in this volume are approaches that provide marginal probability distributions for multivariate discrete observables. This means that methods from categorical data analysis can be used to compare models and to evaluate model data fit.

While some of the models introduced above may be considered approaches for diagnostic classification and may have been used as such, many more sophisticated approaches have been developed since, based on these initial modeling attempts. The aim of the current volume is providing a systematic overview of these more recent approaches.

The *Handbook Diagnostic Classification Models* aims at capturing the current state of research and applications in this domain. While a complete overview of this broad area of research would require a multi-volume effort, we tried to capture a collection of major research streams that have been developed over several years and that continue to produce new results.

The first part of the volume covers major developments of diagnostic models in the form of chapters that introduce the models formally, provide information on parameter estimation and on how to test model-data fit, and applications or extensions of the approach.

The second part of the volume describes special topics and applications. Special topics such as Q-matrix issues are covered, including the data driven improvement and construction, as well as issues around model identifiability. The third part presents applications of diagnostic models, as these are a centerpiece to reasons why not only methodologists but also applied researchers may want to study the volume. These applications show how diagnostic models can be used to derive more fine-grained information about respondents than what traditional methods such as CTT or IRT can provide.

The fourth part of the book includes a range of available software packages, including the use of general purpose statistical software, specialized add-on packages, and available stand-alone software for estimation and testing of CDMs.

In many cases, latent class analysis, customary IRT, and other latent variable models can directly be considered alternatives to diagnostic models, as these are often more parsimonious (in the case of IRT) or do not make as strong (parametric) assumptions about the latent structures and how these structures are related to the conditional response probabilities in the levels of the latent variables. Standard procedure should therefore be used as a comparison of more complex modeling approaches with customary standard examples of latent variable models such as IRT or LCA. Such a practice will ensure that researchers can compare their findings to those obtained from less complex models to check whether the increased model complexity provides added value, through improved model-data fit, and by means of more useful derived quantities such as estimated mastery states.

References

Albert, D., & Lukas, J. (1999). *Knowledge spaces: Theories, empirical research and applications*. Mahwah, NJ: Erlbaum.

Allman, E. S., Matias, C., & Rhodes, J. A. (2009). Identifiability of parameters in latent structure models with many observed variables. *The Annals of Statistics, 37*(6A), 3099–3132. https://doi.org/10.1214/09-AOS689

Croon, M. A. (1990). Latent class analysis with ordered classes. *British Journal of Mathematical and Statistical Psychology, 43*, 171–192.

Dayton, C. M., & Macready, G. (2006). Latent class analysis in psychometrics. In C. R. Rao & S. Sinharay (Eds.), *Handbook of statistics, Vol. 26: Psychometrics* (pp. 45–79). Elsevier Science B.V.: Amsterdam, The Netherlands.

de la Torre, J. (2009, March). DINA Model and parameter estimation: A didactic. *Journal of Educational and Behavioral Statistics, 34*(1), 115–130.

Decarlo, L. T. (2011). On the analysis of fraction subtraction data: The DINA model, classification, latent class sizes, and the Q-matrix. *Applied Psychological Measurement, 35*, 8–26.

Doignon, J. P., & Falmagne, J. C. (1985). Spaces for the assessment of knowledge. *International Journal of Man-Machine Studies, 23*, 175–196.

Doignon, J. P., & Falmagne, J. C. (1998). *Knowledge spaces*. Berlin, Germany: Springer.

Fang, G, Liu, J., & Ying, Z. (2017). *On the identifiability of diagnostic classification models* (Submitted on 5 Jun 2017) arXiv:1706.01240 [math.ST].

Formann, A. K. (1985). Constrained latent class models: Theory and applications. *British Journal of Mathematical and Statistical Psychology, 38*, 87–111.

Formann, A. K. (1992). Linear logistic latent class analysis for polytomous data. *Journal of the American Statistical Association, 87*(418), 476–486. https://doi.org/10.1080/01621459.1992.10475229

Goodman, L. A. (1974). Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika, 61*, 215–231.

Gutmann, M. U., & Hyvärinen, A. (2010). Noise-contrastive estimation: A new estimation principle for unnormalized statistical models. In *International Conference on Artificial Intelligence and Statistics*, 2010.

Gutmann, M. U., & Hyvärinen, A. (2012). Noise-contrastive estimation of unnormalized statistical models, with applications to natural image statistics. *Journal of Machine Learning Research, 13*, 307–361.

Haberman, S. J. (1979). *Qualitative data analysis: Vol. 2. New developments*. New York, NY: Academic.

Haberman, S. J., von Davier, M., & Lee, Y. (2008). Comparison of multidimensional item response models: Multivariate normal ability distributions versus multivariate polytomous ability distributions. RR-08-45. ETS Research Report. https://doi.org/10.1002/j.2333-8504.2008.tb02131.x

Haertel, E. H. (1989). Using restricted latent class models to map the skill structure of achievement items. *Journal of Educational Measurement, 26*, 301–321.

Hagenaars, J. A. (1993). Loglinear models with latent variables. In *Sage University Papers* (Vol. 94). Newbury Park, CA: Sage Publications.

Hagenaars, J. A., & McCutcheon, A. L. (Eds.). (2002). *Applied latent class analysis models*. Cambridge, UK: Cambridge University Press.

Junker, B. W., & Sijtsma, K. (2001). Cognitive assessment models with few assumptions, and connections with nonparametric item response theory. *Applied Psychological Measurement, 25*, 258–272.

Langeheine, R., & Rost, J. (Eds.). (1988). *Latent trait and latent class models*. New York, NY: Plenum Press.

Lazarsfeld, P. F., & Henry, N. W. (1968). *Latent structure analysis*. Boston, MA: Houghton Mifflin Company.

Leighton, J. P., Gierl, M. J., & Hunka, S. M. (2004). The attribute hierarchy model for cognitive assessment: A variation on Tatsuoka's rule-space approach. *Journal of Educational Measurement, 41*, 205–237.

Macready, G. B., & Dayton, C. M. (1977). The use of probabilistic models in the assessment of mastery. *Journal of Educational Statistics, 2*, 99–120.

Maris, E. (1999). Estimating multiple classification latent class models. *Psychometrika, 64*, 187–212.

Marsman, M., Borsboom, D., Kruis, J., Epskamp, S., van Bork, R., Waldorp, L. J., ... Maris, G. (2018). An introduction to network psychometrics: Relating ising network models to item response theory models. *Multivariate Behavioral Research, 53*(1), 15–35. https://doi.org/10.1080/00273171.2017.1379379