

© 2010 Zhushan Li

LOGLINEAR MODELS AS ITEM RESPONSE MODELS

BY

ZHUSHAN LI

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Educational Psychology
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2010

Urbana, Illinois

Doctoral Committee:

Professor Carolyn J. Anderson, Chair
Professor Hua-hua Chang
Professor Jeffrey A. Douglas
Louis A. Roussos, PhD, Measured Progress
Associate Professor Jinming Zhang

Abstract

For analyzing item response data, item response theory (IRT) models treat the discrete responses to the items as driven by underlying continuous latent traits, and consider the form of conditional probability of the response to each item given the latent traits. In a similar fashion, log-linear models directly consider the form of the manifest probability of response patterns. Researchers have been connecting the two paradigms by establishing equivalence relationships between IRT models and log-linear models. This has lead to the notion of obtaining IRT solutions by fitting their equivalent log-linear models.

In this research, I have established a family of log-linear models, log linear-by-linear association (LLLA) models, that incorporate a variety of IRT models, particularly, a family of generalized Rasch models. I have derived an extension of the Dutch Identity theorem to polytomous items and utilized it to develop the models that incorporate item covariates and person covariates. Noteworthy features of the models include both polytomous responses and multiple latent traits.

Along with developing this new family of models, I have conducted extensive research on the development of an accompanying estimation method. Historically, a significant barrier to the application of log-linear models in analyzing item responses has been the high computational cost of maximum likelihood estimation (MLE), due to the fact that the number of response patterns grows exponentially as the number of items increases. To solve this computational problem, a pseudo-likelihood estimation (PLE) method is proposed and it dramatically decreases the computational cost.

To demonstrate the effectiveness of the developed models and the pseudolikelihood estimation method, I will present results of a series of simulation studies. To demonstrate the practical advantages of the methods, I will give a detailed description of an application to a real data set from a study on verbally aggressive behavior.

To my family

Acknowledgments

Completing a doctoral degree has been a fantastic journey for me because of the strong support I received from numerous people over the past years. I am very grateful for the support of families and friends who helped me as I worked on this dissertation. I would like to thank my husband Feng and my daughter Susan for their love. They make my life full, balanced, and happy. My love to them is beyond words I can describe. I would like to thank my parents and my parents-in-law and my siblings in China for their endless encouragement, love and support. I would like to thank the families of Robert and Susan Spellman and their daughters for their tremendous love and support. The love from all my families has been both a trusted guidance and deeply cherished treasure through the years. It is your love that means the world to me.

I am very fortunate to have such a wonderful committee, Carolyn Anderson, Hua-hua Chang, Jeff Douglas, Louis Roussos and Jinming Zhang. They are not only mentors but also friends. I would like to thank my advisor Carolyn who was willing to support me and share her personal stories and advice with me. Her trust and encouragement of my doing good research has given me a lot of confidence. I really appreciate her for encouraging me to think independently. I deeply appreciate my early research advisor and mentor Louis Roussos for his leading me into the field and providing a good foundation for my research. His kindness, love, passion, and willingness to help students deeply influenced me and set a very good model for my future career. I am very grateful to Professor Jeff Douglas for supporting me financially to go to the AERA and NCME conferences. Your wonderful teaching, rich knowledge, useful suggestions, kindness, and support for my research will be remembered forever. Professor Hua-hua Chang's comments on my solid knowledge foundation, good communications skills, and very strong research encouraged me and motivated me to do a better job and helped me to achieve better. His support for my work deeply touched me

and I deeply appreciate his encouragement. Professor Jinming Zhang is a very bright and nice person and gave me many good suggestions.

I want to thank the Department of Educational Psychology and the Department of Statistics for providing financial aid through my doctoral studies. I also want to thank various award committees for giving me awards for my research, which encouraged me to do better and achieve better. The secretaries in our department have been very supportive all these years. I want to thank all of you, especially my friends Helen and Julie who have been there for me many times.

My training in the Department of Statistics will benefit me forever. I deeply appreciate many professors there. In particular, John Marden, John Martinsek, Jeff Douglas, Xuming He, Xiaofeng Shao, Feng Liang, Barbara, Doug Simpson, and Ombao supported me in different ways.

Many friends helped and supported me during my time at University of Illinois. Peter Cao and Amy Li provided their home and food for us on many occasions. Their friendship is more like another family. Ann and Russell helped us a lot when we were in need. My friend Sherman always believed in me and trusted me in doing good research. My friends Cha-chi, Charles, Hongling, Xiangkui, Haiyan, Yiming, Rongchun, EunYoung, Holly Downs, Xiaolin, Xiaofang, Jianxin, Yinping, Min Li, Shiqiu, Shenfang, Shuqing and Peilin provided a lot of support and help. While I am thinking about all my friends, I cannot list all of them here, I feel very blessed to have had such an amazing group of friends during my years at UIUC. My friends Bill and Kathryn Scott and the home group in Rantoul supported us tremendously. I also want to thank Calvary Campus Church, Stone Creek Church and Christian Life Church and CBS friends. I also want to thank all my friends in BCBSG fellowship. Your prayers and love for me has been my strength all these years.

Finally, I would like to give thanks to God for his grace and blessings in my life. May His glory shine through my life.

Table of Contents

List of Tables	ix
List of Figures	xi
List of Abbreviations	xiii
List of Symbols	xiv
Chapter 1 Introduction	1
Item Response Theory Models	1
Log-linear Models	3
Connecting Two Paradigms	6
Research Objectives	7
Adding covariates	9
Polytomous and multidimensional models	11
Summary	12
Chapter 2 LLLA as IRT Models	14
Assumptions for LLLA model	14
LLLA Model as Rasch Model	17
Dutch Identity	20
Fitting Rasch by LLLA	24
Item parameters	24
Person parameters	24
Population latent trait distribution	27
Mixture of conditional normals: demonstrations by simulated data	29
Chapter 3 Pseudolikelihood Estimation	34
MLE for LLLA Model	35
Motivation of PLE for LLLA	37
Introduction to PLE	39
Application of PLE to LLLA	40
Maximize PL Using Logistic Regression	44
Correct Standard Errors	46
Jackknife	46
Bootstrap	47
Robust estimator	48

Chapter 4 LLLA With Covariates	50
IRT Models With Covariates	51
Item covariates	51
Person covariates	51
LLLA With Item Covariates	52
LLLA With Person Covariates	53
Model form	54
Derivation of LLLA with person covariates model	54
Connection and difference with latent regression Rasch model	56
Chapter 5 Polytomous Models	58
Review of Polytomous IRT Models	59
Definition of polytomous logit functions	60
Three classes of polytomous IRT models	64
The family of partial credit models	64
Dutch Identity for Polytomous Models	66
Derivation of the LLLA Model From Bock's Model and the PCM	69
Polytomous Models With Item Covariates	71
Partial credit model with item covariates	71
LLLA model with item covariates	72
Polytomous Models With Person Covariates	74
Partial credit model with person covariates	74
LLLA model with person covariates	75
Chapter 6 Multidimensional Models	77
Multidimensional LLLA Model	78
Dichotomous model	78
Polytomous model	82
Estimating Rasch parameters	90
Multidimensional LLLA Model With Item Covariates	93
Dichotomous model	93
Polytomous model	96
Estimating Rasch parameters	100
Multidimensional LLLA Model With Person Covariates	102
Dichotomous model	102
Polytomous model	106
Estimating Rasch parameters	110
Chapter 7 Simulation Studies	116
Unidimensional Dichotomous Models	119
LLLA with item covariates	119
LLLA with person covariates	123
Unidimensional Polytomous Models	132
Polytomous LLLA with item covariates	132
Polytomous LLLA with person covariates	135

Multidimensional Dichotomous Models	143
Multidimensional LLLA with Item covariates	143
Multidimensional LLLA with person covariates	146
Multidimensional Polytomous Models	151
Multidimensional polytomous LLLA with item covariates	151
Multidimensional polytomous LLLA with person covariates	154
Chapter 8 Real Data Analysis	159
Computation Time	161
Rasch Model	161
LLTM	164
Latent Regression Rasch Model	168
Partial Credit Model	169
Partial Credit Model With Item Covariates	172
Partial Credit Model With Person Covariates	173
Chapter 9 Conclusion	178
Flexibility of the Models	178
Flexibility on Latent Trait Distribution	179
Pseudolikelihood is Fast	180
Non-collapsibility of LLLA Models	181
Interpretation of the LLLA Parameters	181
Differential Item Functioning	182
Estimation for IRT 2PL Model and Bock's Model	182
References	184

List of Tables

1	An Example of a Data Matrix From a 4-item Binary Test	1
2	Different Formulations of the Rasch Model	19
3	Original Response Data Matrix	36
4	Response Data in Count Data Format	37
5	Design Matrix for the LLLA Model	38
6	Stacked Data in Pseudolikelihood Estimation With Logistic Regression Procedures	45
7	Simulation Designs	117
8	MLE and PLE Obtained From Fitting the LLLAi Model on Simulated Data With 5 and 10 Items	120
9	Comparing the Robust Estimates (“Sandwich” Estimator) of the SE’s to True SE’s for the LLLAi models by Monte Carlo Simulation. Each Model was Replicated 10,000 Times	123
10	Parameter Estimates of the LLLAp Model and the LLLA Model Along With the True Parameters for the 10-item Test	127
11	Comparing the Robust Estimates (“Sandwich” Estimator) of the SE’s to True SE’s for the LLLAp models by Monte Carlo Simulation. Each Model was Replicated 10,000 Times. For the Tests With Length 10, 50 and 100 Items, the Item Effects for Only the First 5 Items	131
12	MLE and PLE Obtained From Fitting the LLLAi Model on Simulated (Polytomous) Data With 5 and 10 Items	134
13	Parameter Estimates of the LLLAp Model and the LLLA Model Along With the True Parameters for the 10-item (Poly- tomous) Test	140
14	MLE and PLE Obtained From Fitting the LLLAi 2D Model on Simulated (Polytomous) Data. Test Length $I = 6$ and 10 Items	145
15	MLE and PLE Obtained From Fitting the LLLAi (2D) Model on Simulated Data. Test Length $I = 10$ Items. Polytomous items With 3 Categories	153
16	IRT Models for the Real Data Set	159
17	Time Cost of Fitting IRT Models on Verbal Aggression Data	162
18	Parameter Estimates and Standard Errors for the Rasch Model (Verbal Aggression Data)	163
19	Parameter Estimates and Standard Errors for the LLTM (Ver- bal Aggression Data)	166
20	Parameter Estimates and Standard Errors for the Latent Re- gression Rasch Model (Verbal Aggression Data)	169
21	Parameter Estimates and Standard Errors for the Partial Credit Model (Verbal Aggression Data)	171

22	Parameter Estimates and Standard Errors for the Partial Credit Model With Item Covariates (Verbal Aggression Data)	174
23	Parameter Estimates and Standard Errors for Partial Credit Model With Person Covariates (Verbal Aggression Data)	175
24	Models That can be Fit by R Package ‘plRasch’	179

List of Figures

1	Rasch model's item response function.	3
2	Contingency table for a 2-item test with dichotomous responses	4
3	Extended Rasch models.	8
4	Research strategies.	11
5	Graphical structure of the 1-D LLLA model.	15
6	Population latent trait distributions, standard normal, 10 items.	29
7	Population latent trait distributions, two-component mixture of normal, 10 items.	31
8	Population latent trait distributions, two-component mixture of normal, 50 items.	32
9	Baseline category logits.	61
10	Adjacent category logits.	62
11	Cumulative logits.	63
12	Sequential logits.	63
13	PLE vs MLE for item covariates effects. Test length $I = 5$ and 10 items.	121
14	PLE vs true item covariate effects. Test length $I = 5, 10, 50,$ and 100 items.	122
15	PLE vs true person covariate effects. Test length $I = 5, 10,$ 50, and 100 items.	126
16	Parameter estimates and SEs for λ by the LLLAp and LLLA models.	128
17	Three-way comparison of the true, LLLAp estimated, and LLLA estimated item difficulty parameters β_i	129
18	Three-way comparison of the true, LLLAp estimated, and LLLA estimated person ability parameters θ_p	130
19	PLE vs MLE for item covariates effects. Test length $I = 5$ and 10 items.	133
20	PLE vs true item covariate effects. Test length $I = 5, 10, 50,$ and 100 (polytomous) items.	136
21	PLE vs true person covariate effects. Test length $I = 5, 10,$ 50, and 100 items.	138
22	Parameter estimates and SEs for λ by the LLLAp and LLLA models.	139
23	Three-way comparison of the true, LLLAp estimated, and LLLA estimated item difficulty parameters $\beta_{i(h)}$	141
24	Three-way comparison of the true, LLLAp estimated, and LLLA estimated person ability parameters θ_p	142
25	PLE vs MLE for item covariate effects. 2-D latent traits, Test length $I = 6$ and 10 items.	146
26	PLE vs true item covariate effects. Test length $I = 6, 10, 50,$ and 100 items. 2-D latent traits and dichotomous response.	147

27	PLE vs true person covariate effects. Test length $I = 6, 10,$ 50, and 100 items; 2-D latent traits and dichotomous responses.	149
28	Fitted person covariates vs true parameters with “unbalanced” item loadings. Test length $I = 50$ and 100 items.	151
29	PLE vs MLE for item covariate effects. Test length $I = 6$ and 10 items; 2-dimensional latent traits, 3 category responses.	154
30	PLE vs true item covariate effects. Test length $I = 6, 10, 50,$ and 100 items, 2-dimensional latent traits, 3 category response.	155
31	PLE vs true person covariate effects. Test length $I = 6, 10,$ 50, and 100 items; 2 dimensional traits, 3-category responses.	158
32	Item difficulty estimates for the Rasch model, LLLA (PL) vs GLMM (NLMIXED).	164
33	Item covariate effects estimates for the LLTM, LLLA (PL) vs GLMM (NLMIXED).	166
34	Person covariate effects estimates for the latent regression Rasch model, LLLA (PL) vs GLMM (NLMIXED).	170
35	Item difficulty estimates for the partial credit model, LLLA (PL) vs GLMM (NLMIXED).	172
36	Item covariate effects estimates for the partial credit model with item covariates, LLLA (PL) vs GLMM (NLMIXED).	175
37	Person covariate effects estimates for the partial credit model with person covariates, LLLA (PL) vs GLMM (NLMIXED).	177

List of Abbreviations

CML	Conditional maximum likelihood
GLMM	Generalized linear/nonlinear mixed models
ICRF	Item category response function
IRF	Item response function
IRT	Item response theory
LLLA	Log linear-by-linear association model
LLLAi	Log linear-by-linear association model with item covariates
LLLAp	Log linear-by-linear association model with person covariates
LLTM	Linear logistic test model
LMA	Log multiplicative association model
MLE	Maximum likelihood estimation
MML	Marginal maximum likelihood
PCM	Partial credit model
PCMi	Partial credit model with item covariates
PCMp	Partial credit model with person covariates
PLE	Pseudolikelihood estimation

List of Symbols

Θ, θ	latent traits
Y, y	responses
X, x	item covariates
Z, z	person covariates
a	item discrimination; entries in item-trait adjacency matrix
b	item parameter
β	item covariate effect
γ	person covariate effect
σ^2	variance
σ_0^2	variance conditional on response pattern
λ	intercept in the LLLA model
$\lambda_{i(y_i)}$	item parameter for i th item with response y_i in the LLLA model
$\nu_{i(y_i)}$	score for i th item with response y_i in the LLLA model
Σ	variance-covariance matrix
$\sigma_{dd'}$	entries in the variance-covariance matrix
T, t	total score
p	subscript for p th person
i	subscript for i th item
h	subscript for h th response category
k	subscript for k th item/person covariate
d	subscript for d th latent trait
N	number of persons (examinees) in the test
I	number of items in the test
K	number of item/person covariates

m the highest response category

D number of latent traits

Chapter 1

Introduction

The general purpose of my research is to model item response data in a log-linear model framework. Item response data are generated when people give responses to items in a battery where the goal is to measure some underlying latent trait or ability. Suppose there are N examinees who respond to I items. Let Y_{pi} be the coded response of the p th person to the i th item. Very often the response is dichotomous, and Y_{pi} is coded as 1 (=correct) and 0 (=incorrect). The observed item responses of a person $\mathbf{Y}_p = (Y_{p1}, \dots, Y_{pI})$ are used to evaluate the person's ability that is assumed to be a continuous latent variable Θ .

Table 1 is an example of a data matrix from a 4-item binary test with 1000 examinees ($N = 1000$ and $I = 4$). The data matrix is a 1000 by 4 with 0-1 entries. Each row represents a person's response to the test items, and each column represents an item responded by all the persons.

Table 1

An Example of a Data Matrix From a 4-item Binary Test

Person	Item 1	Item 2	Item 3	Item 4
1	0	1	0	1
2	1	0	1	0
3	1	1	1	1
4	0	1	1	1
\vdots	\vdots	\vdots	\vdots	\vdots
1000	1	0	0	0

Item Response Theory Models

Item Response Theory (IRT) models have been developed to model the structure of the relationship between the manifest (observed) item responses and latent traits (Lord & Novick, 1968; Lord, 1980; Baker & Kim, 2004; Hambleton, Rogers, & Swaminathan, 1995;

Van der Linden & Hambleton, 1997; Embretson & Reise, 2000; Boomsma, Duijn, & Snijders, 2001). A central assumption in IRT models is local independence; that is, given the latent traits, the responses to items are independent of each other. Specifically, local independence implies that the joint distribution of responses to a set of items can be expressed as the product of the probability of each individual response conditional on the latent trait; that is,

$$\begin{aligned}
p(\mathbf{y}_p|\theta_p) &= P(Y_{p1} = y_{p1}, Y_{p2} = y_{p2}, \dots, Y_{pI} = y_{pI} | \Theta_p = \theta_p) \\
&= \prod_{i=1}^I P(Y_{pi} = y_{pi} | \Theta = \theta_p) \\
&= \prod_{i=1}^I p(y_{pi}|\theta_p), \tag{1.1}
\end{aligned}$$

where $P(Y_{pi} = y_{pi} | \Theta = \theta_p)$ is the conditional probability that a person with latent trait θ_p gives one specific response y_{pi} to item i , and $p(\mathbf{y}_p|\theta_p)$ is the conditional probability that person p with latent trait θ_p gives response pattern \mathbf{y}_p to all items. With local independence, we only need to model the conditional probability of responses to each item given the latent trait.

For dichotomous items with coded values of 0 or 1, Y_{pi} follows a Bernoulli distribution. Let $P_i(\theta_p) = P(Y_{pi} = 1|\theta_p)$ and $Q_i(\theta) = P(Y_{pi} = 0|\theta_p) = 1 - P_i(\theta)$, then

$$p(y_{pi}|\theta_p) = P_i(\theta_p)^{y_{pi}} Q_i(\theta_p)^{1-y_{pi}}.$$

The term $P_i(\theta_p)$ that represents a function of the latent trait is called an item response function (IRF).

The Rasch model (Rasch, 1960, 1961) is the simplest yet a very important IRT model. For a description and discussion of recent developments in the Rasch model and related models including the models with covariates that I am going to discuss in this thesis, see Fischer and Molenaar (1995), De Boeck and Wilson (2004), and von Davier and Carstensen

(2007). The Rasch model specifies the IRF for responses to dichotomous items with a single latent trait. The IRF for the Rasch model is given by

$$P_i(\theta_p) = P(Y_{pi} = 1|\theta_p) = \frac{\exp(\theta_p - b_i)}{1 + \exp(\theta_p - b_i)}, \quad (1.2)$$

where b_i is the difficulty parameter of the i th item. Figure 1 shows the Rasch IRF curves for three items with difficulty $b = -1, 0$, and 1 .

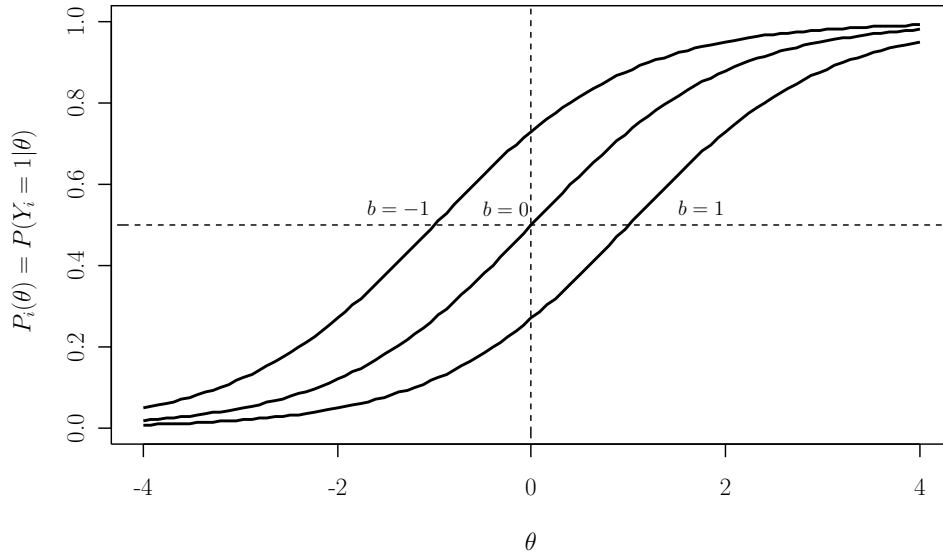


Figure 1. Rasch model's item response function.

The form of the Rasch model reflects the fact that the probability of the response to an item depends not only on the latent trait of the person who answers the item (person parameter θ_p), but it also depends on the characteristics of the item itself (i.e, the difficulty of the item that is represented by the item parameter b_i).

Log-linear Models

Another broad family of statistical models for analyzing discrete response variables is log-linear models (Agresti, 2002). Log-linear or Poisson regression models are regressions for count data that may be entries in a cross-classification by two or more variables. The

dependency structure between the variables is modeled by the linear model that contains, for example, marginal effects, two-way interactions, and three or higher-way interactions. The paradigm of log-linear models has been well developed and is a standard statistical tool for multivariate categorical data analysis.

Item response data can be expressed as cross classifications by items. To see this, consider a simple example of a test with only two items and 1000 examinees (Figure 2). The response data are represented by a 1000×2 matrix and the entries in the matrix are the observed 0-1 responses (Figure 2 (a)). The same data can be represented by a 2×2 contingency table (as shown in Figure 2 (b)), with rows representing the outcomes for the first item and columns for the second item, and the entries in the four cells of the contingency table are the count of persons with each of the response patterns. In this example, there are 100, 200, 300, and 400 persons with the response pattern (0, 0), (0, 1), (1, 0), and (1, 1), respectively. The contingency table is often represented in a “long” form (as shown in Figure 2 (c)), where all the response patterns are listed row by row in a data matrix, and a column of counts is attached to the data matrix. Such “long” form representation is often used for multi-way contingency tables beyond 3-way table.

person	item1	item2
1	0	1
2	1	0
3	1	1
\vdots	\vdots	\vdots
1000	1	0

(a) data matrix

item1	item2	
	0	1
	0	100 200
1	300 400	

(b) 2x2 table

item1	item2	count
0	0	100
1	0	300
0	1	200
1	1	400

(c) long form

Figure 2. Contingency table for a 2-item test with dichotomous responses

In general, for item response data the response to each item is a discrete random variable and responses to all the items can be considered as entries in a multi-way table such that each cell in the table represents a response pattern for a set of items. The cell probability is the probability that a randomly selected person gives response pattern (Y_1, \dots, Y_I) , where

each variable corresponds to a response to an item. The probabilities of response patterns denoted by $P(Y_1, \dots, Y_I)$ are called manifest probabilities. In the previous 4-item binary test example shown in Table 1, there are a total of $2^4 = 16$ response patterns (0000, 1000, 0100, ..., 1111), and each person's actual response to the exam falls into one of the 16 patterns. The data can be considered as a $2 \times 2 \times 2 \times 2$ or four-way table where each dimension of the table represents a dichotomous item in the test. The observed number of persons answering each pattern forms the count data in the table, and the count data can be analyzed using a log-linear model.

Anderson and Vermunt (2000) proposed log multiplicative association (LMA) models for discrete response data that are derived from a latent variable model. For item response data with a unidimensional latent trait, the LMA model has the form

$$\log p(\mathbf{y}) = \lambda + \sum_i \lambda_{i(y_i)} + \sigma^2 \sum_i \sum_{i' > i} \nu_{i(y_i)} \nu_{i'(y_{i'})}, \quad (1.3)$$

where λ is the intercept that ensures that the sum of probabilities over all patterns is 1. The terms $\lambda_{i(y_i)}$, $i = 1, \dots, I$, are the marginal effects of the items, and in a later section we will see they are related to the difficulty of each item. The parameters $\nu_{i(y_i)}$, $i = 1, \dots, I$, are the scores for each item that are related to item discrimination. The term σ^2 is a scale parameter and equals the variance of the latent variable within a response pattern.

The difference between log-linear and item response models is evident by comparing the LMA model for unidimensional and dichotomous items, as given by (1.3), with the Rasch model, as given by (1.1) and (1.2). Although both the LMA model and the Rasch model describe the same underlying structure, log-linear models are expressed in the form of the manifest probabilities $p(\mathbf{y})$, and they do not explicitly include latent traits in the equation. On the other hand, IRT models are expressed as functions of latent variables, and in the form of the conditional probability $p(\mathbf{y}|\theta)$.

The connection between the log-linear models and the IRT models is revealed by the

relationship between the manifest probability and the conditional probability. Integrating over the latent trait θ in the joint distribution of manifest variables and latent variables yields the manifest probabilities,

$$p(\mathbf{y}) = \int p(\mathbf{y}|\theta)p(\theta)d\theta,$$

where $p(\theta)$ is the distribution of the latent trait in the population. Note that both the conditional probability $p(\mathbf{y}|\theta)$ and the latent trait distribution $p(\theta)$ have to be specified in order to get the manifest probabilities.

Connecting Two Paradigms

Since item response data can be analyzed by IRT models and by log-linear models, people have been interested in the relationship between the two approaches. Cressie and Holland (1983) showed that under certain assumptions, the manifest probabilities of the dichotomous item responses that follow a Rasch model will also follow a log-linear model with second order interactions. Holland (1990) further extends the results into what is called the Dutch Identity theorem that provides a general tool to establish the equivalence between the IRT models for dichotomous items and log-linear models under certain conditions. Anderson and Vermunt (2000) proposed the log multiplicative association (LMA) models for the discrete response data that are derived from a statistical graphical model for observed discrete and continuous latent variables. Anderson and Yu (2007) showed that the LMA model is in fact equivalent to the Rasch and the IRT 2-parameter logistic (2PL) models. The equivalence between the IRT models and log-linear models has implications for modeling item response data. It provides a new perspective for fitting the IRT models and opens the doors for new tools for analyzing item response data. Since IRT models have been proven to be useful and have wide applications in educational testing and health outcome research, this provides broad application of log-linear models such as LMA or Log-linear by

linear Association Models (LLA) in these fields.

Although log-linear models as item response models provide vast potentials for the analysis of item response data, computational problems have to be addressed before the log-linear models can be widely used for most applications. The computational cost of maximum likelihood estimation (MLE) of log-linear models is proportional to the number of cells in the multi-way table. As the number of items in the test increases, the total number of responses increases exponentially, and as a result fitting log-linear models soon becomes infeasible. The MLE procedure described in Anderson and Vermunt (2000) as implemented in the statistical package ℓ_{EM} (Vermunt, 1997) cannot be used to fit data sets beyond 12 dichotomous items within a reasonable time or amount of memory. An achievement test may have more than 50 and even a hundred items, so the MLE procedures used by ℓ_{EM} and other programs will not work for moderate to large problems. To solve the computational cost problem, a pseudolikelihood estimation (PLE) approach was discussed for Rasch models by Strauss (1992), Zwiderman (1995), Smit and Kelderman (2000), and Anderson, Li, and Vermunt (2007). Only the method described in Anderson, Li, and Vermunt (2007) can handle polytomous or binary items and single or multiple latent traits. I implemented the PLE procedure in R and published the R package ‘plRasch’ (cran.r-project.org). The paper describing the first version of the R package ‘plRasch’ was published in the Journal of Statistical Software (Anderson, Li, & Vermunt, 2007). We found that the pseudolikelihood approach is computationally efficient, recovers parameters extremely well, and can be used for large number of items (e.g., 100 items).

Research Objectives

Given the success of PLE for models in the Rasch family, my thesis research objective is to extend my previous work on log-linear models for item response data, and implement the PLE procedures in a software package that can be used to analyze item response data with covariates. My strategy is to start with IRT models with covariates (De Boeck & Wilson,

2004), from which I derive the form of the corresponding log-linear model by utilizing the relationship between the IRT models and log-linear models through the Dutch Identity theorem. I will use pseudolikelihood estimation to estimate the parameters in the derived log-linear models. This would solve the estimation problems for large data sets.

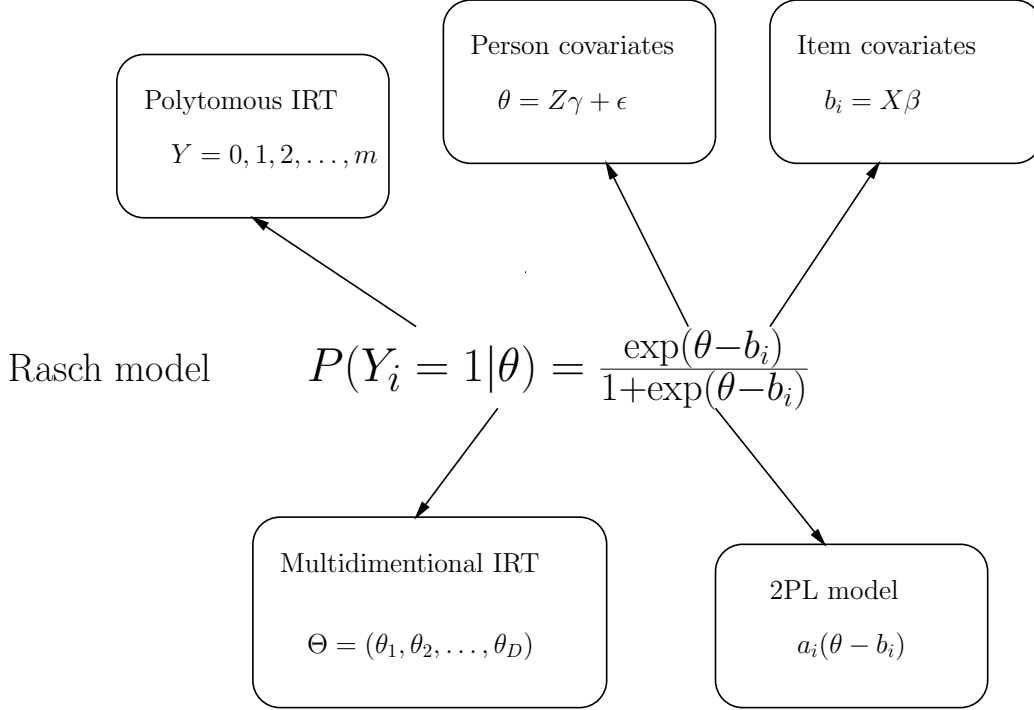


Figure 3. Extended Rasch models.

Many IRT models can be seen as extensions of the Rasch model by modifying different features in the Rasch IRF (Figure 3). For example, the two-parameter logistic (2PL) model (Birnbbaum, 1968) adds another item parameter (a_i) for each item to represent discrimination power that differs across items. By including covariates related to item properties and person properties, the IRT models increases their explanatory power. Well-known examples are the linear logistic test model (LLTM) (Fischer, 1973), and the latent regression Rasch model (Zwinderman, 1991). Polytomous IRT models (Ostini & Nering, 2006; Nering & Ostini, 2010) study items that have multiple response outcomes that is often seen in applications. Some well-known polytomous models are Masters' partial credit model (Masters, 1982), Bock's nominal response model (Bock, 1972), Samejima's graded response model (Samejima,

1970), and the rating scale model (Muraki, 1990). When a single latent variable is not enough to explain all the dependency in the items so that unidimensional local independence is violated, multidimensional IRT models (Reckase, 2009) that assume multiple latent traits may be necessary.

The richness of the family of IRT models is a motivation to draw parallel extensions on the log-linear models for item response data. In the following sections, I will lay out the research objectives of this thesis that include: adding person and item covariates, extending to polytomous items, and to multi-dimensional models.

Adding covariates. With respect to covariates, I intend to extend the log-linear models for item response data to include covariates that are attributes of items and persons. Although the Dutch Identity exists for dichotomous items, I will present an extension so that I can extend log-linear models to polytomous items and multiple latent traits.

Starting with a simple unidimensional dichotomous model, I will show how to add covariates to the model. Item covariates describe the properties of items, such as item type, behavior mode, situation type, and others. Person covariates describe characteristics or attributes of examinees, such as gender, social economic status (SES), ethnicity, and others. In IRT models, the linear logistic test model (LLTM) (Fischer, 1973) was proposed as an extension of the Rasch model that incorporates item covariates, and the latent regression model extends the Rasch model by adding person covariates. Furthermore, models with both item and person covariates have also been proposed. The IRT models that incorporate person and item interactions can be used to model differential item functioning (DIF) (Holland & Wainer, 1993), which is an important research topic on the fairness of test design.

Given the fact that log-linear models are Poisson regression models, it seems that adding covariates is straightforward; however, the log-linear models that I study in this thesis are used as IRT models. The models have a specific structure rather than unrestricted Poisson regression models. Specifically, I will add covariates to the log-linear models with

second order interactions. Since the models will be used as IRT models, the effects of the covariates would retain their interpretation as specified in IRT models.

In the literature there have been efforts in adding covariates to log-linear models with second order interactions. Joe and Liu (1996) proposed a model for multivariate binary response data with covariates. Although Joe and Liu (1996)'s proposal was not IRT based, their model is the same as an LLLA model with person covariates. They specified the model based on compatible conditionally specified logistic regressions. They started with logistic regressions for each variable conditional on other variables and covariates. They derived the conditions to ensure that the set of logistic models were compatible or consistent with some joint distribution, and the joint distribution was shown to be an LLLA model. This conditional specification approach is very closely related to the pseudolikelihood estimation procedure that is described in Anderson, Li, and Vermunt (2007). Conditionally specified logistic regressions can also be used as a tool to derive LLLA models from IRT methodology, as shown in Anderson and Yu (2007), Anderson, Li, and Vermunt (2007), and Anderson, Verkuilen, and Peyton (2010). In Anderson and Vermunt (2000), there is an example of a person variable (i.e., gender) being incorporated into the LMA model using the graphical model approach. In Anderson and Böckenholt (2000), an LMA model with covariates was proposed and illustrated by analyzing $\text{SES} \times \text{program type}$ as a function of student mean achievement test scores. In Tettegah and Anderson (2007), an LLLA model with continuous person covariates was used for binary response data. In Anderson et al. (2010), responses of four polytomous items are analyzed where treatment conditions and item content information are used as item covariates.

What I am proposing in this thesis is a systematic way of developing the log-linear models with covariates starting with IRT models with covariates. The approach I will use (the solid lines in Figure 4) is different from the conditionally specified approach (the dotted lines in Figure 4) used in Joe and Liu (1996); Anderson, Li and Vermunt (2007); Anderson and Yu (2007); and Anderson et al. (2010). Instead of starting with a set of logistic

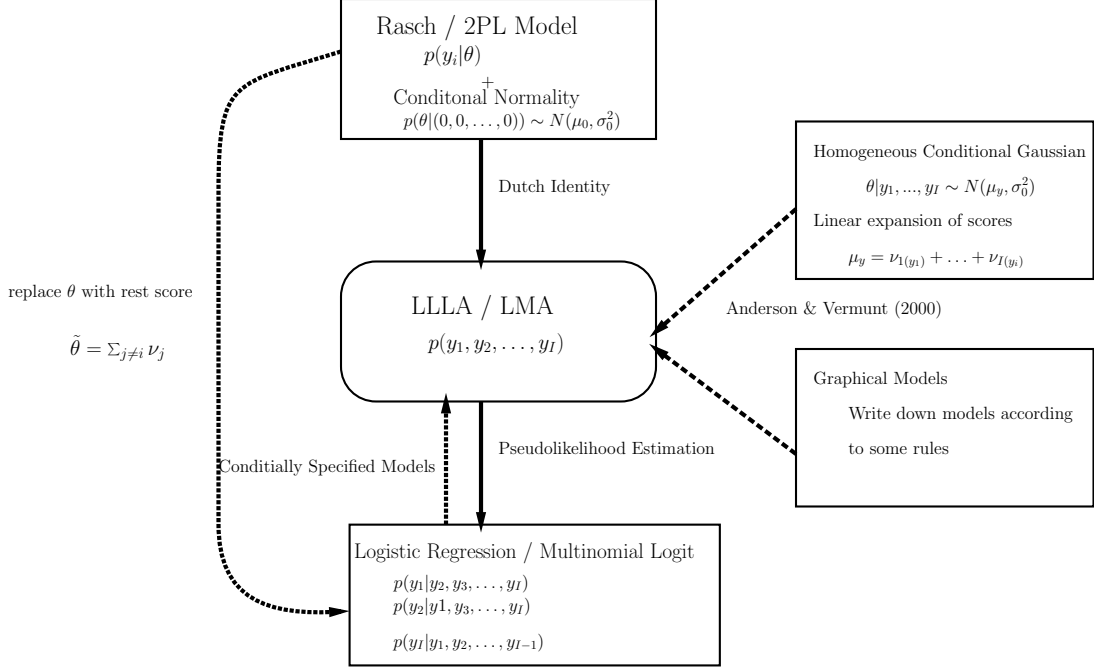


Figure 4. Research strategies.

regressions of each response conditional on other responses and covariates, I start with IRT models with each response conditional on the latent trait and covariates. The advantage of starting directly with IRT models is that the parameters in the models will have a clear interpretation as described in corresponding item response theory models. Furthermore, the pseudolikelihood method provides a computationally efficient way to estimate the parameters and can be used in practical problems.

Polytomous and multidimensional models. I will extend the Dutch Identity theorem to polytomous items and use the polytomous Dutch Identity theorem to derive log-linear models that can be used for ordinal response items that are equivalent to polytomous IRT models such as the partial credit model. In this thesis, I will elaborate how to derive the log-linear models from polytomous IRT models, and develop the log-linear models with covariates that can handle polytomous items.

IRT models with multidimensional latent traits have been an active research area (Ackerman, 1994; Kelderman & Rijkes, 1994; Reckase, 1997a, 1997b, 2009). Log-linear

models estimated by pseudolikelihood methods have a clear advantage in computational feasibility relative to the IRT models estimated by a traditional marginal maximum likelihood (MML) approach. In MML, the latent traits have to be numerically integrated out, assuming multivariate normality or some other distribution for the latent traits. Therefore if the number of the latent traits is large, the computational cost of the numerical integration increases exponentially. The log-linear model approach proposed in my thesis does not suffer this kind of problem because no numerical integration is involved in this method.

Summary. To summarize, the following is what is accomplished in this thesis: the development of log-linear models for item response data that incorporate item and person covariates, with the ability to handle both polytomous items and multidimensional latent traits; the complete presentation, from proof to interpretation of the relationship between log-linear models and the corresponding IRT models; and the development and implementation of computationally efficient estimation procedures using the pseudolikelihood approach.

The remainder of this thesis is structured as follows. In Chapter 2, I will lay out the theoretical basis for this thesis research. I will give an introduction to the LLLA model and show how it can be seen as a equivalent form of the Rasch model. I will introduce the Dutch Identity theorem that is an important tool to prove the equivalence of the log-linear models and the IRT models, and for developing the models in this thesis. In Chapter 3, pseudolikelihood estimation is introduced and I will show how it is applied to the LLLA model. In Chapter 4, I will present the development of LLLA models with item covariates and with person covariates. In Chapter 5, I will present and prove the Dutch Identity theorem for polytomous items and use it to develop polytomous LLLA models that are equivalent to polytomous IRT models, including those with item covariates and person covariates. In Chapter 6, I will derive the LLLA models equivalent to the multidimensional IRT models, including those with item and person covariates. In Chapter 7, simulation studies for all the models developed in this thesis are presented. Chapter 8 contains applications of the

models to a real data set. Finally, the thesis ends with Chapter 9 with a summary and some concluding remarks.

Chapter 2

LLLA as IRT Models

In this chapter, I am going to introduce the equivalence between LLLA models and IRT Rasch models. First I will present the set of assumptions that lead to an LLLA model as presented in the original LMA/LLLA paper (Anderson & Vermunt, 2000). Next I will present another set of assumptions starting with a Rasch model that lead to exact the same LLLA model; thus the LLLA model can be seen as a special form of the Rasch model. Then I will present the Dutch Identity theorem. This theorem is used to derive the LLLA model from the Rasch model. It will be used as the tool to develop the LLLA models with covariates throughout my thesis research. Finally I will show how to estimate the item and person parameters in the Rasch model by fitting the LLLA model.

Assumptions for LLLA model

For a test with I items, the responses of an examinee to the items is a realization of a random vector $\mathbf{Y} = (Y_1, Y_2, \dots, Y_I)$. The value of the random vector is one of the $\prod_{i=1}^I n_i$ possible response patterns, where n_i is the number of possible response options for item i . If all the items are binary, then there are 2^I possible response patterns. Denote $p(\mathbf{y}) = P(\mathbf{Y} = \mathbf{y})$, as the manifest probability for response pattern \mathbf{y} . Note that

$$\sum_{\text{all } \mathbf{y}} p(\mathbf{y}) = 1.$$

The dependence structure of multiple variables can often be represented by a graph, an integral tool in graphical models (Whittaker, 1990). Consider the case with a single latent trait. The dependency among the variables, including the discrete manifest variables Y_1, \dots, Y_I and the continuous latent trait variable Θ , is reflected in the graph shown in Figure 5. In this graphical representation, manifest variables are represented by square boxes;

latent variables are represented by circles; the possible dependence between the variables is represented by lines and paths connecting the variables; and conditional independence is represented by the absence of lines or paths connecting variables.

The concept of local independence is represented in graphical models. As we can see in Figure 5, all the manifest variables Y_i may be dependent on the latent trait variable Θ , as is seen by the edges connecting the manifest variables to the latent variable. Associations among manifest variables are expected to be observed because there are paths connecting the manifest variables through the random latent variable Θ . Since there is no direct edge between any pair of the manifest variables, the manifest variables are independent of each other conditional on the latent trait.

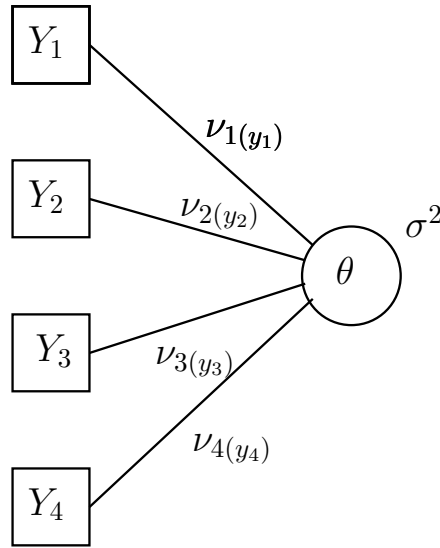


Figure 5. Graphical structure of the 1-D LLLA model.

Anderson and Vermunt (2000) derived the LMA models (of which LLLA models are special cases) for the structure represented in Figure 5 by assuming that:

- The observed variables given the latent traits are conditionally independent, (i.e., local independence):

$$p(\mathbf{y}|\boldsymbol{\theta}) = P(\mathbf{Y} = \mathbf{y}|\boldsymbol{\Theta} = \boldsymbol{\theta}) = \prod_{i=1}^I p(y_i|\boldsymbol{\theta}).$$

- The joint distribution of the discrete variables Y_1, \dots, Y_I and continuous variable Θ is a homogeneous conditional Gaussian (Lauritzen & Wermuth, 1989). A homogeneous conditional Gaussian distribution assumes that the conditional distribution of the continuous variable given the discrete variable is normal with constant variance:

$$\Theta|\mathbf{Y} = \mathbf{y} \sim N(\mu_{\mathbf{y}}, \Sigma).$$

- The mean of the conditional Gaussian distribution is set equal to a linear expansion of scores:

$$\mu_{\mathbf{y}} = \sigma^2 \sum_{i=1}^I \nu_{i(y_i)}$$

In the case of the model with a unidimensional latent trait, the conditional distribution of the latent trait is a univariate normal distribution $\theta|\mathbf{y} \sim N(\mu_{\mathbf{y}}, \sigma^2)$. Under this assumption, Anderson and Vermunt (2000) show that the log manifest probability is given by

$$\log p(\mathbf{y}) = \lambda + \sum_i \lambda_{i(y_i)} + \sigma^2 \sum_i \sum_{i' > i} \nu_{i(y_i)} \nu_{i'(y_{i'})}. \quad (2.1)$$

Model (2.1) is called a log multiplicative association (LMA) model if we assume the scores for each item $\nu_{i(y_i)}$ are unknown and need to be estimated. It is a nonlinear model because of the multiplicative terms $\sigma^2 \nu_{i(y_i)} \nu_{i'(y_{i'})}$. If we assign the scores $\nu_{i(y_i)}$ for each item so they are fixed numbers,¹ then (2.1) is a linear model because the right-hand side of the equation is a linear function of the unknown parameters (i.e., λ , $\lambda_{i(y_i)}$, and σ^2), and the model is called a log linear-by-linear association (LLLA) model (Agresti, 2002).

The distribution $p(\theta|\mathbf{y})$ is the posterior distribution of the latent trait given the

¹ For the assignment of the scores, one popular way is to use the natural score: $\nu_{i(y_i)} = y_i$. For example, for binary items (0, 1 responses) $\nu_{i(0)} = 0$ and $\nu_{i(1)} = 1$; for ternary items (0, 1, 2 responses) $\nu_{i(0)} = 0$, $\nu_{i(1)} = 1$ and $\nu_{i(2)} = 2$. However, other scores are also possible, for example, for binary times $\nu_{i(0)} = -1/\sqrt{2}$ and $\nu_{i(1)} = 1/\sqrt{2}$. It is even possible to have non-uniform (scores dependent on items) but fixed scores to represent the different discriminatory power of the items. In this thesis, I will keep using the notation $\nu_{i(y_i)}$ for the scores instead of a specific score assignment, just keep in mind that the scores are fixed.

response pattern. In practice, this posterior distribution helps us draw inference regarding an examinee's ability from the observed responses to the items; namely, by linear expansion assumption, $E(\theta|\mathbf{y}) = \mu_{\mathbf{y}} = \sigma^2 \sum_{i=1}^I \nu_{i(y_i)}$. Although the latent variable Θ is not explicitly present in model (2.1), once we fit the model we can put the estimates of $\hat{\mu}_{\mathbf{y}}$ and $\hat{\sigma}^2$ into the normal distribution and use it to give a credible interval for the latent trait (i.e., each person's ability) within the response pattern \mathbf{y} .

LLLA Model as Rasch Model

The relationship between an LLLA model and the Rasch model may not seem obvious when we look at the set of assumptions that leads to the LLLA model in (2.1); however, Anderson and Yu (2007) showed that the same LLLA model (2.1) can be derived by starting from the following set of assumptions that Holland (1990) made. These assumptions are

- Local independence:

$$p(y_1, \dots, y_I | \theta) = \prod_i p(y_i | \theta).$$

- The data are generated from Rasch model, i.e., the item response function has the form:

$$P_i(\theta) = P(Y_i = 1 | \theta) = \frac{\exp(\theta - b_i)}{1 + \exp(\theta - b_i)}.$$

- The conditional distribution of θ given one of the response patterns is a normal distribution:²

$$p(\theta | \mathbf{y}_0) \sim N(\mu_{y_0}, \sigma_0^2).$$

As we can see, the first two statements are the assumptions for the Rasch model, as defined by specifying $p(\mathbf{y}|\theta)$; the third assumption states that the posterior distribution of the

²For the variance of the conditional normal distribution I use σ_0^2 . With the subscript '0', I emphasize the fact that the latent trait is conditional on one response pattern. I reserve σ^2 without the subscript later in the MML formulation of the Rasch model, where it stands for the variance of the (unconditional) distribution of the latent traits for the whole population.

latent trait conditional on one of the response patterns follows a normal distribution, which specifies $p(\theta|\mathbf{y}_0)$. With these assumptions, Holland (1990) derived the marginal distribution $p(\mathbf{y})$ by using a proposed tool that he called the “Dutch Identity theorem”, and it is the LLLA model (2.1) (see the next section for details).

Although the third statement only assumes the normality of θ given one of the response patterns, Anderson and Yu (2007) proved that together with the Rasch (or 2PL) model, it is necessarily true that θ conditional on every response pattern has a normal distribution with the same variance σ_0 :

$$p(\theta|\mathbf{y}) \sim N(\mu_{\mathbf{y}}, \sigma_0^2).$$

Recall that this is exactly the homogeneous conditional Gaussian distribution assumption used in the derivation of LLLA model by Anderson and Vermunt (2000). Thus the two sets of assumptions stated in this section and the previous section are actually equivalent.

The relationship between the LLLA model and the Rasch model can be summarized as:

$$\text{LLLA model} = \text{Rasch model} + \text{Conditional Normality of } \theta.$$

The LLLA model can be seen as a Rasch model plus a restriction on the distribution of the latent trait (i.e., conditional normality).

In the literature, the Rasch model is classified to different formulations according to additional assumptions made on θ and the estimation methods (de Leeuw & Verhelst, 1986). The original Rasch model itself does not include any assumption regarding the distribution of the latent traits in the population. The conditional maximum likelihood (CML) method is used to estimate the parameters in the Rasch models because of the existence of sufficient statistics for the latent trait. We can call the Rasch model with no distributional assumption on the latent trait as the CML formulation. Under the CML formulation, the population distribution of the latent traits is actually a nonparametric

distribution (de Leeuw & Verhelst, 1986). Another formulation for the Rasch model is to assume the marginal distribution of the latent trait follows a normal distribution $\theta \sim N(0, \sigma^2)$, and the model is estimated by marginal maximum likelihood (MML) method. Thus we call this the MML formulation. The LLLA model is a formulation of the Rasch model that lies in between: for individuals within the same response pattern, the distribution of their latent trait is assumed to be normal; therefore the (marginal) population distribution of the latent trait is a mixture of normal distributions:

$$p(\theta) = \sum_{\text{all } \mathbf{y}} p(\theta|\mathbf{y})p(\mathbf{y}) = \sum_{\text{all } \mathbf{y}} N(\mu_{\mathbf{y}}, \sigma_0^2)p(\mathbf{y}).$$

Table 2 summarizes the three different formulations of Rasch models as described in the previous paragraph.

Table 2

Different Formulations of the Rasch Model

Model	Model for $p(\mathbf{y} \theta)$	Distributional assumption for $p(\theta)$	Estimation Method
Rasch (CML)	Rasch IRF	no assumption (non-parametric distribution)	Conditional maximum likelihood
Rasch (MML)	Rasch IRF	$\theta \sim N(0, \sigma^2)$	Marginal maximum likelihood
LLLA	Rasch IRF	$\theta \mathbf{y} \sim N(\mu_{\mathbf{y}}, \sigma_0^2)$	Pseudolikelihood estimation, MLE

The conditional normality of the latent trait θ given the response patterns is an important part in the LLLA model that distinguishes it from other formulations of the Rasch model. In many cases, even if the *exact* conditional normality may not hold, very often the conditional normality holds *approximately*, so that the LLLA model will still be useful in such cases. Chang and Stout (1993) proved the asymptotic posterior normality given the response patterns under nonrestrictive nonparametric assumptions and dichotomous IRT

models, and the result was extended to polytomous IRT models in Chang (1996). The main results in Chang and Stout (1993) and Chang (1996) are that for tests with large number of items, the posterior distribution of the latent trait given the response patterns is approximately equal to the normal distribution $N(\hat{\theta}_I, \hat{\sigma}_I^2)$, where $\hat{\theta}_I$ denotes the MLE of θ and $\hat{\sigma}_I$ is the SE of $\hat{\theta}_I$ calculated from the Fisher information. This result suggests that after fitting the LLLA model, if we use the estimated posterior mean $\hat{\mu}_{\mathbf{y}}$ as the estimate for θ , and estimated posterior variance $\hat{\sigma}_0$ as the SE, the estimate and the SE will be very close to the estimates and the SE obtained by MLE from the IRT model. Holland (1990) made a conjecture (Dutch Identity conjecture) that the LLLA model form is a limiting form for *all* “smooth” unidimensional IRT models as length of a test tends to infinity. Zhang and Stout (1997) gave counter examples to show that the Dutch Identity conjecture does not hold in general but has to have some strong conditions to hold.

Dutch Identity

The equivalence of LLLA model and the Rasch model can be established by different approaches (Anderson and Yu, 2007). An early paper that establishes the log-linear form of manifest probability starting with Rasch model is Cressie and Holland (1983). That result is generalized in Holland (1990) in the form of the “Dutch Identity” theorem, which is a general tool to establish the equivalence between the IRT models and the log-linear models. In (Holland, 1990), the Dutch Identity theorem deals with dichotomous items and can handle multiple latent traits. In this thesis, I will extend the Dutch Identity to polytomous items, and use it as a tool in the derivation of LLLA models that can handle covariates, polytomous items, and multiple latent traits. For now, I will review the Dutch Identity theorem and use it to derive the LLLA model from the Rasch model.

Recall that for a test with I items with binary responses, the response vector is $\mathbf{Y} = (Y_1, Y_2, \dots, Y_I)$. Let $\mathbf{y} = (y_1, y_2, \dots, y_I)$ be the value of the response vector, such that

the manifest probability is given by

$$p(\mathbf{y}) = P(\mathbf{Y} = \mathbf{y}) . \quad (2.2)$$

Given the latent trait, by local independence, the conditional probability of a response pattern is

$$p(\mathbf{y}|\theta) = P(\mathbf{Y} = \mathbf{y}|\theta) = \prod_{i=1}^I P(Y_i = y_i|\theta) . \quad (2.3)$$

For each item, the response given the latent trait follows a Bernoulli distribution, given by

$$P(Y_i = y_i|\theta) = P_i(\theta)^{y_i} Q_i(\theta)^{1-y_i} , \quad (2.4)$$

where $P_i(\theta) = P(Y_i = 1|\theta)$ and $Q_i(\theta) = P(Y_i = 0|\theta) = 1 - P_i(\theta)$.

Suppose that the latent trait follows a general distribution with the pdf $p(\theta)$, then we can calculate the manifest probability by

$$p(\mathbf{y}) = \int \prod_{i=1}^I P_i(\theta)^{y_i} Q_i(\theta)^{1-y_i} p(\theta) d\theta . \quad (2.5)$$

The Dutch Identity is given by the following theorem:

Theorem 2.1. (*Dutch Identity, Holland 1990*) *If the manifest probabilities $p(\mathbf{y})$ satisfy (2.5), then for any fixed response pattern \mathbf{y}_0 ,*

$$\frac{p(\mathbf{y})}{p(\mathbf{y}_0)} = E\{\exp[(\mathbf{y} - \mathbf{y}_0)^T \boldsymbol{\delta}(\theta)] | \mathbf{Y} = \mathbf{y}_0\} , \quad (2.6)$$

where $\boldsymbol{\delta}(\theta) = (\delta_1(\theta), \dots, \delta_I(\theta))^T$ and

$$\delta_i(\theta) = \log \frac{P_i(\theta)}{Q_i(\theta)} . \quad (2.7)$$

Note that in (2.6) the expectation is taken with respect to the random variable θ . So

(2.6) equals

$$\int \exp \left[\sum_{i=1}^I (y_i - y_0) \delta_i(\theta) \right] p(\theta | \mathbf{Y} = \mathbf{y}_0) d\theta, \quad (2.8)$$

where $p(\theta | \mathbf{Y} = \mathbf{y}_0)$ is the conditional distribution of the latent trait θ given the response pattern \mathbf{y}_0 .

The Dutch Identity tells us that for an IRT model that satisfies local independence, if we know the manifest probability for some reference response pattern \mathbf{y}_0 (i.e, $p(\mathbf{y}_0)$), and the posterior probability of the latent trait under the reference pattern $p(\theta | \mathbf{Y} = \mathbf{y}_0)$, then we can calculate the manifest probability for any response pattern.

The following corollary established the relationship between the Rasch model and the log-linear model with quadratic terms, for which the LLLA model is a special case.

Corollary 2.2. *(Holland 1990) If for some \mathbf{y}_0 we have a normal posterior distribution for the D -dimensional trait*

$$\boldsymbol{\theta} | \mathbf{Y} = \mathbf{y}_0 \sim N_D(\mu_{y_0}, \Sigma_{y_0}), \quad (2.9)$$

and the logit of IRF is a linear function of θ

$$\delta_i(\boldsymbol{\theta}) = \delta_i(\mu_{y_0}) + \mathbf{a}_i^T (\boldsymbol{\theta} - \mu_{y_0}), \quad (2.10)$$

then

$$\log p(\mathbf{y}) = \log p(\mathbf{y}_0) + (\mathbf{y} - \mathbf{y}_0)^T \boldsymbol{\delta}(\mu_{y_0}) + \frac{1}{2} (\mathbf{y} - \mathbf{y}_0)^T A^T \Sigma_{y_0} A (\mathbf{y} - \mathbf{y}_0), \quad (2.11)$$

where the matrix $A = [a_1 | a_2 | \dots | a_I]$.

A direct application of the above corollary is for the special case of the unidimensional (1-D) Rasch model where

$$\delta_i(\theta) = \log \frac{P_i(\theta)}{Q_i(\theta)} = \theta - b_i. \quad (2.12)$$

If we assume for some reference response \mathbf{y}_0 that

$$\theta|\mathbf{y}_0 \sim N(\mu_{y_0}, \sigma_{y_0}^2), \quad (2.13)$$

then we will have

$$\log p(\mathbf{y}) = \log p(\mathbf{y}_0) + \sum_{i=1}^I (y_i - y_{0i})(-b_i + \mu_{y_0}) + \frac{1}{2}\sigma_{y_0}^2 \left(\sum_i (y_i - y_{0i}) \right)^2. \quad (2.14)$$

Now we can show that (2.14) has the same form as the LLLA model. Since \mathbf{y}_0 is arbitrary, let the reference response be the response of all 0s (i.e., $\mathbf{y}_0 = \mathbf{0} = (0, 0, \dots, 0)$), then

$$\log p(\mathbf{y}) = \log p(\mathbf{0}) + \sum_{i=1}^I [(-b_i + \mu_0)y_i + \frac{1}{2}\sigma_0^2 y_i^2] + \sigma_0^2 \left(\sum_i \sum_{i'>i} y_i y_{i'} \right). \quad (2.15)$$

Comparing this with the LLLA model

$$\log p(\mathbf{y}) = \lambda + \sum_{i=1}^I \lambda_{i(y_i)} + \sigma^2 \left(\sum_i \sum_{i'>i} \nu_{i(y_i)} \nu_{i'(y_{i'})} \right), \quad (2.16)$$

the correspondence between the parameters in the LLLA model (2.16) and the Rasch model parameters in (2.15) is as follows:

- $\lambda = \log p(\mathbf{0})$, the intercept is equal to the log of the manifest probability of the all-0 responses.
- $\lambda_{i(0)} = 0$ and $\lambda_{i(1)} = -b_i + \mu_0 + 1/2\sigma_0^2$, the main item effect is equal to the negative of the item difficulty (i.e., easiness) plus some constant, and this constant is the posterior mean plus half of the variance of the latent trait with all-0 response.
- $\sigma^2 = \sigma_0^2$, the posterior variance of the latent trait given all the responses are 0.
- $\nu_{i(y_i)} = y_i$, the raw scores.

Now we can draw the conclusion that for the Rasch model, if we assume that the posterior distribution of θ given all-0 response is a normal distribution, then the manifest probability follows the LLLA model.

Fitting Rasch by LLLA

The equivalence relationship between the Rasch model and the LLLA model provide one way to fit the Rasch model through fitting the corresponding LLLA models. After fitting the LLLA model to the item response data, we transform the parameter estimates in the LLLA model into the item and person parameters in the Rasch model.

Item parameters. For the item difficulty parameters b_i , $i = 1, \dots, I$ in the Rasch model, we can calculate them from the LLLA parameters by

$$b_i = -\lambda_{i(1)} + \mu_0 + \frac{1}{2}\sigma_0^2. \quad (2.17)$$

In the above equation, μ_0 is not in the LLLA model equation. As stated in equation (2.13), μ_0 is the mean of latent trait θ conditional on the reference response pattern $\mathbf{Y} = \mathbf{0}$. It is not identifiable unless constraints are applied. Usually one can constraint μ_0 such that the sample mean of the estimated person parameter $\hat{\theta}$ for the persons in the data set is 0 (this is called “anchoring” according to the person parameters; see Embretson and Reise (2000, pp 129-131)). The formula for estimating μ_0 is given later in this section (see Equation (2.27)).

Person parameters. In order to estimate the person parameters θ , we need to derive the conditional distribution $p(\theta|y)$. It is given by the following theorem.

Theorem 2.3. *(Conditional normality for any pattern) Under the assumptions of (A) the Rasch IRF, see Equation (2.12), and (B) conditional normality on the reference pattern, see*

Equation (2.13), for any response pattern \mathbf{y} , the conditional distribution of $\theta|\mathbf{y}$ is

$$\theta|\mathbf{y} \sim N(\mu_0 + \sigma_0^2(T - T_0), \sigma_0^2), \quad (2.18)$$

where $T = T(\mathbf{y}) = \sum_{i=1}^I y_i$ is the total score for response pattern \mathbf{y} , and $T_0 = T(\mathbf{y}_0) = \sum_{i=1}^I y_{0i}$ is the total score for the reference pattern \mathbf{y}_0 .

Proof. From (2.13), the pdf for $p(\theta|\mathbf{y}_0)$ is

$$p(\theta|\mathbf{y}_0) = \frac{1}{\sqrt{2\pi}\sigma_0} \exp\left(-\frac{(\theta - \mu_0)^2}{2\sigma_0^2}\right). \quad (2.19)$$

Since $p(\mathbf{y}, \theta) = p(\mathbf{y}|\theta)p(\theta) = p(\theta|\mathbf{y})p(\mathbf{y})$, we have

$$\frac{p(\mathbf{y}|\theta)}{p(\mathbf{y}_0|\theta)} = \frac{p(\theta|\mathbf{y})}{p(\theta|\mathbf{y}_0)} \frac{p(\mathbf{y})}{p(\mathbf{y}_0)}. \quad (2.20)$$

The ratio of the likelihood functions for pattern \mathbf{y} and the reference pattern \mathbf{y}_0 is

$$\begin{aligned} & \frac{p(\mathbf{y}|\theta)}{p(\mathbf{y}_0|\theta)} \\ &= \frac{\prod_i P_i(\theta)^{y_i} Q_i(\theta)^{1-y_i}}{\prod_i P_i(\theta)^{y_{0i}} Q_i(\theta)^{1-y_{0i}}} \\ &= \prod_i \left(\frac{P_i(\theta)}{Q_i(\theta)} \right)^{(y_i - y_{0i})} \\ &= \prod_i \exp\{(\theta - b_i)(y_i - y_{0i})\} \\ &= \exp \left[\theta \left(\sum_i y_i - \sum_i y_{0i} \right) + \sum_i b_i (y_i - y_{0i}) \right]. \end{aligned} \quad (2.21)$$

Substituting (2.19), (2.21) into (2.20), we get

$$\begin{aligned}
& p(\theta|\mathbf{y}) \\
&= p(\theta|\mathbf{y}_0) \frac{p(\mathbf{y}|\theta)}{p(\mathbf{y}_0|\theta)} \frac{p(\mathbf{y})}{p(\mathbf{y}_0)} \\
&\propto \exp \left[-\frac{1}{2\sigma_0^2}(\theta - \mu_0)^2 + \theta \left(\sum_i y_i - \sum_i y_{0i} \right) \right] \\
&\propto \exp \left[-\frac{[\theta - \mu_0 - \sigma_0^2(\sum_i y_i - \sum_i y_{0i})]^2}{2\sigma_0^2} \right]
\end{aligned} \tag{2.22}$$

We can clearly see that this is a pdf of a normal distribution,

$$\theta|\mathbf{y} \sim N(\mu_0 + \sigma_0^2[T(\mathbf{y}) - T(\mathbf{y}_0)], \sigma_0^2). \tag{2.23}$$

where $T(\mathbf{y}) = \sum_i y_i$ and $T(\mathbf{y}_0) = \sum_i y_{0i}$ are total scores. □

If we use all-zero pattern as the reference: $\mathbf{y}_0 = \mathbf{0}$, Equation (2.24)

$$\theta|\mathbf{y} \sim N(\mu_0 + \sigma_0^2 T, \sigma_0^2). \tag{2.24}$$

This provides a way of scoring the persons, (i.e., estimating the latent trait for each individual). For a person with response pattern \mathbf{y} , we can use the estimated posterior mean to estimate the latent trait

$$\hat{\theta} = E(\theta|\mathbf{y}) = \hat{\mu}_0 + \hat{\sigma}_0^2 T, \tag{2.25}$$

and use the estimated standard deviation of $\theta|\mathbf{y}$ as the standard error,

$$se(\hat{\theta}) = \hat{\sigma}_0. \tag{2.26}$$

Since μ_0 is unidentifiable, constraints are imposed such that the sample mean of $\hat{\theta}$ is 0 (this is called “anchoring” according to the person parameters, see Embretson and Reise (2000,

pp 129-131)). Under this constraint, it is derived that

$$\hat{\mu}_0 = -\hat{\sigma}_0^2 \bar{T}, \quad (2.27)$$

where $\bar{T} = \sum_{p=1}^N T(\mathbf{y}_p)/N$ is the sample mean of the total scores over all the persons in the data set. Therefore the estimate of the latent traits is given by

$$\hat{\theta} = \hat{\sigma}_0(T - \bar{T}). \quad (2.28)$$

The estimate is a linear function of the total scores. The sample mean of $\hat{\theta}_p$, $p = 1, \dots, N$ is 0.

Population latent trait distribution. We have seen two ways to specify the latent trait distribution. The first way is to directly specify a normal (marginal) distribution for the latent trait, $\theta \sim N(0, \sigma^2)$. This method is popular and is used in the MML formulation of the Rasch model. However, this latent distribution will not lead to an LLLA model for the manifest probability.

The second way is to assume the conditional distribution of the latent trait given some response pattern is a normal distribution. We have seen that under this assumption the manifest probability follows an LLLA model. Further more, the conditional distribution of the latent trait given any pattern is also a normal distribution with the same variance as in the reference pattern but with a different mean. This means that it essentially assumes the latent trait has a distribution of a mixture of normal distributions with the same variance.

$$\theta \sim \sum_{all \mathbf{y}} p(\mathbf{y}) N(\mu_{\mathbf{y}}, \sigma_0^2) \quad (2.29)$$

Theoretically a finite mixture of normal distributions is different from a normal distribution. However, in practice, the difference between modeling the population latent trait

with a single normal distribution and modeling with a mixture of normal distribution may not be that important. Both ways can be good choices for the population distribution and we may get similar item and person parameter estimates under the two ways. See the simulation studies in the next section.

Now consider the situation that the two distributions of the latent trait are similar to each other,

$$N(0, \sigma^2) \approx \sum_{all \mathbf{y}} p(\mathbf{y}) N(\mu_{\mathbf{y}}, \sigma_0^2). \quad (2.30)$$

What is the relationship between the global variance σ^2 and the conditional variance σ_0^2 ? To answer the question, we can match the variance of the two distributions, and get

$$\sigma^2 = \text{var}(\sum_{all \mathbf{y}} p(\mathbf{y}) N(\mu_{\mathbf{y}}, \sigma_0^2)). \quad (2.31)$$

To calculate the variance of the mixture of normal on the right-hand side of (2.31), we can use

$$\begin{aligned} & \text{var}(\sum_{all \mathbf{y}} p(\mathbf{y}) N(\mu_{\mathbf{y}}, \sigma_0^2)) \\ = & E(\text{var}(\theta|\mathbf{y})) + \text{var}(E(\theta|\mathbf{y})) \\ = & E(\sigma_0^2) + \text{var}(\mu_{\mathbf{y}}) \\ = & \sigma_0^2 + \text{Var}(\mu_0 + \sigma_0^2(T(\mathbf{y}) - T(\mathbf{y}_0))) \\ = & \sigma_0^2 + \sigma_0^4 \text{var}(T(\mathbf{y})). \end{aligned} \quad (2.32)$$

Now we get the relationship

$$\sigma^2 = \sigma_0^2 + \sigma_0^4 \text{var}(T(\mathbf{y})). \quad (2.33)$$

Solving (2.33) for σ_0^2 , we will get

$$\sigma_0^2 = \frac{\sqrt{1 + 4\sigma^2 \text{var}(T(\mathbf{y}))} - 1}{2 \text{var}(T(\mathbf{y}))}. \quad (2.34)$$

With Equation (2.33), one can estimate the population variance σ^2 for the latent trait once the conditional variance σ_0^2 is estimated from the LLLA model.

Mixture of conditional normals: demonstrations by simulated data. To demonstrate how the mixture of conditional normal distributions (2.29) in the LLLA model are used to approximate the population latent trait distribution, I use several simulated data from the Rasch model with different specified latent trait distributions.

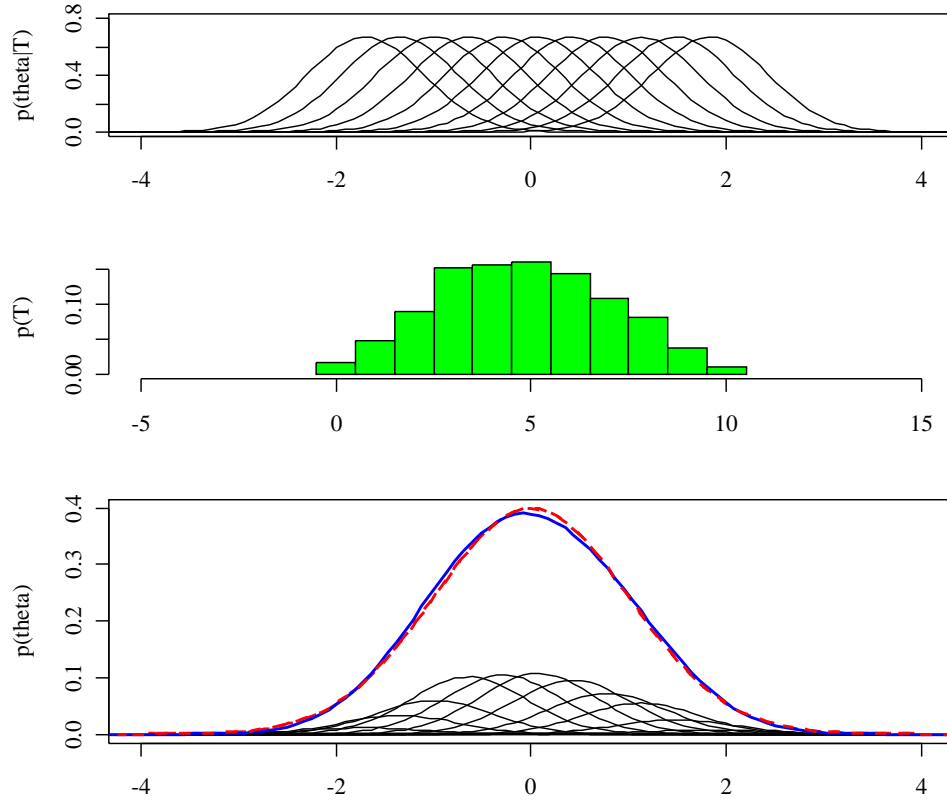


Figure 6. Population latent trait distributions, standard normal, 10 items.

A normal latent trait distribution. In the first simulation, the ability of 1000 persons are simulated from the standard normal distribution (as assumed in the MML Rasch

model),

$$\theta_p \sim N(0, 1), \quad p = 1, \dots, 1000, \quad (2.35)$$

and a 10-item test is simulated with difficulty generated from the standard normal distribution, $b_i \sim N(0, 1)$, $i = 1, \dots, 10$. The 1000×10 response data matrix is then generated from the Rasch model.

The persons are grouped according to their total scores into 11 groups, and the histogram of the total scores is shown in the middle panel of Figure 6. The persons in the same total-score group would have the same conditional latent trait distribution $p(\theta|\mathbf{y}) = p(\theta|\mathbf{T}) = N(\sigma_0^2 T, \sigma_0^2)$, and σ_0^2 is estimated by fitting the LLLA model to the data. The upper panel of Figure 6 shows the conditional distributions $p(\theta|T)$ of the 11 groups, represented by the 11 normal curves with the same variance $\hat{\sigma}_0^2$ and evenly spaced means $\hat{\mu}_T = \hat{\sigma}_0^2 T$. In the lower panel of Figure 6, the conditional latent distribution for each group of the upper panel is weighted by the corresponding group probability from the middle panel, resulting in $p(\theta|T)p(T)$ as shown by the thin solid curves. These weighted conditional distributions are summed up to produce the mixture-of-normal latent trait distribution $\hat{p}(\theta)$ as assumed by the LLLA model, which is shown by the thick solid curve. The true latent trait distribution $p(\theta)$ from which the data are simulated is the standard normal distribution, as shown by the thick dashed curve in the lower panel of Figure 6. We can see that the estimated latent trait distribution (the mixture of 11 normals) produced by the LLLA model is very close to the true latent trait distribution, so the mixture-of-normal distribution is a good approximation of the true latent distribution in this data set.

A bi-modal normal mixture latent trait distribution. In the second simulation, we deviate from using a homogeneous latent trait distribution as used in the previous simulation, and generate a heterogeneous population of persons. The ability of 1000 persons

are simulated from a mixture of two normal distributions

$$\theta_p \sim \frac{1}{2}N(-1, 1) + \frac{1}{2}N(1, 1), \quad p = 1, \dots, 1000. \quad (2.36)$$

The true latent trait distribution $p(\theta)$ is shown as the bimodal thick dashed curve in the lower panel of Figure 7, and the two normal components are also shown as the two thin solid curves.

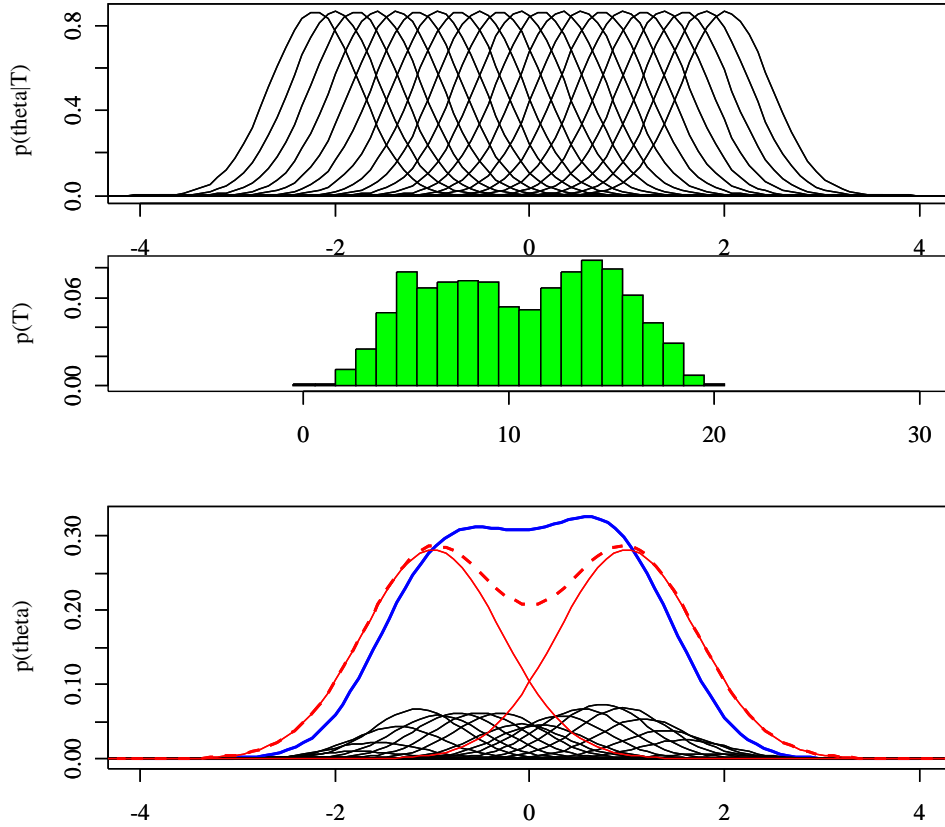


Figure 7. Population latent trait distributions, two-component mixture of normal, 10 items.

The same 10-item test with difficulty generated from the standard normal distribution, $b_i \sim N(0, 1)$, $i = 1, \dots, 10$, as in the previous simulation is used, and the response data matrix is generated from the Rasch model. While the conditional distributions $p(\theta|T)$ are similar to those in the previous simulation (compare upper panels of Figure 6 and Figure 7), the distribution of the total scores $p(T)$ now is a different distribution with a bi-modal

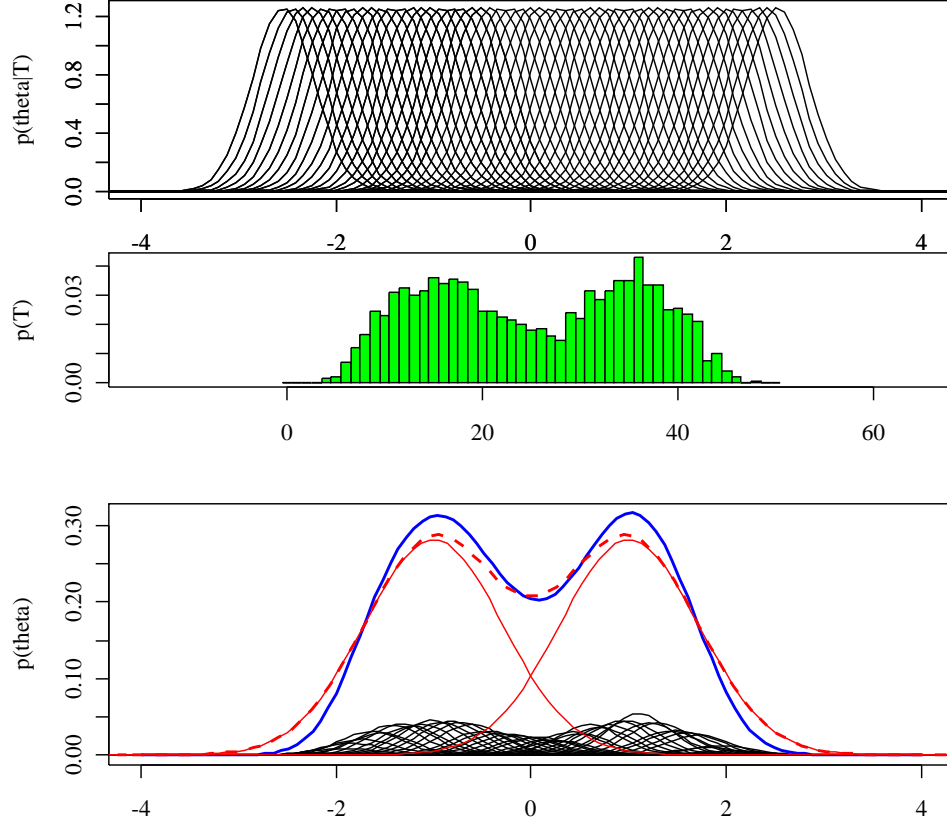


Figure 8. Population latent trait distributions, two-component mixture of normal, 50 items.

feature (compare middle panels of Figure 6 and Figure 7). The resulting mixture-of-normal latent trait distribution $\hat{p}(\theta)$ produced by fitting the LLLA model (Figure 7, lower panel, thick solid curve) also has a bi-modal feature. Although the approximation of $\hat{p}(\theta)$ to $p(\theta)$ in this data set is not as good as shown in the previous simulation, it is better than using a normal distribution, as assumed in the MML Rasch model, to approximate the true latent trait distribution.

If we increase the number of test items to 50, with difficulty generated from the standard normal distribution, the results are shown in Figure 8. The persons are now grouped into 51 groups according to the total scores, and the shape of $p(T)$ has a better resemblance of the latent trait distribution than in the 10-item test. The conditional distribution $p(\theta|\mathbf{y})$ has smaller variance σ_0^2 than in the 10-item test. The resulting mixture-of-normal latent trait distribution $\hat{p}(\theta)$ produced by fitting the LLLA model (Figure 8, lower panel, thick

solid curve) is now a much better approximation to the true latent trait distribution (Figure 8, lower panel, thick dashed curve).

As demonstrated by the simulation studies, the LLLA model has a greater flexibility in describing the latent trait distribution than the MML Rasch model. When the assumptions of the MML Rasch model are true (i.e., the latent trait follows a normal distribution), the LLLA model will produce a latent trait distribution very close to the true distribution; in this case we expect the LLLA model will have a performance nearly as good as the MML Rasch model. When the assumption of the MML Rasch model on the latent trait distribution is violated and the latent trait follows a distribution other than a normal distribution, the LLLA model will have a better approximation on the latent trait distribution than the MML Rasch model.

Chapter 3

Pseudolikelihood Estimation

It is known that maximum likelihood estimation can be computationally prohibitive for many models. LLLA models for data sets with large numbers of items fall into this category. Pseudolikelihood estimation (PLE) is a computationally efficient alternative estimation procedure to MLE. The idea of pseudolikelihood estimation was originated by Besag (1975) for spatial data analysis. Arnold and Strauss (1991) proved the consistency and asymptotic normality of the pseudolikelihood estimator, and it was pointed out that PLE will be less efficient than MLE (see also Geys, Molenberghs, and Ryan (2002a)). Strauss (1992) demonstrated how to use logistic regression procedures to maximize the pseudolikelihood function for different models, and interestingly, the Rasch model was among the examples he discussed. Zwinderman (1995) proposed a pseudolikelihood method for Rasch models based on comparing responses to pairs of items irrespective of other items. Smit and Kelderman (2000) used pseudolikelihood estimation for the Rasch model for dichotomous items and a single latent trait, and their simulation results show the strong similarity between the pseudolikelihood estimates and those from the conditional maximum likelihood estimation. Most recently, the pseudolikelihood estimation of LLLA model in Anderson, Li, and Vermunt (2007) is directly related to my thesis. In that paper, I implemented in R the pseudolikelihood estimation procedures for the LLLA model for dichotomous and polytomous items, with single or many latent traits.

In this section, I will first introduce the maximum likelihood estimation procedure for the LLLA model and point out its limitation in terms of computation. Subsequently, I will give the definition of a pseudolikelihood function and derive the pseudolikelihood function used for the LLLA model. I will show how to maximize the pseudolikelihood function by using logistic regression, so that we can conduct the pseudolikelihood estimation (PLE). Finally I will present ways to get correct standard errors for the PLE.

MLE for LLLA Model

Before we go into to PLE, I will first introduce the maximum likelihood estimation (MLE) for LLLA models. MLE is one of the most popular estimation methods in statistics. For many statistical models, MLE enjoys the property of consistency (converges to true value as sample size increases) and efficiency (has the smallest standard error asymptotically); see, e.g., Lehmann and Casella (1998). When we propose a new estimation method such as the pseudolikelihood estimation procedure, we use MLE as a reference to compare with.

As a subfamily of log-linear models, MLE for LLLA models can be calculated using log-linear or Poisson regression procedures, which are available in almost all standard statistical packages (e.g., R, S-PLUS, SAS, Stata and SPSS) that are capable of fitting generalized linear models. For an introduction to the MLE for log-linear models, including the use of the Newton-Raphson method to numerically calculate the MLE, see Agresti (2002).

Suppose we have N persons respond to I items, and the responses are recorded as an N by I matrix Y with entries y_{pi} , where y_{pi} is the response of the person p to item i . For now let us just consider binary responses, so y_{pi} can only take on values 0 or 1.

The likelihood function is the joint pdf of the data matrix Y . In LLLA model, we assume that persons are iid, so we have

$$L = \prod_{p=1}^N P(y_{p1}, y_{p2}, \dots, y_{pI}). \quad (3.1)$$

On the other hand, every person's response is one of the total 2^I possible response patterns from $(0, 0, \dots, 0)$ to $(1, 1, \dots, 1)$. If we denote the number of persons that produce the response pattern \mathbf{y} as $n(\mathbf{y})$, then the distribution of the 2^I by 1 count vector $\{n(\mathbf{y})\} = [n(0, 0, \dots, 0), \dots, n(1, 1, \dots, 1)]$ follows a multinomial distribution with parameters N and $p(\mathbf{y})$.

The likelihood of the multinomial count data is

$$L = \prod_{\mathbf{y}} p(\mathbf{y})^{n(\mathbf{y})}, \quad (3.2)$$

and the log likelihood is

$$l = \log L = \sum_{\mathbf{y}} n(\mathbf{y}) \log p(\mathbf{y}). \quad (3.3)$$

The likelihood derived from the data matrix Y (3.1) and the likelihood derived from the count vector $\{n(\mathbf{y})\}$ (3.2) are exactly the same. The ML fitting of the LLLA model is essentially the ML fitting of the log-linear model for multinomial data; therefore we can use the procedures for log-linear models (i.e., Poisson regressions) to get the ML estimates for the LLLA model.

To demonstrate how the MLE for an LLLA model is calculated, let's take a look at a simple example for a 4-item test. Suppose N persons take the test and their binary responses ($y_{pi} = 1$ or 0) are collected in a data matrix Y shown in Table 3.

Table 3

Original Response Data Matrix

	Item 1	Item 2	Item 3	Item 4
Person 1	y_{11}	y_{12}	y_{13}	y_{14}
Person 2	y_{21}	y_{22}	y_{23}	y_{24}
Person 3	y_{31}	y_{32}	y_{33}	y_{34}
\vdots	\vdots	\vdots	\vdots	\vdots
Person N	y_{N1}	y_{N2}	y_{N3}	y_{N4}

From the data matrix we can construct the count vector of $2^4 = 16$ response patterns (Table 4).

Now that we have the count \mathbf{n} as response variable, the next step is to get the design matrix that corresponds to the 1-D binary model (Table 5). The length of the count vector

Table 4

Response Data in Count Data Format

Response		Patterns		Count
Item 1	Item 2	Item 3	Item 4	\mathbf{n}
0	0	0	0	n_1
1	0	0	0	n_2
0	1	0	0	n_3
1	1	0	0	n_4
\vdots	\vdots	\vdots	\vdots	\vdots
1	1	1	1	n_{16}
$\sum_{i=1}^{16} n_i = N$				

is 16, and the number of non-redundant parameters is $1 + 4 + 1 = 6$. Therefore the design matrix is a 16 by 6 matrix. The first column of the data matrix corresponds to the intercept and is a vector of 1's. The second to the fifth columns correspond to the main effects of each item. The dummy coding of these vectors are the same as the 0-1 responses to each item because we use response 0 as the reference level. The last column in the design matrix is the interaction term, which is calculated by the sum of the products of pairs of scores ($\nu_{1(y_1)}\nu_{1(y_2)} + \nu_{1(y_1)}\nu_{3(y_3)} + \dots + \nu_{3(y_3)}\nu_{4(y_4)}$); and here we use item raw scores (i.e., the scores are the same as 0-1 responses).

With the response variable the counts \mathbf{n} and the explanatory variables in the form of the design matrix, we can use software for Poisson regression to fit the model and get the MLE for the parameters. In R, the function we used to fit the Poisson regression model is 'glm' with Poisson family and log link function.

Motivation of PLE for LLLA

As we saw in the previous section, the maximum likelihood estimation (MLE) for the LLLA models requires the count vector as response variable, and the length of the

Table 5

Design Matrix for the LLLA Model

Count \mathbf{n}	Design matrix					
	Intercept	item 1	item 2	item 3	item 4	Interaction
	λ	$\lambda_{1(1)}$	$\lambda_{2(1)}$	$\lambda_{2(1)}$	$\lambda_{4(1)}$	σ^2
n_1	1	0	0	0	0	0
n_2	1	1	0	0	0	0
n_3	1	0	1	0	0	0
n_4	1	1	1	0	0	1
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
n_{16}	1	1	1	1	1	6

count vector grows exponentially. For example, for a 4-item test with binary responses, the count vector has $2^4 = 16$ entries, and for a 10-item test, the length of the count vector is $2^{10} = 1024$, and for a 20-item test, $2^{20} = 1,048,576$, and for a 100-item test, $2^{100} = 1.267651 \times 10^{30}$. Therefore the computational costs (time and space) for the ML estimation increase drastically as the number of items increases. For polytomous items, the costs increase even faster. In reality, the MLE of the LLLA model is not practical beyond 20 dichotomous items using a standard personal computer with R or SAS.

To address this problem, pseudolikelihood estimation has been proposed. Instead of maximizing the likelihood function, a pseudolikelihood function is defined and the estimates that maximize the pseudolikelihood are derived. The PL estimates are consistent but less efficient than ML estimates; however, the computational cost decreases a lot with small sacrifice of efficiency.

Introduction to PLE

In maximum likelihood estimation method for LLLA, the basic building block in the likelihood function given in (3.1) is the joint distribution of the items for each person:

$$P(y_{p1}, y_{p2}, \dots, y_{pI}).$$

Pseudolikelihood estimation method replaces the above function by the product of conditional distributions:

$$P(y_{p1}|y_{p2}, \dots, y_{pI})P(y_{p2}|y_{p1}, y_{p3}, \dots, y_{pI}) \dots P(y_{pI}|y_{p1}, \dots, y_{p,I-1}).$$

In this way, a pseudolikelihood function is defined as follows for the response data matrix \mathbf{Y} ,

$$PL(\boldsymbol{\beta}; \mathbf{Y}) = \prod_{p=1}^N \prod_{i=1}^I P(y_{pi}|y_{p(-i)}), \quad (3.4)$$

where $\boldsymbol{\beta}$ is the vector of parameters in the model; and $P(y_{pi}|y_{p(-i)})$ is the conditional probability of person p 's response to item i given the responses to the rest items $y_{p(-i)} = (y_{p1}, \dots, y_{p,i-1}, y_{p,i+1}, \dots, y_{pI})$.

Similarly, we can define the log-pseudolikelihood as the logarithm of the pseudolikelihood,

$$pl = \log PL(\boldsymbol{\beta}; \mathbf{Y}) = \sum_{p=1}^N \sum_{i=1}^I \log P(y_{pi}|y_{p(-i)}). \quad (3.5)$$

Analogous to MLE, the pseudolikelihood estimator (PLE) is defined as the $\boldsymbol{\beta}$ that maximizes the pseudolikelihood function (3.4), which can be obtained by setting the first derivative of the log-pseudolikelihood to 0.

It was proved by Arnold and Strauss (1991) that under some regularity conditions, there is at least one solution to the pseudolikelihood equations that is consistent and asymptotically normal. This result means that it is valid to use a pseudolikelihood estimator

because it is consistent, and by the asymptotic normality we can construct the asymptotic confidence intervals for the estimator.

Application of PLE to LLLA

Now we apply pseudolikelihood estimation to the LLLA model. For each person, we can write the pseudolikelihood function as the product of conditional probabilities of item responses given the rest of the items. As an example, suppose we have 4 items, (for simplicity, let's ignore the subscript p in y_{pi} for now), so that we have:

$$\begin{aligned}
PL_p &= P(Y_1 = y_1 | Y_2 = y_2, Y_3 = y_3, Y_4 = y_4) \\
&\quad \times P(Y_2 = y_2 | Y_1 = y_1, Y_3 = y_3, Y_4 = y_4) \\
&\quad \times P(Y_3 = y_3 | Y_1 = y_1, Y_2 = y_2, Y_4 = y_4) \\
&\quad \times P(Y_4 = y_4 | Y_1 = y_1, Y_2 = y_2, Y_3 = y_3) \\
&= P(y_1 | y_2, y_3, y_4) P(y_2 | y_1, y_3, y_4) P(y_3 | y_1, y_2, y_4) P(y_4 | y_1, y_2, y_3) \\
&= P(y_1 | y_{-1}) P(y_2 | y_{-2}) P(y_3 | y_{-3}) P(y_4 | y_{-4}), .
\end{aligned} \tag{3.6}$$

For illustration, consider the first conditional probability in the above equation, and note that

$$P(Y_1 = y_1 | y_2, y_3, y_4) = \frac{P(y_1, y_2, y_3, y_4)}{P(y_2, y_3, y_4)}. \tag{3.7}$$

The joint probability in the numerator is given by the LLLA model, resulting in

$$\begin{aligned}
P(y_1, y_2, y_3, y_4) &= \exp[\lambda + \lambda_{1(y_1)} + \lambda_{2(y_2)} + \lambda_{3(y_3)} + \lambda_{4(y_4)} + \\
&\quad \sigma^2(\nu_{1(y_1)}\nu_{2(y_2)} + \nu_{1(y_1)}\nu_{3(y_3)} + \dots + \nu_{3(y_3)}\nu_{4(y_4)})].
\end{aligned} \tag{3.8}$$

The marginal probability in the denominator can then be calculated as

$$\begin{aligned}
P(y_2, y_3, y_4) &= \sum_{y_1} P(y_1, y_2, y_3, y_4) = \\
&= \sum_{y_1=0}^1 \exp[\lambda + \lambda_{1(y_1)} + \lambda_{2(y_2)} + \lambda_{3(y_3)} + \lambda_{4(y_4)} + \\
&\quad \sigma^2(\nu_{1(y_1)}\nu_{2(y_2)} + \nu_{1(y_1)}\nu_{3(y_3)} + \dots + \nu_{3(y_3)}\nu_{4(y_4)})] \\
&= \exp[\lambda + \lambda_{2(y_2)} + \lambda_{3(y_3)} + \lambda_{4(y_4)} + \\
&\quad \sigma^2(\nu_{2(y_2)}\nu_{3(y_3)} + \nu_{2(y_2)}\nu_{4(y_4)} + \nu_{3(y_3)}\nu_{4(y_4)})] \\
&\quad \sum_{y_1=0}^1 \exp[\lambda_{1(y_1)} + \sigma^2(\nu_{1(y_1)}\nu_{2(y_2)} + \nu_{1(y_1)}\nu_{3(y_3)} + \nu_{1(y_1)}\nu_{4(y_4)})]. \tag{3.9}
\end{aligned}$$

Substituting (3.8) and (3.9) into (3.7), the terms that do not contain y_1 in the numerator and denominator will cancel out, and we will have

$$P(y_1|y_2, y_3, y_4) = \frac{\exp[\lambda_{1(y_1)} + \sigma^2\nu_{1(y_1)}(\nu_{2(y_2)} + \nu_{3(y_3)} + \nu_{4(y_4)})]}{\sum_{y_1=0}^1 \exp[\lambda_{1(y_1)} + \sigma^2\nu_{1(y_1)}(\nu_{2(y_2)} + \nu_{3(y_3)} + \nu_{4(y_4)})]}. \tag{3.10}$$

Thus we have the conditional probability in terms of LLLA model parameters.

Actually there is a quicker way to write down the same conditional probability (3.10) without explicitly writing down the marginal probability $P(y_2, y_3, y_4)$. Notice that for the conditional probability $P(y_1|y_2, y_3, y_4)$ what we are interested in is how it changes as a function of y_1 , and only the joint probability in the numerator of (3.7) is a function of y_1 . Therefore, we only need the terms that contains y_1 in the joint probability (3.8), which allows us to write

$$\begin{aligned}
&P(Y_1 = y_1|y_2, y_3, y_4) \\
&\propto P(y_1, y_2, y_3, y_4) \\
&\propto \exp[\lambda_{1(y_1)} + \sigma^2\nu_{1(y_1)}(\nu_{2(y_2)} + \nu_{3(y_3)} + \nu_{4(y_4)})]. \tag{3.11}
\end{aligned}$$

The next step is to calculate the normalizing constant of the probability function, so that the sum of the probability over all possible outcomes of y_1 (i.e., 0 and 1) is one. We get the same answer as in (3.10).

Similarly, we can write down $P(Y_2 = y_2|y_1, y_3, y_4)$, $P(Y_3 = y_3|y_1, y_2, y_4)$, and $P(Y_4 = y_4|y_1, y_2, y_3)$, and they all have the same form as in (3.10). Substituting them into (3.6), we get the pseudolikelihood function for a person. The pseudolikelihood function for the data set is given by the product of pseudolikelihood function for each person: $PL = \prod_{p=1}^N PL_p$.

Now I will show that the conditional probability we have derived has the same form as a logistic regression model. This fact will justify using the MLE estimation of logistic regression procedure as a means to maximize the pseudolikelihood function, which saves the effort of implementing the optimization procedure for finding the maximum of pseudolikelihood from scratch.

Let us write down the conditional probability of $Y_1 = 1$ given the responses of the rest of the items, by substituting $y_1 = 1$ into (3.10), and then dividing both the numerator and denominator by $P(Y_1 = 0|y_2, y_3, y_4)$:

$$\begin{aligned}
& P(Y_1 = 1|y_2, y_3, y_4) \\
&= \frac{P(Y_1 = 1|y_2, y_3, y_4)}{P(Y_1 = 0|y_2, y_3, y_4) + P(Y_1 = 1|y_2, y_3, y_4)} \\
&= \frac{\exp[(\lambda_{1(1)} - \lambda_{1(0)}) + \sigma^2(\nu_{1(1)} - \nu_{1(0)})(\nu_{2(y_2)} + \nu_{3(y_3)} + \nu_{4(y_4)})]}{1 + \exp[(\lambda_{1(1)} - \lambda_{1(0)}) + \sigma^2(\nu_{1(1)} - \nu_{1(0)})(\nu_{2(y_2)} + \nu_{3(y_3)} + \nu_{4(y_4)})]} .
\end{aligned} \tag{3.12}$$

To see more clearly that the above equation has the same form as a logistic regression, we can further transform the left-hand side of (3.12) into the logit function:

$$\begin{aligned}
& \log \frac{P(Y_1 = 1|y_2, y_3, y_4)}{P(Y_1 = 0|y_2, y_3, y_4)} \\
&= (\lambda_{1(1)} - \lambda_{1(0)}) + \sigma^2(\nu_{1(1)} - \nu_{1(0)})(\nu_{2(y_2)} + \nu_{3(y_3)} + \nu_{4(y_4)}) .
\end{aligned} \tag{3.13}$$

Now the right-hand side of the equation is a linear combination of main effect of item 1 and the effect of rest-score for item 1. The main effect is $(\lambda_{1(1)} - \lambda_{1(0)})$; the rest-score is $(\nu_{2(y_2)} + \nu_{3(y_3)} + \nu_{4(y_4)})$; $(\nu_{1(1)} - \nu_{1(0)})$ is a fixed number since the scores are known; and σ^2 is the unknown coefficient of the rest-score.

Recall that to make the model identifiable we always have the constraint for the parameters for the response 0: $\lambda_{i(0)} = 0$ and $\nu_{i(0)} = 0$. Denote $\lambda_i = \lambda_{i(1)} - \lambda_{i(0)}$, and $\nu_i = \nu_{i(1)} - \nu_{i(0)}$ to remove the redundant parameters in the model.

Now we can extend from what we obtained in the 4-item example to the general I -item test. Generally, for a test with I items, denote $y_{-i} = (y_1, \dots, y_{i-1}, y_{i+1}, \dots, y_I)$, the response vector without entry y_i . The conditional distribution of an item response given the rest item responses is given by

$$P(Y_i = 1|y_{-i}) = \frac{\exp[\lambda_i + \sigma^2 \nu_i \sum_{l \neq i} \nu_{l(y_l)}]}{1 + \exp[\lambda_i + \sigma^2 \nu_i \sum_{l \neq i} \nu_{l(y_l)}]} \quad (3.14)$$

and

$$P(Y_i = 0|y_{-i}) = \frac{1}{1 + \exp[\lambda_i + \sigma^2 \nu_i \sum_{l \neq i} \nu_{l(y_l)}]} \quad (3.15)$$

We can write the above two equation into one equation

$$P(Y_i = y_i|y_{-i}) = \frac{\exp[y_i(\lambda_i + \sigma^2 \nu_i \sum_{l \neq i} \nu_{l(y_l)})]}{1 + \exp[\lambda_i + \sigma^2 \nu_i \sum_{l \neq i} \nu_{l(y_l)}]} \quad (3.16)$$

Now for the data matrix $Y = y_{pi}, p = 1, \dots, N, i = 1, \dots, I$, the pseudolikelihood function is

$$PL = \prod_{p=1}^N \prod_{i=1}^I P(y_{pi}|y_{p(-i)}) \quad (3.17)$$

Substituting (3.16) into (3.17), we get

$$\begin{aligned}
PL &= \prod_{p=1}^N \prod_{i=1}^I P(y_{pi}|y_{p(-i)}) \\
&= \prod_{p=1}^N \prod_{i=1}^I \left[\frac{\exp[y_{pi}(\lambda_i + \sigma^2 \nu_i \sum_{l \neq i} \nu_{l(y_{pl})})]}{1 + \exp[\lambda_i + \sigma^2 \nu_i \sum_{l \neq i} \nu_{l(y_{pl})}]} \right]. \tag{3.18}
\end{aligned}$$

Next step is to find the parameters $(\lambda_1, \lambda_2, \dots, \lambda_I, \sigma^2)$ that maximize the pseudolikelihood function PL .

As was discussed before in this section, the pseudolikelihood function in (3.18) has the same form of the likelihood function of a logistic regression model. In this logistic regression model, there are NI observations; λ_i are the main effects that are associated with dummy-coded item indicators; and σ^2 is the effect of rest total scores $r_{pi} = \nu_i \sum_{l \neq i} \nu_{l(y_{pl})}$. Therefore, maximizing PL is the same as maximizing the likelihood function in the logistic regression problem. This motivates us to use logistic regression procedures to get the PLE, as described in next section.

Maximize PL Using Logistic Regression

The similarity of the conditional distribution function and the logistic regression distribution allows us to use logistic regression procedures to obtain the pseudolikelihood estimates by constructing a new data matrix for the equivalent logistic regression problem, then applying a logistic regression procedure to maximize the likelihood.

We can construct the new data matrix with the following variables:

- Response $Y = \{y_{pi}\}$ is an NI by 1 vector.
- Item variable $L = \{l_{pi}\}$, with $l_{pi} = i$.
- Rest total score $R = \{r_{pi}\}$, with $r_{pi} = \nu_i \sum_{l \neq i} \nu_{l(y_{pl})}$. Usually we use scores $\nu_{i(1)} = \nu_i = 1$, and $\nu_{i(0)} = 0$.

We call this new NI by 3 data matrix “stacked” data (Table 6).

Table 6

Stacked Data in Pseudolikelihood Estimation With Logistic Regression Procedures

Person	Item	Response	Rest Total
1	1	y_{11}	r_{11}
1	2	y_{12}	r_{12}
1	3	y_{13}	r_{13}
1	4	y_{14}	r_{14}
2	1	y_{21}	r_{21}
2	2	y_{22}	r_{22}
2	3	y_{23}	r_{23}
2	4	y_{24}	r_{24}
\vdots	\vdots	\vdots	\vdots
N	1	y_{N1}	r_{N1}
N	2	y_{N2}	r_{N2}
N	3	y_{N3}	r_{N3}
N	4	y_{N4}	r_{N4}

The maximization of the pseudo likelihood function in (3.17) is equivalent to the maximization of the likelihood function of a logistic regression model on the stacked data, with Y as response variable, and L and R as explanatory variables.

To fit the logistic regression in R, I use function “glm” with “binomial” family option and “logit” link option. The formula of the model is

$$\text{Response} \sim \text{factor}(\text{Item}) + \text{RestTotal} - 1$$

Here **Response** is the 0-1 response column; **factor(Item)** indicates that column **Item** is treated as a factor or categorical variable; **-1** indicates that there is no intercept term in the model.

Correct Standard Errors

The logistic regression procedure gives us the estimates of the parameters that maximize the pseudolikelihood function. Unfortunately standard errors reported by the logistic regression procedure are not correct because the rows of the stacked data matrix are dependent, rather than independent. Note that the stack data matrix (Table 6) is constructed such that each person has I rows in the stacked data set, and these I observations are dependent because they are from the same person. Between persons, different observations (rows) in the stacked data are independent, but within a person, the observations (rows) are dependent.

Since the observed responses from a person are likely to be positively correlated, the estimated standard errors assuming independent observations will be too small. The standard errors reported by the standard MLE of a logistic regression model will underestimate the true standard errors for the PLE.

To obtain the correct standard errors, we can use following choices of methods: jackknife, bootstrap, and sandwich estimator.

Jackknife and bootstrap (Efron & Tibshirani, 1997; Shao & Tu, 1995) estimate standard errors by perturbing the original data set many times, and using the parameter estimates from those perturbed data sets to estimate the standard errors. Jackknife and Bootstrap use different ways to perturb the original data set. Jackknife generates multiple data sets by “leave-one-observation-out” of the original data. Bootstrap generates multiple data sets by “sample-with-replacement” from the original data.

Jackknife. Suppose there are N observations in the original data set. In the Jackknife method, a perturbed data set is constructed by removing one observation from the original data set. The estimation procedure is applied to the perturbed data set of size $N - 1$ to get the parameter estimates. Repeating the procedure by letting each observation take a turn as the left-out observation results in N sets of parameter estimates $\hat{\beta}^{(1)}, \hat{\beta}^{(2)}, \dots, \hat{\beta}^{(N)}$.

The jackknife estimate of standard error is given by

$$SE_{jack} = \sqrt{N-1} \text{sd}(\hat{\beta}^{(1)}, \hat{\beta}^{(2)}, \dots, \hat{\beta}^{(N)}) \quad (3.19)$$

where $\text{sd}(\hat{\beta}^{(1)}, \hat{\beta}^{(2)}, \dots, \hat{\beta}^{(N)})$ is the standard deviation of the parameter estimates:

$$\sqrt{\frac{1}{N} \sum_{j=1}^N (\hat{\beta}^{(j)} - \bar{\hat{\beta}}^{(\cdot)})^2}.$$

In the LLLA model, the vector responses given by one person, $\mathbf{y}_p = (y_{p1}, \dots, y_{pI})$ is considered as a single observation, that is, a row in the original $N \times I$ data matrix of item responses. To obtain the jackknife estimate of standard errors, we can construct the perturbed data set by removing one person from the original data set, and apply the PLE procedure to the $(N-1) \times I$ perturbed data set, which includes the steps of constructing the “stacked” data set of $(N-1)I$ rows and applying logistic regression to maximize the pseudolikelihood function. For each parameter, we can obtain the jackknife estimate of standard error of PLE by using formula (3.19).

Bootstrap. In the bootstrap method, a perturbed data set that is called a bootstrap sample is constructed by sampling with replacement N data points from the original data set. Repeat the bootstrap sample many times (B times). By applying the estimation procedure to each of the B bootstrap sample, B sets of parameter estimates are obtained: $\hat{\beta}^{(1)}, \hat{\beta}^{(2)}, \dots, \hat{\beta}^{(B)}$. The bootstrap estimate of the standard error is given by

$$SE_{boot} = \text{sd}(\hat{\beta}^{(1)}, \hat{\beta}^{(2)}, \dots, \hat{\beta}^{(B)}) = \sqrt{\frac{1}{B} \sum_{j=1}^B (\hat{\beta}^{(j)} - \bar{\hat{\beta}}^{(\cdot)})^2}. \quad (3.20)$$

In an LLLA model, we can construct a bootstrap sample by sampling N persons (that is, N rows) with replacement from the original data matrix. By treating the response vector of a person as a single observation, the dependency structure of the original data set

(being dependent within a person, and independent between persons) is preserved in the bootstrap sample. Each bootstrap sample is an $N \times I$ data matrix that has the same size as the original data matrix. We can calculate the PLE for these bootstrap samples and use (3.20) to get the bootstrap estimate of SE for each parameter estimated by PLE.

Robust estimator. Jackknife and bootstrap are computationally intensive methods, they repeat the same procedures on many perturbed data sets. It is possible to correct the standard errors by using the robust or “sandwich” estimator (White, 1982) for the standard errors, which does not require fitting models to the resampled data. Therefore the sandwich estimator is computationally more efficient than either the jackknife or bootstrap. Geys et al. (2002a) discussed about the “sandwich” estimator in the context of the pseudolikelihood estimation, and the form of the “sandwich” estimator for PLE was first proved by Arnold and Strauss (1991). In this research, the sandwich estimator is implemented as a way to obtain adjusted or correct SE for the pseudolikelihood estimators.

The sandwich estimator is a robust method to estimate the covariance matrix of the maximum likelihood estimates. According to the theory of MLE, the covariance matrix for the MLE is the inverse of the Fisher information matrix. There are two ways to define the Fisher information matrix, one is that the Fisher information is the covariance matrix of the gradient (i.e., the vector of the first derivatives of the log-likelihood function with respect to the parameter):

$$I_n(\boldsymbol{\beta}) = Cov\left(\frac{\partial l_n}{\partial \boldsymbol{\beta}}\right) = nK_1(\boldsymbol{\beta}), \quad (3.21)$$

where $K_1(\boldsymbol{\beta}) = Cov(\frac{\partial l_1}{\partial \boldsymbol{\beta}})$ is the contribution of one observation in the data set to the Fisher information.

The other definition of Fisher information is the negative of the expected value of the Hessian (i.e., the matrix of the second derivatives of the log-likelihood function):

$$I_n(\boldsymbol{\beta}) = E\left(-\frac{\partial^2 l_n}{\partial \boldsymbol{\beta}^2}\right) = \left\{E\left(-\frac{\partial^2 l_n}{\partial \beta_j \partial \beta_k}\right)\right\} = nJ_1(\boldsymbol{\beta}) \quad (3.22)$$

where $J_1(\boldsymbol{\beta}) = E(-\frac{\partial^2 l_1}{\partial \boldsymbol{\beta}^2})$ is the contribution of one observation. If the data are generated from the specified model, the two definitions of the Fisher information (3.21) and (3.22) are equivalent: $I_1(\boldsymbol{\beta}) = J_1(\boldsymbol{\beta}) = K_1(\boldsymbol{\beta})$.

Usually, the estimated Fisher information is given by the inverse of the observed Fisher information matrix, which is the negative Hessian with plugged-in ML estimates, i.e., $\widehat{\text{cov}}(\hat{\boldsymbol{\beta}}) = \hat{J}_n^{-1}(\hat{\boldsymbol{\beta}}) = -H^{-1}(\hat{\boldsymbol{\beta}})$. The matrix $\hat{J}_n^{-1}(\hat{\boldsymbol{\beta}})$ is given by the standard MLE procedures.

If the model is misspecified, then the two matrices $J_1(\boldsymbol{\beta})$ and $K_1(\boldsymbol{\beta})$ will not agree, the covariance of the MLE under the misspecified model is given by the sandwich estimator

$$\widehat{\text{cov}}(\hat{\boldsymbol{\beta}}) = \hat{J}_n^{-1}(\hat{\boldsymbol{\beta}}) \hat{K}_n(\hat{\boldsymbol{\beta}}) \hat{J}_n^{-1}(\hat{\boldsymbol{\beta}}). \quad (3.23)$$

where the matrix $\hat{K}_n(\hat{\boldsymbol{\beta}})$ is given by

$$\hat{K}_n(\hat{\boldsymbol{\beta}}) = \sum_{i=1}^n g_i(\boldsymbol{\beta}) g_i(\boldsymbol{\beta})', \quad (3.24)$$

where $g_i(\boldsymbol{\beta}) = \frac{\partial l_i}{\partial \boldsymbol{\beta}}$ is the contribution of the i -th observation to the gradient vector. To obtain the sandwich estimator, we have to use (3.24) to calculate $\hat{K}_n(\hat{\boldsymbol{\beta}})$ and use it to correct the covariance matrix given by standard procedures.

Chapter 4

LLLA With Covariates

In studies involving item response data, it is often desirable to address the effect of covariates on the responses of the person to the items. As an item response model involves two entities, persons and items, the covariates can be classified as item covariates and person covariates. An item covariate describes a property of the associated item and a person covariate relates to the person's characteristics. Generally speaking, the inclusion of covariates in IRT models improves the explanatory power of the models and measurement precision (De Boeck & Wilson, 2004).

There has been extensive research in the IRT field to include covariates in IRT models. De Boeck and Wilson (2004) present a summary of extensions of various Rasch models that incorporate covariates. Although many instances and proposals have been made to incorporate collateral information into LMA models, (e.g. Anderson and Böckenholt (2000); Tettegah and Anderson (2007); Anderson et al. (2010); Goodman (1974); Clogg (1981); Böckenholt and Böcknholt (1990); Gilula and Haberman (1994, 1995); Wong (1995)), comparatively, these extensions with covariates in LMA/LLLA are not used as item response models. In Anderson and Vermunt (2000), there is an example of an LMA model that takes person's gender into consideration. In Tettegah and Anderson (2007), an LMA model with two person covariates is proposed and applied to the analysis of coded text data.

The equivalence of Rasch models with LLLA models provides a framework or a tool to systematically develop the LLLA models with covariates for use as IRT models. We can start with an extended Rasch model with covariates and find its manifest probability, which leads to the equivalent LLLA model with covariates extension.

Once we have the LLLA models with covariates, we provide a method and tool to fit the corresponding IRT-with-covariates models.

IRT Models With Covariates

Let us review what has been done in adding covariates to Rasch models.

Item covariates. With respect to adding item covariates, the linear logistic test model (LLTM) was proposed as an extension of the Rasch model (Fischer, 1973). Suppose that the item difficulty parameter for the i th item in a Rasch model is b_i , and the item predictors associated with the properties of the i th item are $X_{ik}, k = 1, \dots, K$. In the LLTM model, we write the item difficulty as a linear combination of the item predictors,

$$b_i = \sum_{k=0}^K \beta_k X_{ik}, \quad (4.1)$$

where β_0 is the intercept, and $X_{i0} = 1$.

Note that in the original Rasch model, there are I item parameters (b_1, b_2, \dots, b_I) , and the parameter b_i is associated with item i . In the LLTM model, there are $K + 1$ item parameters $(\beta_0, \beta_1, \beta_2, \dots, \beta_K)$, and the parameter β_k is associated with the k -th item covariate. The number of item covariate K cannot exceed the number of items I , otherwise the model will not be identifiable.

Since $K \leq I$, we have a smaller number of parameters in a LLTM model than in a Rasch model. In fact, the Rasch model is a special case under LLTM model with $K = I$, and there is no intercept, and X_{ik} is the item indicator variable: $X_{ik} = 1$ if $k = i$ and $X_{ik} = 0$ if $k \neq i$.

Person covariates. With respect to adding person covariates, the Latent Regression Rasch Model has been proposed (Zwinderman, 1991). Denoting the person covariates as $\mathbf{Z} = (Z_1, \dots, Z_K)'$, a latent regression Rasch model specifies the relationship between the latent trait and person covariates as a linear regression model:

$$\theta | \mathbf{Z} = \mathbf{Z}'\boldsymbol{\gamma} + \epsilon, \quad \text{where } \epsilon \sim N(0, \sigma_\epsilon^2), \quad (4.2)$$

where $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_K)'$ are the effects of the person covariates; ϵ is the residual for the latent trait that is not explained by the person covariates. The conditional probability $p(Y_i|\theta)$ follows a Rasch model. The inclusion of person covariates in the Rasch model serves two purposes. The first purpose is explanatory power in that it helps to answer questions regarding how these person covariates relate to the latent traits. For example, we can use (4.2) to study how reading ability (the latent trait) is related to gender (the person covariate). The second purpose is the measurement, that is, by considering these covariates, we will have better prediction of the person's ability θ_p .

LLLA With Item Covariates

Adding item covariates in an LLLA model is similar to adding item covariates in the Rasch model as described in the previous section. The relationship between the item covariates and the item parameters is specified as a linear model, and this linear model is incorporated into the LLLA model.

Suppose we have K item covariates that are associated with item i , and denote them by a vector $\mathbf{X}_i = (X_{i1}, \dots, X_{iK})'$. The item parameters in the LLLA model are $\lambda_{i(y_i)}$, which correspond to item easiness parameters.

The LLLA model with item covariates (LLLAi) can be treated as a (two-part) model that consists of a base model and a submodel. The base model is the LLLA model

$$\text{Base model : } \log p(\mathbf{Y}) = \lambda + \sum_i \lambda_{i(y_i)} + \sigma^2 \sum_{i \neq j} \nu_{i(y_i)} \nu_{j(y_j)}. \quad (4.3)$$

The submodel specifies the relationship between the item parameters and the item covariates.

In the submodel, we can model the item parameters as linear combinations of the

item covariates, as was done in the LLTM model,

$$\text{Submodel : } \quad \lambda_{i(y_i)} = \sum_{k=0}^K \beta_{k(y_i)} X_{ik}. \quad (4.4)$$

Combining the submodel (4.3) with the base model (4.4) yields the LLLA model with item covariates in a single equation

$$\log P(\mathbf{Y} = \mathbf{y} | \mathbf{X}) = \lambda + \sum_i \sum_k X_{ik} \beta_{k(y_i)} + \sigma^2 \sum_{i \neq i'} \nu_{i(y_i)} \nu_{i'(y_{i'})}. \quad (4.5)$$

Since each item parameter $\lambda_{i(y_i)}$ is a linear function of K ($K < I$) item covariates, the number of item parameters is reduced from I to K parameters.

To prove the equivalence of the LLTM with the proposed LLLA model with item covariates is rather straightforward because the additional submodel (4.3) does not involve any random effects that need to be integrated out. The exact relationship between the parameters in the LLLA model with item covariates and the parameters in the LLTM is given by

$$\lambda = \log p(\mathbf{0}), \quad (4.6)$$

$$\beta_{0(1)} = \beta_0 + \mu_0 + \frac{1}{2} \sigma_0^2, \quad (4.7)$$

$$\beta_{k(1)} = \beta_k, \quad k = 1, \dots, K, \quad (4.8)$$

$$\sigma^2 = \sigma_0^2, \quad (4.9)$$

$$\nu_{i(y_i)} = y_i. \quad (4.10)$$

LLLA With Person Covariates

In this section, after stating the general form of the LLLA model with person covariates (LLLAp), I will present how to derive this model from the latent regression Rasch model by taking a similar approach used to derive the LLLA model from the Rasch model.

Model form. The LLLA model with person covariates \mathbf{Z} is given by

$$\log P(\mathbf{Y} = \mathbf{y}|\mathbf{Z}) = \lambda(\mathbf{Z}) + \sum_i \lambda_{i(y_i)} + \mathbf{Z}'\boldsymbol{\gamma} \sum_i \nu_{i(y_i)} + \sigma_{0\epsilon}^2 \sum_{i \neq i'} \nu_{i(y_i)} \nu_{i'(y'_i)}. \quad (4.11)$$

Comparing (4.11) with the LLLA model without any covariates, there are two differences: (a) there is an addition term $\mathbf{Z}'\boldsymbol{\gamma} \sum_i \nu_{i(y_i)}$ that is the linear combination of the person covariates multiplied by the total score; (b) The intercept is a function of \mathbf{Z} . The reason for these differences is that the model needs to satisfy the sum-to-one constraint on the conditional distribution $\sum_{(y_1, \dots, y_I)} p(\mathbf{Y} = (y_1, \dots, y_I) | \mathbf{Z}) = 1$.

Derivation of LLLA with person covariates model. Now I show how to derive the LLLAp model (4.11) starting from a latent regression Rasch model. For simplicity, let's consider only the 1-D model with binary items.

Let us assume that the reference response is the response pattern with all 0's, $\mathbf{y}_0 = \mathbf{0}$. Without considering the person covariates \mathbf{Z} , if we assume that
(A) the responses follow a Rasch model, with item response function

$$P_i(\theta) = P(Y_i = 1 | \theta) = \frac{\exp(\theta - b_i)}{1 + \exp(\theta - b_i)}, \quad (4.12)$$

or, in the logit form

$$\lambda_i(\theta) = \log \frac{P_i(\theta)}{Q_i(\theta)} = \theta - b_i; \quad (4.13)$$

and (B) the latent trait conditional on the reference response follows a normal distribution

$$\theta | \mathbf{Y} = \mathbf{0} \sim N(\mu_0, \sigma_0^2). \quad (4.14)$$

Given these two assumptions, applying the Dutch Identity gives the manifest probability for

any pattern \mathbf{y} in the following form, (see Equation (2.15))

$$\log p(\mathbf{y}) = \log p(\mathbf{0}) + \sum_{i=1}^I (-b_i) y_i + \mu_0 \sum_{i=1}^I (y_i) + \frac{1}{2} \sigma_0^2 \left(\sum_i y_i \right)^2. \quad (4.15)$$

To this, we can add the person covariates \mathbf{Z} in the model, which means the model is conditional on the person covariates. The assumption (A), the response follows a Rasch model, is the same as before. For the assumption (B), the conditional distribution of the latent trait given the reference response pattern, we assume a linear regression relationship between the latent trait and the person covariates. The conditional distribution of the latent trait given the reference response pattern is

$$\text{Given } \mathbf{Y} = \mathbf{0} : \quad \theta = \mu_0 + \mathbf{Z}'\boldsymbol{\gamma} + \epsilon, \quad \epsilon \sim N(0, \sigma_{0\epsilon}^2). \quad (4.16)$$

Applying the Dutch Identity theorem yields the manifest probability given the person covariates

$$\begin{aligned} \log p(\mathbf{y}|\mathbf{Z}) &= \log p(\mathbf{0}|\mathbf{Z}) + \sum_{i=1}^I (-b_i) y_i + (\mu_0 + \mathbf{Z}'\boldsymbol{\gamma}) \sum_{i=1}^I (y_i) + \frac{1}{2} \sigma_{0\epsilon}^2 \left(\sum_i y_i \right)^2 \\ &= \log p(\mathbf{0}|\mathbf{Z}) + \sum_{i=1}^I [(-b_i + \mu_0) y_i + \frac{1}{2} \sigma_{0\epsilon}^2 y_i^2] + \mathbf{Z}\boldsymbol{\gamma} \sum_{i=1}^I (y_i) + \sigma_{0\epsilon}^2 \sum_{i \neq i'} y_i y_{i'}. \end{aligned} \quad (4.17)$$

Equation (4.17) is in the form of the LLLA model with person covariates. If we let

$$\lambda(\mathbf{Z}) = \log p(\mathbf{0}|\mathbf{Z}), \quad (4.18)$$

$$\lambda_{i(y_i)} = (-b_i + \mu_0) y_i + \frac{1}{2} \sigma_{0\epsilon}^2 y_i^2, \quad (4.19)$$

$$\sigma_{0\epsilon}^2 = \sigma_{0\epsilon}^2, \quad (4.20)$$

$$\nu_{i(y_i)} = y_i, \quad (4.21)$$

then we get the following LLLA model with person covariates:

$$\log P(\mathbf{Y} = \mathbf{y}|\mathbf{Z}) = \lambda(\mathbf{Z}) + \sum_i \lambda_{i(y_i)} + \mathbf{Z}'\boldsymbol{\gamma} \sum_i \nu_{i(y_i)} + \sigma_{0\epsilon}^2 \sum_{i \neq i'} \nu_{i(y_i)} \nu_{i'(y'_i)}. \quad (4.22)$$

Connection and difference with latent regression Rasch model. A latent regression Rasch model is given by a two-level model with the base model being the Rasch model,

$$\text{Base model : } P_i(\theta) = P(Y_i = 1|\theta) = \frac{\exp(\theta - b_i)}{1 + \exp(\theta - b_i)}, \quad (4.23)$$

and the submodel being a linear regression of the latent trait on the person covariates,

$$\text{Submodel : } \theta = \mathbf{Z}'\boldsymbol{\gamma} + \epsilon, \quad \epsilon \sim N(0, \sigma_\epsilon^2). \quad (4.24)$$

In other words, θ follows a normal distribution: $\theta \sim N(\mathbf{Z}'\boldsymbol{\gamma}, \sigma_\epsilon^2)$.

The above latent regression Rasch model is not exactly equivalent to the LLLA model (4.22), because in the LLLA model the distribution of the latent trait for the whole population is a mixture of normal distributions with a common variance,

$$\theta \sim \sum_{\mathbf{y}} p(\mathbf{y}) N(\mu_{\mathbf{y}} + \mathbf{Z}'\boldsymbol{\gamma}, \sigma_{0\epsilon}^2). \quad (4.25)$$

A mixture-of-normals distribution cannot be a normal distribution, therefore the population distribution of the latent trait described by (4.25) is not the same as the distribution in (4.24). However, the mixture-of-normals distribution (4.25) used by the LLLA model can approximate the normal distribution (4.24) used by the latent regression Rasch model. When applied to a data set with underlying person's ability following a normal distribution, the two models give very similar item parameter estimates (item difficulties). In this case, we

can show that the parameters in the two distributions have the following relationship:

$$\sigma_{\epsilon}^2 = \sigma_{0\epsilon}^2 + \sigma_{0\epsilon}^4 \text{var}(T(\mathbf{y})), \quad (4.26)$$

where $T(\mathbf{y}) = \sum_{i=1}^I y_i$ is the total score for a response pattern $\mathbf{y} = (y_1, y_2, \dots, y_I)$.

Note that if the actual population distribution for the latent trait $p(\theta)$ does not follow a normal distribution, a mixture-of-normals distributions may be a better approximation to $p(\theta)$ because it is more flexible than a normal distribution. Then LLLA will do better than latent regression Rasch model.

Chapter 5

Polytomous Models

One of the research objectives in this thesis is to extend the log-linear-as-IRT models to polytomous items. Under the log-linear model framework, response variables with multiple nominal levels involve no more theoretical complexity than the response variables with just two levels. A procedure to fit log-linear models by default applies to polytomous response data.

In the item response theory field, models for polytomous items are not trivial extension from IRT models for dichotomous items. A procedure for dichotomous items usually does not work for polytomous items. Polytomous IRT models such as polytomous Rasch model have to be developed for polytomous items (Ostini & Nering, 2006; Nering & Ostini, 2010). More complicated case is when the data contain both dichotomous and polytomous items, or polytomous items with different number of levels.

Log-linear models for item response data have clear advantages in modeling polytomous item data, in that a single procedure can handle both dichotomous and polytomous items, and even mixed items. Thus it is important and beneficial to the IRT field to extend the LLLA models for polytomous item response data.

In the first version of R package ‘plRasch’, I implemented the procedure to fit a polytomous Rasch model by fitting the equivalent LLLA model with pseudolikelihood estimation. A proof of the equivalence of the polytomous Rasch model to the LLLA model was given in Anderson, Li, and Vermunt (2007). That proof is based on the method of conditionally specified models as given in Joe and Liu (1996), and this method was used to prove the equivalence between the Rasch model with dichotomous items and the LLLA model in Anderson and Yu (2007). In this thesis, I have been taking a different route by using the Dutch Identity to establish the equivalence between IRT models and LLLA models. I am going to provide a proof for polytomous items using the similar method. However, the Dutch

Identity theorem presented by Holland (1990) only assumes dichotomous items, and no one has published similar results for polytomous items (Holland, personal communication). So I will first extend the Dutch Identity theorem to polytomous items, and then use it as a tool to establish the equivalence of the polytomous Rasch model (the partial credit model) with the LLLA model. The LLLA models with item covariates and with person covariates for polytomous data are derived using the polytomous Dutch Identity theorem.

Review of Polytomous IRT Models

In practice, polytomous items (i.e., items with multiple response options) are very common. Polytomous items can be classified into nominal items and ordinal items. For nominal items, there is no ordering relationship among the multiple choices or categorized responses. For ordinal items, there is an ordering relationship among the multiple choices or graded responses.

Most polytomous IRT models can be seen as extensions of the dichotomous IRT models such as the Rasch model and the 2PL model. Since the Rasch model is a special case of 2PL model, here we write down the 2PL model and see how it is extended to different polytomous models. A 2PL model has an item response function of

$$P(Y_{pi} = 1|\theta_p) = \frac{\exp[a_i(\theta_p - b_i)]}{1 + \exp[a_i(\theta_p - b_i)]}, \quad (5.1)$$

where a_i is the discrimination parameter and b_i is the difficulty parameter for item i . The Rasch model is a special case of 2PL when $a_i = 1$. The item response function can be written in another form by taking the logit transformation on the left hand-side of (5.1),

$$\text{logit}[P(Y_{pi} = 1|\theta_p)] = \log \frac{P(Y_{pi} = 1|\theta_p)}{P(Y_{pi} = 0|\theta_p)} = a_i(\theta_p - b_i). \quad (5.2)$$

The IRF in (5.2) can be summarized in a general form:

$$\text{logit} = \text{linear function of } \theta.$$

Most polytomous IRT models have the similar form. These models can be distinguished by how the logits for the polytomous items are defined.

Definition of polytomous logit functions. For a polytomous response with $(m + 1)$ categories, we can code the response categories into $\{0, 1, \dots, m\}$. A dichotomous item corresponds to the case $m = 1$. For a dichotomous item, the logit function is defined as the log ratio of the probability of response 1 vs the probability of response 0,

$$\text{logit} = \log \frac{P(Y_{pi} = 1 | \theta_p)}{P(Y_{pi} = 0 | \theta_p)} \quad (5.3)$$

that we represent as $\log(p(1)/p(0))$. There are several ways to extend this and define the logit functions for polytomous items (see Agresti, 2002, Chapter 7, pp. 267-313; Mellenbergh, 1995; van der Ark, 2001). They are baseline category logits, adjacent category logits, cumulative logits, and sequential logits. In each of these definitions, for a polytomous item with $(m + 1)$ categories, m non-redundant logit functions need to be specified. All these definitions reduce to the binary logit function in (5.3) when $m = 1$.

Baseline category logit. To define the baseline category logit functions, one of the categories is set as a reference. If the category 0 is selected as the reference category, the baseline category logits are defined by

$$\log \frac{P(Y_{pi} = k | \theta_p)}{P(Y_{pi} = 0 | \theta_p)}, \quad k = 1, \dots, m \quad (5.4)$$

that we represent as $\log(p(k)/p(0))$. Once the m baseline category logits are specified, logits defined by log ratio of any two categories can be calculated from the baseline logits. For

example, consider an item with 4 response categories: 0, 1, 2 and 3. The baseline category logits are

$$\log \frac{p(1)}{p(0)}, \quad \log \frac{p(2)}{p(0)}, \quad \log \frac{p(3)}{p(0)}. \quad (5.5)$$

The baseline category logits are demonstrated in Figure 9 (Mellenbergh, 1995). The four categories are represented by the four cells in the leftmost column in the sketch. The remaining three columns illustrate how the three logits are defined. For each logit, the numerator is represented by a white box, and the denominator is represented by a shaded box.

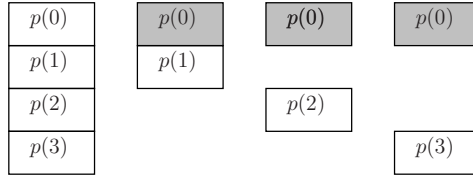


Figure 9. Baseline category logits.

It is possible to derive other baseline category logits with the reference level being other than category 0, solely from the sets of the three logits in (5.5). For example, the logit function defined by the log ratio of category 2 and category 1 is given by $\log(p(2)/p(1)) = \log(p(2)/p(0)) - \log(p(1)/p(0))$.

Adjacent category logit. Adjacent category logits are defined by taking the log ratio of consecutive categories.

$$\log \frac{P(Y_{pi} = k | \theta_p)}{P(Y_{pi} = k - 1 | \theta_p)}, \quad k = 1, \dots, m. \quad (5.6)$$

Similar to baseline category logits, other logits can be calculated from adjacent category logits. For example, the adjacent category logits of an item with 4 categories are

$$\log \frac{p(1)}{p(0)}, \quad \log \frac{p(2)}{p(1)}, \quad \log \frac{p(3)}{p(2)}. \quad (5.7)$$

A sketch of the adjacent category logits is shown in Figure 10 (Mellenbergh, 1995).

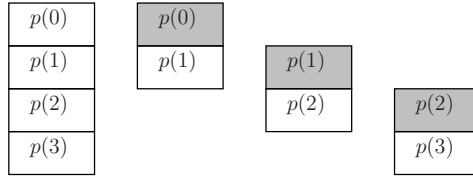


Figure 10. Adjacent category logits.

Any logit function that is defined by the log ratio of two categories can be derived from the set of the three adjacent logit functions in (5.7). For example if we want to calculate the logit of category 3 vs category 1, then the logit $\log(p(3)/p(1)) = \log(p(3)/p(2)) + \log(p(2)/p(1))$.

Based on this fact, adjacent category logits are equivalent to baseline category logits, because they can be derived from each other. Although baseline category logits are often used in the models for nominal response data, and adjacent category logits are often used in the models for ordinal response data, these models can be written down in both the form of baseline logits and the form of adjacent logits, because these two forms of logits are equivalent.

Cumulative logit. Cumulative logits are defined by taking the ratios of the probability above a category (including the category) and the probability below the category.

$$\log \frac{P(Y_{pi} \geq k | \theta_p)}{P(Y_{pi} < k | \theta_p)}, \quad k = 1, \dots, m. \quad (5.8)$$

For a response of 4 categories. the cumulative logits are

$$\log \frac{p(1) + p(2) + p(3)}{p(0)}, \quad \log \frac{p(2) + p(3)}{p(0) + p(1)}, \quad \log \frac{p(3)}{p(0) + p(1) + p(2)}. \quad (5.9)$$

A sketch of the cumulative logits is shown in Figure 11 (Mellenbergh, 1995).

Cumulative logits are different from the logits defined by taking the log ratio of individual categories, such as baseline category logits and adjacent category logits. There is

$p(0)$	$p(0)$	$p(0)$	$p(0)$
$p(1)$	$p(1)$	$+$ $p(1)$	$+$ $p(1)$
$p(2)$	$+$ $p(2)$	$+$ $p(2)$	$+$ $p(2)$
$p(3)$	$+$ $p(3)$	$+$ $p(3)$	$p(3)$

Figure 11. Cumulative logits.

no simple linear relationship between the cumulative logits in (5.9) and the baseline category logits in (5.5) (or the adjacent category logits in (5.7)). Therefore, the models defined by cumulative logits are usually not equivalent to the models defined by baseline category logits (or adjacent category logits).

Sequential logit. Sequential logits (also called continuation ratio logits) are defined by

$$\log \frac{P(Y_{pi} \geq k | \theta_p)}{P(Y_{pi} = k - 1 | \theta_p)}, \quad k = 1, \dots, m. \quad (5.10)$$

For a response of 4 categories, the sequential logits are

$$\log \frac{p(1) + p(2) + p(3)}{p(0)}, \quad \log \frac{p(2) + p(3)}{p(1)}, \quad \log \frac{p(3)}{p(2)}. \quad (5.11)$$

A sketch of the sequential logits is shown in Figure 12 (Mellenbergh, 1995).

$p(0)$	$p(0)$		
$p(1)$	$p(1)$	$p(1)$	
$p(2)$	$+$ $p(2)$	$+$ $p(2)$	$p(2)$
$p(3)$	$+$ $p(3)$	$+$ $p(3)$	$p(3)$

Figure 12. Sequential logits.

Sequential logits share features of both cumulative logits and adjacent category logits. If we take the numerators from the cumulative logits in (5.9) and the denominators from the adjacent category logits in (5.7), and form the log ratios, we will get sequential logits in (5.11).

Three classes of polytomous IRT models. According to the logits used in the model, polytomous IRT models can be classified into three classes (van der Ark, 2001): partial credit models (PCM), graded response models (GRM), and sequential models (SM). Partial credit models use adjacent/baseline category logits. Graded response models use cumulative logits. Sequential models use sequential logits. Assuming unidimensional latent trait and logits linear in the latent trait, these models are special cases of the following form

$$\text{logit} = a_{ik}\theta_p + b_{ik}, \quad k = 1, \dots, m.$$

The family of graded response models include the graded response model (Samejima, 1969). The family of sequential models include the sequential model (Tutz, 1990) and the sequential rating scale model (Tutz, 1990). The family of partial credit models include Bock's model (Bock 1972), the partial credit model (Masters, 1982), the generalized partial credit model (Muraki, 1992), and the rating scale model (Andrich, 1978).

The family of partial credit models. In this thesis, I am going to focus on the family of partial credit models. As I will show later in this chapter, similar to what we have seen in dichotomous models, there is a direct equivalent relationship between the log-linear IRT models and the PCMs for polytomous items. Bock's model is equivalent to the LMA model, and the partial credit model is equivalent to the LLLA model.

Bock's model. The most general form of this family is the Bock (1972) model. Bock's model is given by

$$P(Y_{pi} = k | \theta_p) = \frac{\exp(a_{ik}\theta_p + b_{ik})}{\sum_{h=0}^m \exp(a_{ih}\theta_p + b_{ih})}, \quad (5.12)$$

with the identification constraint for the reference category that $a_{i0} = 0$ and $b_{i0} = 0$. The numerator $\exp(a_{ik}\theta_p + b_{ik})$ can be seen as the weight for the category k and the denominator is the total sum of the weights over all the response categories for item i (reference category 0)

has weight 1). Therefore the family of PCMs are also called divide-by-total models (Thissen & Steinberg, 1986).

Bock's model was proposed as a polytomous model for nominal response data. We can write it in the form of baseline category logits as

$$\log \frac{p(k)}{p(0)} = \log \frac{P(Y_{pi} = k|\theta_p)}{P(Y_{pi} = 0|\theta_p)} = a_{ik}\theta_p + b_{ik}, k = 1, \dots, m. \quad (5.13)$$

As we mentioned before, the baseline category logits are equivalent to adjacent category logits. So Bock's model in the form of adjacent category logits is

$$\begin{aligned} \log \frac{p(k)}{p(k-1)} &= \log \frac{P(Y_{pi} = k|\theta_p)}{P(Y_{pi} = k-1|\theta_p)} \\ &= (a_{ik} - a_{i,k-1})\theta_p + (b_{ik} - b_{i,k-1}) \\ &= a_{ik}^*\theta_p + b_{ik}^*, k = 1, \dots, m, \end{aligned} \quad (5.14)$$

where $a_{ik}^* = a_{ik} - a_{i,k-1}$ and $b_{ik}^* = b_{ik} - b_{i,k-1}$.

The probability of choosing category k in Bock's model is

$$P(Y_{pi} = k|\theta_p) = \frac{\exp(a_{ik}\theta_p + b_{ik})}{\sum_{h=0}^m \exp(a_{ih}\theta_p + b_{ih})}. \quad (5.15)$$

Partial credit model. The partial credit model (Masters, 1982) is a special case of Bock's model by setting $a_{ik}^* = 1$ in (5.14).

$$\log \frac{p(k)}{p(k-1)} = \log \frac{P(Y_{pi} = k|\theta_p)}{P(Y_{pi} = k-1|\theta_p)} = \theta_p + b_{ik}^*, k = 1, \dots, m. \quad (5.16)$$

If we write the PCM in the baseline category logits, the model is in the following form:

$$\log \frac{p(k)}{p(0)} = \log \frac{P(Y_{pi} = k|\theta_p)}{P(Y_{pi} = 0|\theta_p)} = k\theta_p + b_{ik}, k = 1, \dots, m. \quad (5.17)$$

So it is Bock's model with constraint $a_{ik} = k$.

The probability of choosing category k in the PCM is

$$P(Y_{pi} = k | \theta_p) = \frac{\exp(k\theta_p + b_{ik})}{\sum_{h=0}^m \exp(h\theta_p + b_{ih})}. \quad (5.18)$$

Note that if $m = 1$, the PCM becomes the Rasch model. Therefore, the PCM is a direct extension of the Rasch model to polytomous items.

Rating scale model. The rating scale model (Andrich, 1978) is a special case of the PCM model with b_{ik}^* in (5.16) being restricted by

$$b_{ik}^* = \delta_i + \tau_k, \quad (5.19)$$

with constraint $\sum_k \tau_k = 0$.

Generalized partial credit model. The generalized PCM (Muraki, 1992) is obtained by setting Bock's model with $a_{ik}^* = a_i$ or $a_{ik} = ka_i$ (i.e., the slope is constant within the item, but different across the items).

Dutch Identity for Polytomous Models

The Dutch Identity theorem (Holland, 1990) is the central tool in this thesis to establish the equivalence of the Rasch models to their corresponding log-linear models. Holland (1990) only considers dichotomous items (i.e., with binary responses 0 or 1). For polytomous models dealing with responses with more than 2 categories, we need to extend the Dutch Identity to polytomous responses.

Suppose that we have polytomous response data given by an $N \times I$ matrix \mathbf{Y} . The value of the p th person's response to i th item Y_{pi} is from $\{0, 1, \dots, m\}$. The dichotomous response corresponds to the special case $m = 1$. Let $\mathbf{Y} = (Y_1, \dots, Y_I)$ be the response vector of a randomly selected person from the population, and θ is the person's latent trait.

Given the latent trait, Y_i follows a multinomial distribution with $n = 1$. To write down the distribution function, we define

$$U_{ih} = \begin{cases} 1, & \text{if } Y_i = h, \ h = 0, 1, \dots, m \\ 0, & \text{otherwise} \end{cases} \quad (5.20)$$

In other words, $Y_i = h$ corresponds to the $(m + 1)$ vector $(U_{i0}, U_{i1}, \dots, U_{ih}, \dots, U_{im}) = (0, 0, \dots, 1, \dots, 0)$, where only the h th element is 1 and all other elements are 0. It is obvious that $\sum_{h=0}^m U_{ih} = 1$.

The $I \times 1$ vector of the response pattern

$$\mathbf{y} = (y_1, y_2, \dots, y_I)'$$

corresponds to the $I(m + 1) \times 1$ vector of the dummy variables

$$\begin{aligned} \mathbf{u} &= (\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_I) \\ &= (\underbrace{u_{10}, u_{12}, \dots, u_{1m}}_{\text{item 1}}, \underbrace{u_{20}, u_{22}, \dots, u_{2m}}_{\text{item 2}}, \dots, \underbrace{u_{I0}, u_{I2}, \dots, u_{Im}}_{\text{item } I})'. \end{aligned}$$

Define $P_{ih}(\theta) = P(Y_i = h|\theta)$. The pdf of Y_i is given by

$$P(Y_i = y_i|\theta) = P_{i(y_i)}(\theta) = \prod_{h=0}^m P_{ih}(\theta)^{u_{ih}}. \quad (5.21)$$

According to the principle of local independence, the conditional probability of a response pattern given the latent trait is

$$p(\mathbf{y}|\theta) = P[\mathbf{Y} = \mathbf{y}|\theta] = \prod_{i=1}^I P(Y_i = y_i|\theta) = \prod_{i=1}^I \prod_{h=0}^m P_{ih}(\theta)^{u_{ih}}. \quad (5.22)$$

Suppose that the latent trait follows a general distribution with the cumulative dis-

tribution function (CDF) $F(\theta)$, then we can calculate the manifest probability by

$$p(\mathbf{y}) = \int P(\mathbf{Y} = \mathbf{y}|\theta) dF(\theta) = \int \prod_{i=1}^I \prod_{h=0}^m P_{ih}(\theta)^{u_{ih}} dF(\theta). \quad (5.23)$$

The Dutch Identity for polytomous response is given by the following theorem:

Theorem 5.4. (The Dutch Identity extended to polytomous response) *If the manifest probabilities $p(\mathbf{y})$ satisfy (5.23), then for any fixed response pattern \mathbf{y}_0 ,*

$$\frac{p(\mathbf{y})}{p(\mathbf{y}_0)} = E\{\exp[(\mathbf{u} - \mathbf{u}_0)^T \boldsymbol{\delta}(\boldsymbol{\theta})] | \mathbf{Y} = \mathbf{y}_0\}, \quad (5.24)$$

where $\boldsymbol{\delta}(\boldsymbol{\theta}) = (\boldsymbol{\delta}_1(\theta), \dots, \boldsymbol{\delta}_I(\theta))^T$ and $\boldsymbol{\delta}_i(\theta) = (\delta_{i0}(\theta), \delta_{i1}(\theta), \dots, \delta_{im}(\theta))^T$ and

$$\delta_{ih}(\theta) = \log \frac{P_{ih}(\theta)}{P_{i0}(\theta)}. \quad (5.25)$$

Proof. The proof follows closely to the proof provided for the original Dutch Identity Theorem in Holland (1990), with slight modification for the polytomous data.

According to Bayes theorem, we have

$$dF(\theta) = \frac{p(\mathbf{y}_0) dF(\theta | \mathbf{Y} = \mathbf{y}_0)}{P(\mathbf{Y} = \mathbf{y}_0 | \theta)}. \quad (5.26)$$

Substituting (5.26) into (5.23), we have

$$\begin{aligned} p(\mathbf{y}) &= p(\mathbf{y}_0) \int \frac{P(\mathbf{Y} = \mathbf{y} | \theta)}{P(\mathbf{Y} = \mathbf{y}_0 | \theta)} dF(\theta | \mathbf{Y} = \mathbf{y}_0) \\ &= p(\mathbf{y}_0) \int \frac{\prod_{i=1}^I \prod_{h=0}^m P_{ih}(\theta)^{u_{ih}}}{\prod_{i=1}^I \prod_{h=0}^m P_{ih}(\theta)^{u_{0ih}}} dF(\theta | \mathbf{Y} = \mathbf{y}_0) \\ &= p(\mathbf{y}_0) \int \frac{\prod_{i=1}^I \frac{\prod_{h=0}^m P_{ih}(\theta)^{u_{ih}}}{p_{i0}(\theta)}}{\prod_{i=1}^I \frac{\prod_{h=0}^m P_{ih}(\theta)^{u_{0ih}}}{p_{i0}(\theta)}} dF(\theta | \mathbf{Y} = \mathbf{y}_0). \end{aligned} \quad (5.27)$$

Since $\sum_{h=0}^m u_{ih} = 1$,

$$p_{i0}(\theta) = p_{i0}(\theta)^{\sum_{h=0}^m u_{ih}} = \prod_{h=0}^m p_{i0}(\theta)^{u_{ih}}. \quad (5.28)$$

Similarly, $\sum_{h=0}^m u_{0ih} = 1$, we have

$$p_{i0}(\theta) = p_{i0}(\theta)^{\sum_{h=0}^m u_{0ih}} = \prod_{h=0}^m p_{i0}(\theta)^{u_{0ih}}. \quad (5.29)$$

Therefore

$$\begin{aligned} p(\mathbf{y}) &= p(\mathbf{y}_0) \int \frac{\prod_{i=1}^I \prod_{h=0}^m [P_{ih}(\theta)/P_{i0}(\theta)]^{u_{ih}}}{\prod_{i=1}^I \prod_{h=0}^m [P_{ih}(\theta)/P_{i0}(\theta)]^{u_{0ih}}} dF(\theta | \mathbf{Y} = \mathbf{y}_0) \\ &= p(\mathbf{y}_0) \int \prod_{i=1}^I \prod_{h=0}^m \left(\frac{P_{ih}(\theta)}{P_{i0}(\theta)} \right)^{u_{ih} - u_{0ih}} dF(\theta | \mathbf{Y} = \mathbf{y}_0) \\ &= p(\mathbf{y}_0) \int \exp \left[\sum_{i=1}^I \sum_{h=0}^m (u_{ih} - u_{0ih}) \delta_{ih}(\theta) \right] dF(\theta | \mathbf{Y} = \mathbf{y}_0) \end{aligned} \quad (5.30)$$

from which we have (5.24) □

Derivation of the LLLA Model From Bock's Model and the PCM

With the polytomous Dutch Identity theorem for polytomous models, we can now derive the equivalence between the LLLA model (more precisely, the LMA model) to Bock's model. Since the PCM is a special case of Bock's model, this will also lead to the equivalence between the LLLA model and the PCM.

Suppose the polytomous response data follow Bock's model

$$P_{ih}(\theta) = P(Y_i = h | \theta) = \frac{\exp(a_{ih}\theta + b_{ih})}{\sum_{l=0}^m \exp(a_{il}\theta + b_{il})}, \quad (5.31)$$

with the constraint that $a_{i0} = 0$ and $b_{i0} = 0$, then

$$\delta_{ih}(\theta) = \log \frac{P_{ih}(\theta)}{P_{i0}(\theta)} = a_{ih}\theta + b_{ih}. \quad (5.32)$$

Also let's use all-zero response as the reference response pattern, $\mathbf{y}_0 = \mathbf{0}$. Applying the Dutch Identity theorem, then we have

$$p(\mathbf{y}) = p(\mathbf{y}_0) \exp \left[\sum_i \sum_h b_{ih} u_{ih} \right] \int \exp \left[\theta \sum_{i=1}^I \sum_{h=0}^m a_{ih} u_{ih} \right] dF(\theta | \mathbf{Y} = \mathbf{0}) \quad (5.33)$$

Take additional assumption that $\theta | \mathbf{Y} = \mathbf{0} \sim N(\mu_0, \sigma_0^2)$, then

$$\begin{aligned} & E \left\{ \exp \left[\theta \sum_{i=1}^I \sum_{h=0}^m a_{ih} u_{ih} \right] \middle| \mathbf{Y} = \mathbf{0} \right\} \\ &= \exp \left[\mu_0 \sum_{i=1}^I \sum_{h=0}^m a_{ih} u_{ih} + \frac{1}{2} \sigma_0^2 \left(\sum_{i=1}^I \sum_{h=0}^m a_{ih} u_{ih} \right)^2 \right] \end{aligned} \quad (5.34)$$

Therefore

$$\log p(\mathbf{y}) = \log p(\mathbf{0}) + \sum_i \sum_h b_{ih} u_{ih} + \mu_0 \sum_{i=1}^I \sum_{h=0}^m a_{ih} u_{ih} + \frac{1}{2} \sigma_0^2 \left(\sum_{i=1}^I \sum_{h=0}^m a_{ih} u_{ih} \right)^2 \quad (5.35)$$

$$= \log p(\mathbf{0}) + \sum_i b_{i(y_i)} + \mu_0 \sum_{i=1}^I a_{i(y_i)} + \frac{1}{2} \sigma_0^2 \left(\sum_{i=1}^I a_{i(y_i)} \right)^2 \quad (5.36)$$

which is essentially the LLLA model,

$$\log p(\mathbf{y}) = \lambda + \sum_{i=1}^I \lambda_{i(y_i)} + \sigma^2 \left(\sum_{i \neq i'} \nu_{i(y_i)} \nu_{i'(y'_i)} \right), \quad (5.37)$$

where

$$\lambda = \log p(\mathbf{0}), \quad (5.38)$$

$$\lambda_{i(y_i)} = b_{i(y_i)} + \mu_0 a_{i(y_i)} + \frac{1}{2} \sigma_0^2 a_{i(y_i)}^2, \quad (5.39)$$

$$\sigma^2 = \sigma_0^2, \quad (5.40)$$

$$\nu_{i(y_i)} = a_{i(y_i)}. \quad (5.41)$$

In Bock's model, a_{ih} are unknown discrimination parameters, so the corresponding LLLA score parameters $\nu_{i(h)}$ are unknown parameters, which makes the log-linear model an LMA model. Since the PCM is a special case of Bock's model where $a_{ih} = h$, it is equivalent to the LLLA model with score parameters $\nu_{i(h)} = h$.

Polytomous Models With Item Covariates

For dichotomous items, the LLTM is obtained by modeling the item parameters in the Rasch model as a linear combination of the item covariates. We have introduced the LLLA-with-item-covariates (LLLAi) model that is equivalent to the LLTM. To extend these models to polytomous models, we just take the similar method as in the dichotomous models. The Rasch model for polytomous items is the partial credit model (PCM), so we just take the similar modification as in LLTM to obtain the PCM with item covariates (PCMi). The corresponding equivalent model for the manifest probability of the item responses is the same LLLA-with-item-covariates (LLLAi) model we have seen in the dichotomous case.

Partial credit model with item covariates. Adding item covariates to the PCM for polytomous response data is similar to adding item covariates to the Rasch model for binary response data. The original PCM is modified by modeling the item difficulty parameters as a linear combination of the item covariates.

Suppose that Y_{pi} , $p = 1, \dots, N$, $i = 1, \dots, I$ are the $(m + 1)$ -category responses, and

X_{ik} , $k = 1, \dots, K$ are the item covariates associated with the i -th item. Based on PCM, the item category response functions for the p -th person responding i -th item with response h is

$$P_{pih}(\theta_p) = P(Y_{pi} = h|\theta_p) = \frac{\exp(h\theta_p + b_{ih})}{\sum_{l=0}^m \exp(l\theta_p + b_{il})}, \quad h = 0, 1, \dots, m, \quad (5.42)$$

where b_{i0} , which corresponds to the reference response 0, is always 0; and the item difficulty parameters are $(b_{i1}, b_{i2}, \dots, b_{im})$. We write the item difficulty parameters as linear combinations of the item covariates,

$$b_{ih} = \beta_{0h} + \sum_{k=1}^K \beta_{kh} X_{ik}, \quad (5.43)$$

where β_{0h} is the intercept associated with response category 0; and β_{kh} is the effect of the k -th item covariate associated with the response category h .

Combining (5.42) and (5.43), we obtain the partial credit model with item covariates (PCM_i),

$$P_{pih}(\theta_p) = P(Y_{pi} = h|\theta_p) = \frac{\exp(h\theta_p + \beta_{0h} + \sum_{k=1}^K \beta_{kh} X_{ik})}{\sum_{l=0}^m \exp(l\theta_p + \beta_{0l} + \sum_{k=1}^K \beta_{kl} X_{ik})}, \quad h = 0, 1, \dots, m. \quad (5.44)$$

When $m = 1$, the item responses are dichotomous, and the model (5.44) reduces to the LLTM. The PCM is a special case of the PCM_i when item covariates X_{ik} are the item indicator variables: $X_{ik} = 1$ if $k = i$ and $X_{ik} = 0$ if $k \neq i$; and in this case the intercept term in (5.43) is not needed because it is redundant with the item indicator variables.

The number of covariates K should always be less than the total number of items I , otherwise the model is over-parameterized and not identifiable. The PCM_i is a more restricted model than the PCM.

LLA model with item covariates. For polytomous items, the LLLA model with item covariates (LLLA_i) has the same form as what was previously seen for the dichotomous

items. It is a two-level model with the base model (LLLA) given by

$$\log p(\mathbf{Y}) = \lambda + \sum_i \lambda_{i(y_i)} + \sigma_0^2 \sum_{i \neq j} \nu_{i(y_i)} \nu_{j(y_j)} , \quad (5.45)$$

and the submodel (linear combination for item parameter) given by

$$\lambda_{i(y_i)} = \beta_{0(y_i)} + \sum_{k=1}^K \beta_{k(y_i)} X_{ik} . \quad (5.46)$$

Combining the base model (5.45) and the submodel (5.46) yields the LLLAi model in a single equation

$$\log P(\mathbf{Y} = \mathbf{y} | \mathbf{X}) = \lambda + \sum_{i=1}^I \sum_{k=0}^K X_{ik} \beta_{k(y_i)} + \sigma_0^2 \sum_{i \neq i'} \nu_{i(y_i)} \nu_{i'(y_{i'})} , \quad (5.47)$$

where $X_{i0} = 1$ represents the covariate corresponding to the intercept.

It is straightforward to prove the equivalence between the PCMi and the LLLAi model for polytomous items. Assuming that the response data follows the PCMi, and conditional on some reference response pattern the latent trait follows a normal distribution, we can use the polytomous Dutch Identity theorem to derive the form of the manifest distribution, which turns out to be the LLLAi model. The exact relationship between the parameters in the LLLAi model and the PCMi is given by

$$\lambda = \log p(\mathbf{0}) , \quad (5.48)$$

$$\beta_{0(h)} = \beta_{0h} + h\mu_0 + \frac{1}{2}\sigma_0^2 h^2 , \quad (5.49)$$

$$\beta_{k(h)} = \beta_{kh}, \quad k = 1, \dots, K , \quad (5.50)$$

$$\sigma_0^2 = \sigma_0^2 , \quad (5.51)$$

$$\nu_{i(h)} = h . \quad (5.52)$$

Polytomous Models With Person Covariates

For dichotomous items, the latent regression Rasch model is obtained by modeling the person parameters in the Rasch model as a linear combination of the person covariates plus noise. We have derived the LLLA-with-person-covariates (LLLAp) model that is equivalent to the latent regression Rasch model. To extend these models to polytomous models, we just take the similar method as in the dichotomous models. The Rasch model for polytomous items is the partial credit model (PCM), so we just take the similar modification as in the latent regression Rasch model to obtain the PCM with person covariates (PCMp). The corresponding equivalent model for the manifest probability of the item responses is the same LLLAp model we have seen in the dichotomous case.

Partial credit model with person covariates. Adding person covariates to the PCM for polytomous response data is similar to adding person covariates to the Rasch model for binary response data. The original PCM is modified by modeling the latent trait as a linear regression model of the person covariates.

Suppose that Y_{pi} , $p = 1, \dots, N$, $i = 1, \dots, I$ are the $(m + 1)$ -category responses, and $\mathbf{Z} = \{Z_{pk}\}$, $k = 1, \dots, K$ are the person covariates associated with the p -th person. Based on PCM, the item category response function (ICRF) (Chang & Mazzeo, 1994) for the p -th person responding i -th item with response h is

$$P_{pih}(\theta_p) = P(Y_{pi} = h | \theta_p) = \frac{\exp(h\theta_p + b_{ih})}{\sum_{l=0}^m \exp(l\theta_p + b_{il})}, \quad h = 0, 1, \dots, m, \quad (5.53)$$

where b_{i0} , which corresponds to the reference response 0, is always 0; and the item ability parameters are $(b_{i1}, b_{i2}, \dots, b_{im})$. We write the person ability parameter as a linear regression model of the person covariates,

$$\theta | \mathbf{Z} = \mathbf{Z}\boldsymbol{\gamma} + \epsilon, \quad \text{where } \epsilon \sim N(0, \sigma_\epsilon^2), \quad (5.54)$$

where $\boldsymbol{\gamma}$ are the effects of the person covariates \mathbf{Z} .

Similar to the latent regression Rasch model, by including person covariates, the PCMp helps to answer questions regarding how the person covariates relate the latent trait; it also helps to get better prediction of each person's latent trait (ability).

LLA model with person covariates. The development of the LLLAp model for polytomous items is similar to that of the PCMi in the previous section, but with a different assumption on the linear regression models of the ability on the person covariates. Instead of modeling the population distribution of the latent trait as in (5.54), the linear regression model is modeled conditional on the reference response pattern $\mathbf{Y} = \mathbf{0}$.

$$\theta|(\mathbf{Z}, \mathbf{Y} = \mathbf{0}) = \mu_0 + \mathbf{Z}\boldsymbol{\gamma}_0 + \epsilon, \quad \text{where } \epsilon \sim N(0, \sigma_{0\epsilon}^2). \quad (5.55)$$

Together with assumption of PCM model (5.53), and applying the polytomous Dutch Identity theorem, the manifest probability given the person covariates is

$$\begin{aligned} \log p(\mathbf{y}|\mathbf{Z}) &= \log p(\mathbf{0}|\mathbf{Z}) + \sum_{i=1}^I b_{i(y_i)} + (\mu_0 + \mathbf{Z}\boldsymbol{\gamma}_0) \sum_{i=1}^I (y_i) + \frac{1}{2}\sigma_{0\epsilon}^2 \left(\sum_i y_i \right)^2 \\ &= \log p(\mathbf{0}|\mathbf{Z}) + \sum_{i=1}^I [b_{i(y_i)} + \mu_0 y_i + \frac{1}{2}\sigma_{0\epsilon}^2 y_i^2] + \mathbf{Z}\boldsymbol{\gamma}_0 \sum_{i=1}^I (y_i) + \sigma_{0\epsilon}^2 \sum_{i \neq i'} y_i y_{i'}. \end{aligned} \quad (5.56)$$

Equation (5.56) is in the form of the LLLA model with person covariates. If we let

$$\lambda(\mathbf{Z}) = \log p(\mathbf{0}|\mathbf{Z}) \quad (5.57)$$

$$\lambda_{i(y_i)} = b_{i(y_i)} + \mu_0 y_i + \frac{1}{2}\sigma_{0\epsilon}^2 y_i^2 \quad (5.58)$$

$$\sigma_\epsilon^2 = \sigma_{0\epsilon}^2 \quad (5.59)$$

$$\nu_{i(y_i)} = y_i, \quad (5.60)$$

then we get the following LLLA model with person covariates:

$$\log P(\mathbf{Y} = \mathbf{y}|\mathbf{Z}) = \lambda(\mathbf{Z}) + \sum_i \lambda_{i(y_i)} + \mathbf{Z}\boldsymbol{\gamma}_0 \sum_i \nu_{i(y_i)} + \sigma_\epsilon^2 \sum_{i \neq i'} \nu_{i(y_i)} \nu_{i'(y'_i)}. \quad (5.61)$$

Chapter 6

Multidimensional Models

In this chapter, I will develop the LLLA models for multidimensional IRT models. Multidimensional IRT models (Reckase, 2009) assume multiple latent traits. The specific multidimensional IRT models studied in this thesis are the family of multidimensional Rasch models. The multidimensional Rasch models extend the unidimensional Rasch models by replacing the latent trait θ with the linear combinations of the multiple latent traits $\boldsymbol{\theta} = (\theta_1, \dots, \theta_D)$. For example, the item response function for the multidimensional Rasch model for dichotomous items is given by

$$\begin{aligned} P_i(\boldsymbol{\theta}) &= P(Y_i = 1|\boldsymbol{\theta}) \\ &= \frac{\exp(\mathbf{a}_i'\boldsymbol{\theta} + b_i)}{1 + \exp(\mathbf{a}_i'\boldsymbol{\theta} + b_i)} \\ &= \frac{\exp(a_{i1}\theta_1 + \dots + a_{iD}\theta_D + b_i)}{1 + \exp(a_{i1}\theta_1 + \dots + a_{iD}\theta_D + b_i)}, \end{aligned}$$

where a_{id} is the 0-1 indicator of whether item i measures latent trait θ_d . The unidimensional Rasch models that we have seen in previous chapters, for dichotomous items or polytomous items, with item covariates or person covariates, can all be extended to the multidimensional models in the same way.

I will derive the LLLA models for these multidimensional Rasch models. There are three types of models: (a) multidimensional LLLA models without covariates, (b) multidimensional LLLA models with item covariates, and (c) multidimensional LLLA models with person covariates. For each type of the models, I will derive the model for dichotomous items, then the model for polytomous items, and I will show how to estimate the parameters in the Rasch models after LLLA parameters are obtained.

Multidimensional LLLA Model

Dichotomous model.

Assumptions. To derive the model for the manifest probability $p(\mathbf{y})$, we start from the following assumptions:

- (A) *The data follow the multidimensional Rasch model.* First, local independence is satisfied. Conditional on the latent traits $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_D)$, the distribution of the manifest variables satisfies

$$p(\mathbf{y}|\boldsymbol{\theta}) = p(y_1, y_2, \dots, y_I | \theta_1, \dots, \theta_D) = \prod_{i=1}^I p(y_i | \theta_1, \dots, \theta_D). \quad (6.1)$$

The distribution $p(y_i|\boldsymbol{\theta})$ is a Bernoulli distribution

$$p(y_i|\boldsymbol{\theta}) = P(Y_i = y_i|\boldsymbol{\theta}) = P_i(\boldsymbol{\theta})^{y_i} Q_i(\boldsymbol{\theta})^{1-y_i}, \quad (6.2)$$

where $P_i(\boldsymbol{\theta}) = P(Y_i = 1|\boldsymbol{\theta})$ is the item response function (IRF) and $Q_i(\boldsymbol{\theta}) = 1 - P_i(\boldsymbol{\theta})$.

The IRF of the multidimensional Rasch model is given by

$$P_i(\boldsymbol{\theta}) = P(Y_i = 1|\boldsymbol{\theta}) = \frac{\exp(\mathbf{a}'_i \boldsymbol{\theta} + b_i)}{1 + \exp(\mathbf{a}'_i \boldsymbol{\theta} + b_i)} = \frac{\exp(a_{i1}\theta_1 + \dots + a_{iD}\theta_D + b_i)}{1 + \exp(a_{i1}\theta_1 + \dots + a_{iD}\theta_D + b_i)} \quad (6.3)$$

where a_{id} , $d = 1, \dots, D$, is the item-trait adjacency value which indicates whether the i -th item Y_i measures the d -th latent trait θ_d ; or whether Y_i is connected to θ_d ($a_{id} = 1$, connected; $a_{id} = 0$, not connected) in the graphical model representation of the item-trait structure. We assume the item-trait structure is already known. The unknown item parameter b_i is the threshold or difficulty parameter, similar to that in the unidimensional Rasch model.

- (B) *Conditional normality of $\boldsymbol{\theta}$.* Given that the responses to the items are the reference response pattern $\mathbf{Y} = \mathbf{y}_0 = \mathbf{0} = (0, \dots, 0)$, the distribution of the latent traits $\boldsymbol{\theta}$

follows a multivariate normal distribution

$$\boldsymbol{\theta}|\mathbf{Y} = \mathbf{y}_0 \sim N(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0), \quad (6.4)$$

with the $D \times 1$ mean vector $\boldsymbol{\mu}_0 = (\mu_{01}, \dots, \mu_{0D})'$ and the $D \times D$ covariance matrix

$$\boldsymbol{\Sigma}_0 = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1D} \\ \sigma_{12} & \sigma_{22} & \dots & \sigma_{2D} \\ \vdots & & & \vdots \\ \sigma_{1D} & \sigma_{2D} & \dots & \sigma_{DD} \end{bmatrix}.$$

The LLLA model. The equivalence of the multidimensional Rasch model and the LLLA model is given in the following theorem.

Theorem 6.5. (*Multidimensional Rasch/LLLA equivalence*) *If the assumptions (A) and (B) are satisfied, then for the manifest probability $p(\mathbf{y}) = p(y_1, \dots, y_I) = P(Y_1 = y_1, \dots, Y_I = y_I)$,*

$$\log p(\mathbf{y}) = \log p(\mathbf{0}) + \mathbf{y}'\mathbf{b} + \mathbf{y}'\mathbf{A}\boldsymbol{\mu}_0 + \frac{1}{2}\mathbf{y}'\mathbf{A}\boldsymbol{\Sigma}_0\mathbf{A}'\mathbf{y} \quad (6.5)$$

that can be written in the form of the LLLA model

$$\log p(\mathbf{y}) = \lambda + \sum_{i=1}^I \lambda_{i(y_i)} + \sum_{d=1}^D \sigma_{dd} \sum_{i=1}^I \sum_{i' > i} \nu_{i(y_i)d} \nu_{i'(y'_i)d} + \sum_{d=1}^D \sum_{d' > d} \sigma_{dd'} \sum_{i=1}^I \sum_{i' \neq i} \nu_{i(y_i)d} \nu_{i'(y'_i)d'}. \quad (6.6)$$

In equation (6.5), \mathbf{b} is the $I \times 1$ vector of item difficulty parameters, $\mathbf{b} = (b_1, \dots, b_I)'$; \mathbf{A} is the $I \times D$ item-trait adjacency matrix, $\mathbf{A} = \{a_{id}\}$, $i = 1, \dots, I$, $d = 1, \dots, D$.

Proof. From assumptions (A) and (B), (6.5) is derived by applying the Dutch Identity theorem (Theorem 2.1). Then (6.6) can be derived by collecting and renaming the terms in (6.5) (see detailed derivation below). \square

Relationship between LLLA and Rasch parameters. I will present the detailed derivation of (6.6) from (6.5). In the end it will give the relationship between the parameters in the LLLA model and the parameters in the Rasch model.

The terms in (6.5) are in matrix forms. Let's expand the equation in scalar forms term by term. For the second term,

$$\mathbf{y}'\mathbf{b} = \sum_{i=1}^I b_i y_i. \quad (6.7)$$

In the third term, let $\mathbf{t} = \mathbf{A}'\mathbf{y}$ be the $D \times 1$ vector of total scores, $\mathbf{t} = (t_1, t_2, \dots, t_D)$. The total score for the d -th latent trait is

$$t_d = \mathbf{y}'\mathbf{a}_d = \sum_{i=1}^I a_{id} y_i. \quad (6.8)$$

In other words, t_d is the total score by adding up the scores of all the items that measure the d -th latent trait. The third term

$$\mathbf{y}'\mathbf{A}\boldsymbol{\mu}_0 = \mathbf{t}'\boldsymbol{\mu}_0 \quad (6.9)$$

$$= \sum_{d=1}^D \mu_{0d} t_d \quad (6.10)$$

$$= \sum_{d=1}^D \mu_{0d} \sum_{i=1}^I a_{id} y_i. \quad (6.11)$$

The fourth term

$$\frac{1}{2}\mathbf{y}'\mathbf{A}\Sigma_0\mathbf{A}'\mathbf{y} = \frac{1}{2}\mathbf{t}'\Sigma_0\mathbf{t} \quad (6.12)$$

$$= \frac{1}{2} \left(\sum_{d=1}^D \sigma_{dd} t_d^2 + 2 \sum_{d=1}^D \sum_{d'>d} \sigma_{dd'} t_d t_{d'} \right) \quad (6.13)$$

$$= \frac{1}{2} \sum_{d=1}^D \sigma_{dd} \left(\sum_{i=1}^I a_{id} y_i \right)^2 + \sum_{d=1}^D \sum_{d'>d} \sigma_{dd'} \left(\sum_{i=1}^I a_{id} y_i \right) \left(\sum_{i'=1}^I a_{i'd'} y_{i'} \right) \quad (6.14)$$

$$= \frac{1}{2} \sum_{d=1}^D \sigma_{dd} \left(\sum_{i=1}^I a_{id}^2 y_i^2 + 2 \sum_{i=1}^I \sum_{i'>i} a_{id} a_{i'd} y_i y_{i'} \right) \\ + \sum_{d=1}^D \sum_{d'>d} \sigma_{dd'} \left(\sum_{i=1}^I a_{id} a_{i'd'} y_i^2 + \sum_{i=1}^I \sum_{i' \neq i} a_{id} a_{i'd'} y_i y_{i'} \right). \quad (6.15)$$

By substituting (6.7), (6.11), and (6.15) into (6.5), and rearranging the terms so that all the terms containing y_i only (including y_i^2) are collected together, the collection will be the main effects; if all the terms containing $y_i y_{i'}$ are collected together, the collection will be the interaction terms. The resulting equation is

$$\begin{aligned} \log p(\mathbf{y}) &= \log p(\mathbf{0}) && \text{(Intercept)} \\ &+ \sum_{i=1}^I b_i y_i + \sum_{d=1}^D \mu_{0d} \sum_{i=1}^I a_{id} y_i + \frac{1}{2} \sum_{d=1}^D \sigma_{dd} \sum_{i=1}^I a_{id}^2 y_i^2 + \sum_{d=1}^D \sum_{d'>d} \sigma_{dd'} \sum_{i=1}^I a_{id} a_{i'd'} y_i^2 \\ &&& \text{(Main effects)} \\ &+ \sum_{d=1}^D \sigma_{dd} \sum_{i=1}^I \sum_{i'>i} a_{id} a_{i'd} y_i y_{i'} + \sum_{d=1}^D \sum_{d'>d} \sigma_{dd'} \sum_{i=1}^I \sum_{i' \neq i} a_{id} a_{i'd'} y_i y_{i'}. && \text{(Interactions)} \end{aligned} \quad (6.16)$$

We obtain the LLLA model (6.6) by renaming the terms in (6.16) as follows,

$$\lambda = \log p(\mathbf{0}) \quad (6.17)$$

$$\lambda_{i(0)} = 0 \quad (6.18)$$

$$\lambda_{i(1)} = b_i + \sum_{d=1}^D \mu_{0d} a_{id} + \frac{1}{2} \sum_{d=1}^D \sigma_{dd} a_{id}^2 + \sum_{d=1}^D \sum_{d'>d} \sigma_{dd'} a_{id} a_{id'} \quad (6.19)$$

$$\nu_{i(0)d} = 0 \quad (6.20)$$

$$\nu_{i(1)d} = a_{id} \quad (6.21)$$

$$\sigma_{dd} = \sigma_{dd} \quad (6.22)$$

$$\sigma_{dd'} = \sigma_{dd'} , \quad (6.23)$$

and noting that $y_i = y_i^2$ since y_i can only take values 0 or 1; and the fact $\nu_{i(y_i)d} = \nu_{i(1)d} y_i$.

The above equations (6.17)-(6.23) present the exact relationship between the parameters in the LLLA model and those in the multidimensional polytomous Rasch model. On the left-hand side are the parameters in the LLLA model, and on the right-hand side are the parameters in the multidimensional Rasch model. Based on these equations we can interpret the parameters in the LLLA model according to the interpretation in the multidimensional Rasch models. After fitting the LLLA model to the item response data, we can use these equations to obtain the estimates of the parameters in the corresponding multidimensional Rasch model.

Polytomous model.

Assumptions. To derive the model for the manifest probability $p(\mathbf{y})$, we start from the following assumptions:

- (A) *The data follow the multidimensional polytomous Rasch model.* First, local independence is satisfied. Conditional on the latent traits $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_D)$, the distribution

of the manifest variables satisfies

$$p(\mathbf{y}|\boldsymbol{\theta}) = p(y_1, y_2, \dots, y_I | \theta_1, \dots, \theta_D) = \prod_{i=1}^I p(y_i | \theta_1, \dots, \theta_D). \quad (6.24)$$

The distribution $p(y_i|\boldsymbol{\theta})$ is a multinomial distribution

$$p(y_i|\boldsymbol{\theta}) = P(Y_i = y_i | \boldsymbol{\theta}) = \prod_{h=0}^m P_{ih}(\boldsymbol{\theta})^{I(y_i=h)}, \quad (6.25)$$

where $P_{ih}(\boldsymbol{\theta}) = P(Y_i = h | \boldsymbol{\theta})$ is the item category response function (ICRF) and $\sum_{h=0}^m P_{ih}(\boldsymbol{\theta}) = 1$; $I(\cdot)$ is the indicator function.

The ICRF of the multidimensional polytomous Rasch model is given by

$$\begin{aligned} P_{ih}(\boldsymbol{\theta}) &= P(Y_i = h | \boldsymbol{\theta}) \\ &= \frac{\exp(\mathbf{a}'_{ih}\boldsymbol{\theta} + b_{ih})}{\sum_{l=0}^m \exp(\mathbf{a}'_{il}\boldsymbol{\theta} + b_{il})}, \end{aligned} \quad (6.26)$$

where

$$\begin{aligned} \mathbf{a}_{ih} &= h\mathbf{a}_i \\ &= (ha_{i1}, \dots, ha_{iD})', \quad h = 0, \dots, m. \end{aligned} \quad (6.27)$$

Alternatively, we can write these as a single equation

$$\begin{aligned} P_{ih}(\boldsymbol{\theta}) &= P(Y_i = h | \boldsymbol{\theta}) \\ &= \frac{\exp(h\mathbf{a}'_i\boldsymbol{\theta} + b_{ih})}{\sum_{l=0}^m \exp(l\mathbf{a}'_i\boldsymbol{\theta} + b_{il})} \\ &= \frac{\exp(ha_{i1}\theta_1 + \dots + ha_{iD}\theta_D + b_{ih})}{\sum_{l=0}^m \exp(la_{i1}\theta_1 + \dots + la_{iD}\theta_D + b_{il})}, \end{aligned} \quad (6.28)$$

where a_{id} , $d = 1, \dots, D$, is the item-trait adjacency value which indicates whether

the i -th item Y_i measures the d -th latent trait θ_d ; or whether Y_i is connected to θ_d ($a_{id} = 1$, connected; $a_{id} = 0$, not connected) in the graphical model representation of the item-trait structure. We assume the item-trait structure is already known. The unknown item parameter b_{ih} is the threshold or difficulty parameter, similar to that in the unidimensional Rasch model.

(B) *Conditional normality of $\boldsymbol{\theta}$.* Given that the responses to the items are the reference response pattern $\mathbf{Y} = \mathbf{y}_0 = \mathbf{0} = (0, \dots, 0)$, the distribution of the latent traits $\boldsymbol{\theta}$ follows a multivariate normal distribution

$$\boldsymbol{\theta} | \mathbf{Y} = \mathbf{y}_0 \sim N(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0), \quad (6.29)$$

with the $D \times 1$ mean vector $\boldsymbol{\mu}_0 = (\mu_{01}, \dots, \mu_{0D})'$ and the $D \times D$ covariance matrix

$$\boldsymbol{\Sigma}_0 = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1D} \\ \sigma_{12} & \sigma_{22} & \dots & \sigma_{2D} \\ \vdots & & & \vdots \\ \sigma_{1D} & \sigma_{2D} & \dots & \sigma_{DD} \end{bmatrix}.$$

The LLLA model. The equivalence of the multidimensional polytomous Rasch model and the LLLA model is given in the following theorem.

Theorem 6.6. (*Multidimensional polytomous Rasch/LLLA equivalence*) *If the assumptions (A) and (B) are satisfied, then for the manifest probability $p(\mathbf{y}) = p(y_1, \dots, y_I) = P(Y_1 = y_1, \dots, Y_I = y_I)$,*

$$\log p(\mathbf{y}) = \log p(\mathbf{0}) + (\mathbf{u} - \mathbf{u}_0)' \mathbf{b} + \mathbf{y}' \mathbf{A} \boldsymbol{\mu}_0 + \frac{1}{2} \mathbf{y}' \mathbf{A} \boldsymbol{\Sigma}_0 \mathbf{A}' \mathbf{y}, \quad (6.30)$$

which can be written in the form of the LLLA model

$$\log p(\mathbf{y}) = \lambda + \sum_{i=1}^I \lambda_{i(y_i)} + \sum_{d=1}^D \sigma_{dd} \sum_{i=1}^I \sum_{i' > i} \nu_{i(y_i)d} \nu_{i'(y'_i)d} + \sum_{d=1}^D \sum_{d' > d} \sigma_{dd'} \sum_{i=1}^I \sum_{i' \neq i} \nu_{i(y_i)d} \nu_{i'(y'_i)d'} . \quad (6.31)$$

In (6.30), \mathbf{u} is the dummy coding vector for the response vector \mathbf{y} , \mathbf{u}_0 is the dummy coding vector for the reference pattern $\mathbf{y}_0 = \mathbf{0}$, and \mathbf{b} is the vector of difficulty parameters, see their definitions in the detailed proof below.

Proof. From assumptions (A) and (B), (6.30) is derived by applying the polytomous Dutch Identity theorem (Theorem 5.4). Given this, (6.31) can be derived by collecting and renaming the terms in (6.30). See the detailed proof below. \square

Step 1 of the detailed proof: applying polytomous Dutch Identity theorem. For each $(m+1)$ category response Y_i , define the vector of indicator variables (or dummy variables) as

$$\mathbf{u}_i = (u_{i0}, u_{i1}, \dots, u_{im})', \quad (6.32)$$

where

$$u_{ih} = I(Y_i = h) = \begin{cases} 1, & \text{if } Y_i = h, \ h = 0, 1, \dots, m \\ 0, & \text{otherwise} \end{cases} \quad (6.33)$$

In other words, $Y_i = h$ corresponds to the $m+1$ vector $(u_{i0}, u_{i1}, \dots, u_{ih}, \dots, u_{im}) = (0, 0, \dots, 1, \dots, 0)$, where only the h th element is 1 and all other elements are 0.

The $I \times 1$ vector of the response pattern

$$\mathbf{y} = (y_1, y_2, \dots, y_I)'$$

corresponds to the $(m+1)I \times 1$ vector of the dummy variables

$$\begin{aligned}\mathbf{u} &= (\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_I) \\ &= (u_{10}, u_{12}, \dots, u_{1m}, \mid u_{20}, u_{22}, \dots, u_{2m}, \mid \dots, \mid u_{I0}, u_{I2}, \dots, u_{Im})' .\end{aligned}$$

According to assumption (A), from (6.28) we can calculate the baseline logit function for the ICRF,

$$\begin{aligned}\delta_{ih}(\boldsymbol{\theta}) &= \log \frac{P_{ih}(\boldsymbol{\theta})}{P_{i0}(\boldsymbol{\theta})} = h\mathbf{a}_i'\boldsymbol{\theta} + b_{ih} \\ &= \sum_{d=1}^D ha_{id}\theta_d + b_{ih} \\ &= h(a_{i1}\theta_1 + a_{i2}\theta_2 + \dots + a_{iD}\theta_D) + b_{ih} .\end{aligned}\tag{6.34}$$

The $(m+1)I \times 1$ vector of logits is

$$\begin{aligned}\boldsymbol{\delta}(\boldsymbol{\theta}) &= (\boldsymbol{\delta}_1(\boldsymbol{\theta}), \boldsymbol{\delta}_2(\boldsymbol{\theta}), \dots, \boldsymbol{\delta}_I(\boldsymbol{\theta})) \\ &= (\delta_{10}(\theta), \delta_{12}(\theta), \dots, \delta_{1m}(\theta), \mid \delta_{20}(\theta), \delta_{22}(\theta), \dots, \delta_{2m}(\theta), \\ &\quad \mid \dots, \mid \delta_{I0}(\theta), \delta_{I2}(\theta), \dots, \delta_{Im}(\theta))' \\ &= \boldsymbol{\Phi}\boldsymbol{\theta} + \mathbf{b},\end{aligned}$$

where $\mathbf{b} = (\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_I)' = \{b_{ih}\}$ is an $(m+1)I \times 1$ vector, and $\boldsymbol{\Phi}$ is the $(m+1)I \times D$ matrix

$$\boldsymbol{\Phi} = \mathbf{A} \otimes \mathbf{h},\tag{6.35}$$

where $\mathbf{A} = \{a_{ij}\}$ is the $I \times D$ item-trait adjacency matrix, and \mathbf{h} is the $(m+1) \times 1$ vector with $\mathbf{h} = (0, 1, \dots, m)'$. The symbol \otimes stands for Kronecker product, and $\mathbf{A} \otimes \mathbf{h} = \{a_{id}\mathbf{h}\}$, $i = 1, \dots, I$, $d = 1, \dots, D$.

As stated in assumption (B), the reference response pattern is $\mathbf{y}_0 = \mathbf{0}$, and the corresponding dummy variable vector is $\mathbf{u}_0 = (1, 0, \dots, 0, 1, 0, \dots, 0, \dots, 1, 0, \dots, 0)'$.

Applying the polytomous Dutch Identity theorem (Theorem 5.4),

$$\begin{aligned} p(\mathbf{y}) &= p(\mathbf{y}_0) E\{\exp[(\mathbf{u} - \mathbf{u}_0)' \boldsymbol{\delta}(\boldsymbol{\theta})] | \mathbf{Y} = \mathbf{y}_0\} \\ &= p(\mathbf{y}_0) \exp[(\mathbf{u} - \mathbf{u}_0)' \mathbf{b}] \int \exp[(\mathbf{u} - \mathbf{u}_0)' \boldsymbol{\Phi} \boldsymbol{\theta}] dF(\boldsymbol{\theta} | \mathbf{Y} = \mathbf{0}). \end{aligned} \quad (6.36)$$

Taking the assumption (B) that $\boldsymbol{\theta} | \mathbf{Y} = \mathbf{0} \sim N(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$,

$$\begin{aligned} &E \left\{ \exp[(\mathbf{u} - \mathbf{u}_0)' \boldsymbol{\Phi} \boldsymbol{\theta}] \middle| \mathbf{Y} = \mathbf{0} \right\} \\ &= \int \exp[(\mathbf{u} - \mathbf{u}_0)' \boldsymbol{\Phi} \boldsymbol{\theta}] dF(\boldsymbol{\theta} | \mathbf{Y} = \mathbf{0}) \\ &= \exp \left[(\mathbf{u} - \mathbf{u}_0)' \boldsymbol{\Phi} \boldsymbol{\mu}_0 + \frac{1}{2} (\mathbf{u} - \mathbf{u}_0)' \boldsymbol{\Phi} \boldsymbol{\Sigma}_0 \boldsymbol{\Phi}' (\mathbf{u} - \mathbf{u}_0) \right]. \end{aligned} \quad (6.37)$$

Therefore,

$$\begin{aligned} \log p(\mathbf{y}) &= \log p(\mathbf{0}) + (\mathbf{u} - \mathbf{u}_0)' \mathbf{b} + (\mathbf{u} - \mathbf{u}_0)' \boldsymbol{\Phi} \boldsymbol{\mu}_0 + \frac{1}{2} (\mathbf{u} - \mathbf{u}_0)' \boldsymbol{\Phi} \boldsymbol{\Sigma}_0 \boldsymbol{\Phi}' (\mathbf{u} - \mathbf{u}_0). \end{aligned} \quad (6.38)$$

It is easy to see by expanding into algebraic form that

$$\boldsymbol{\Phi}' \mathbf{u} = \mathbf{A}' \mathbf{y} \quad (6.39)$$

and

$$\boldsymbol{\Phi}' \mathbf{u}_0 = \mathbf{A}' \mathbf{0} = \mathbf{0}. \quad (6.40)$$

Thus we have proved (6.30) is true.

Step 2 of the detailed proof: the relationship between LLLA and Rasch parameters. I will present the detailed derivation of (6.31) from (6.30). In the end it will give the relationship between the parameters in the LLLA model and the parameters in the Rasch model.

The terms in (6.30) are in matrix forms. Let's expand the equation in scalar forms term by term. For the second term,

$$(\mathbf{u} - \mathbf{u}_0)' \mathbf{b} = \sum_i \sum_h u_{ih} b_{ih} - 0 = \sum_i b_{i(y_i)}. \quad (6.41)$$

In the third term, let $\mathbf{t} = \mathbf{A}' \mathbf{y}$ be the $D \times 1$ vector of total scores, $\mathbf{t} = (t_1, t_2, \dots, t_D)$. The total score for the d -th latent trait is

$$t_d = \mathbf{y}' \mathbf{a}_d = \sum_{i=1}^I a_{id} y_i. \quad (6.42)$$

In other words, t_d is the total score by adding up the scores of all the items that measure the d -th latent trait. The third term

$$\mathbf{y}' \mathbf{A} \boldsymbol{\mu}_0 = \mathbf{t}' \boldsymbol{\mu}_0 \quad (6.43)$$

$$= \sum_{d=1}^D \mu_{0d} t_d \quad (6.44)$$

$$= \sum_{d=1}^D \mu_{0d} \sum_{i=1}^I a_{id} y_i. \quad (6.45)$$

The fourth term

$$\frac{1}{2}\mathbf{y}'\mathbf{A}\Sigma_0\mathbf{A}'\mathbf{y} = \frac{1}{2}\mathbf{t}'\Sigma_0\mathbf{t} \quad (6.46)$$

$$= \frac{1}{2} \left(\sum_{d=1}^D \sigma_{dd} t_d^2 + 2 \sum_{d=1}^D \sum_{d'>d} \sigma_{dd'} t_d t_{d'} \right) \quad (6.47)$$

$$= \frac{1}{2} \sum_{d=1}^D \sigma_{dd} \left(\sum_{i=1}^I a_{id} y_i \right)^2 + \sum_{d=1}^D \sum_{d'>d} \sigma_{dd'} \left(\sum_{i=1}^I a_{id} y_i \right) \left(\sum_{i'=1}^I a_{i'd'} y_{i'} \right) \quad (6.48)$$

$$= \frac{1}{2} \sum_{d=1}^D \sigma_{dd} \left(\sum_{i=1}^I a_{id}^2 y_i^2 + 2 \sum_{i=1}^I \sum_{i'>i} a_{id} a_{i'd} y_i y_{i'} \right) \\ + \sum_{d=1}^D \sum_{d'>d} \sigma_{dd'} \left(\sum_{i=1}^I a_{id} a_{i'd'} y_i^2 + \sum_{i=1}^I \sum_{i' \neq i} a_{id} a_{i'd'} y_i y_{i'} \right). \quad (6.49)$$

By substituting (6.41), (6.45), and (6.49) into (6.30), and rearranging the terms so that all the terms containing y_i only (including y_i^2) are collected together, the collection will be the main effects; if all the terms containing $y_i y_{i'}$ are collected together, the collection will be the interaction terms. The resulting equation is

$$\begin{aligned} \log p(\mathbf{y}) &= \log p(\mathbf{0}) && \text{(Intercept)} \\ &+ \sum_{i=1}^I b_{i(y_i)} + \sum_{d=1}^D \mu_{0d} \sum_{i=1}^I a_{id} y_i + \frac{1}{2} \sum_{d=1}^D \sigma_{dd} \sum_{i=1}^I a_{id}^2 y_i^2 + \sum_{d=1}^D \sum_{d'>d} \sigma_{dd'} \sum_{i=1}^I a_{id} a_{i'd'} y_i^2 \\ &&& \text{(Main effects)} \\ &+ \sum_{d=1}^D \sigma_{dd} \sum_{i=1}^I \sum_{i'>i} a_{id} a_{i'd} y_i y_{i'} + \sum_{d=1}^D \sum_{d'>d} \sigma_{dd'} \sum_{i=1}^I \sum_{i' \neq i} a_{id} a_{i'd'} y_i y_{i'}. && \text{(Interactions)} \end{aligned} \quad (6.50)$$

We obtain the LLLA model (6.31) by renaming the terms in (6.50) as follows,

$$\lambda = \log p(\mathbf{0}) , \quad (6.51)$$

$$\lambda_{i(y_i)} = b_{i(y_i)} + \sum_{d=1}^D \mu_{0d} a_{id} y_i + \frac{1}{2} \sum_{d=1}^D \sigma_{dd} a_{id}^2 y_i^2 + \sum_{d=1}^D \sum_{d'>d} \sigma_{dd'} a_{id} a_{id'} y_i^2 , \quad (6.52)$$

$$\nu_{i(y_i)d} = a_{id} y_i , \quad (6.53)$$

$$\sigma_{dd} = \sigma_{dd} , \quad (6.54)$$

$$\sigma_{dd'} = \sigma_{dd'} . \quad (6.55)$$

The above equations (6.51)-(6.55) present the exact relationship between the parameters in the LLLA model and those in the multidimensional polytomous Rasch model. On the left-hand side are the parameters in the LLLA model, and on the right-hand side are the parameters in the multidimensional polytomous Rasch model. Based on this we can interpret the parameters in the LLLA model according to the interpretation in the multidimensional polytomous Rasch models.

Estimating Rasch parameters. The equivalence relationship between the LLLA model and the multidimensional dichotomous/polytomous Rasch model provide one way to fit the multidimensional Rasch models through fitting the corresponding LLLA models. After fitting the LLLA model to the item response data, we transform the parameter estimates in the LLLA model into the item and person parameters in the multidimensional Rasch model.

Item parameters. For the item difficulty parameters b_{ih} , $i = 1, \dots, I$, one can solve (6.52) and obtain

$$b_{i(y_i)} = \lambda_{i(y_i)} - \sum_{d=1}^D \mu_{0d} \nu_{i(y_i)d} - \frac{1}{2} \sum_{d=1}^D \sigma_{dd} \nu_{i(y_i)d}^2 - \sum_{d=1}^D \sum_{d'>d} \sigma_{dd'} \nu_{i(y_i)d} \nu_{i(y_i)d'} . \quad (6.56)$$

In the above equation, μ_{0d} are not in the LLLA model equation. Actually μ_{0d} is the mean of d -th latent trait θ_d conditional on the reference response pattern $\mathbf{Y} = \mathbf{0}$. It is not identifiable

unless constraints are applied. Usually one can constraint μ_{0d} such that the sample mean of the estimated person parameter $\hat{\theta}_d$ for the persons in the data set is 0 (this is called “anchoring” according to the person parameters). The formula for estimating μ_{0d} is given later in this section (see Equation (6.58)).

Latent traits. To estimate the person parameters $\boldsymbol{\theta}$, we need to derive the conditional distribution $p(\boldsymbol{\theta}|\mathbf{y})$. It is given by the following theorem.

Theorem 6.7. *(Conditional normality for any pattern) Under the assumptions (A) and (B) in Theorem 6.6, for any response pattern \mathbf{y} , the conditional distribution of $\boldsymbol{\theta}|\mathbf{y}$ is*

$$\boldsymbol{\theta}|\mathbf{y} \sim N(\boldsymbol{\mu}_0 + \boldsymbol{\Sigma}_0 \mathbf{t}, \boldsymbol{\Sigma}_0), \quad (6.57)$$

where $\mathbf{t} = \mathbf{A}'\mathbf{y}$ is the $D \times 1$ vector of total scores.

Proof. The proof is similar to the unidimensional case. □

Based on the conditional normal distribution (6.57), we can estimate the latent traits by $\hat{\boldsymbol{\theta}} = \boldsymbol{\mu}_0 + \hat{\boldsymbol{\Sigma}}_0 \mathbf{t}$, and the corresponding standard errors are obtained from the covariance matrix $\hat{\boldsymbol{\Sigma}}_0$. Since $\boldsymbol{\mu}_0$ is unidentifiable, constraints are imposed such that the sample mean of $\hat{\boldsymbol{\theta}}$ is $\mathbf{0}$. Under this constraint, it is derived that

$$\boldsymbol{\mu}_0 = -\hat{\boldsymbol{\Sigma}}_0 \bar{\mathbf{t}}, \quad (6.58)$$

where $\bar{\mathbf{t}} = \sum_{p=1}^N \mathbf{t}_p / N$ is the sample mean of the total scores over all the persons in the data set. Therefore the estimate of the latent traits is given by

$$\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\Sigma}}_0 (\mathbf{t} - \bar{\mathbf{t}}). \quad (6.59)$$

The estimate is a linear function of the total scores.

Population latent trait distribution. Under the LLLA model, the population distribution of the latent traits $p(\boldsymbol{\theta})$ is a mixture of the conditional distributions $p(\boldsymbol{\theta}|\mathbf{y})$, which are multivariate normal as given by (6.57),

$$p(\boldsymbol{\theta}) = \sum_{\text{All } \mathbf{y}} p(\mathbf{y})p(\boldsymbol{\theta}|\mathbf{y}) = \sum_{\text{All } \mathbf{y}} p(\mathbf{y})N(\boldsymbol{\theta}|\boldsymbol{\mu}_0 + \boldsymbol{\Sigma}_0\mathbf{t}, \boldsymbol{\Sigma}_0), \quad (6.60)$$

where $p(\mathbf{y})$ is given by the LLLA model.

The mean of the population distribution of $\boldsymbol{\theta}$ is

$$\begin{aligned} E(\boldsymbol{\theta}) &= E(E(\boldsymbol{\theta}|\mathbf{y})) \\ &= E(\boldsymbol{\mu}_0 + \boldsymbol{\Sigma}_0\mathbf{t}) \\ &= \boldsymbol{\mu}_0 + \boldsymbol{\Sigma}_0E(\mathbf{t}), \end{aligned} \quad (6.61)$$

and the covariance matrix of $\boldsymbol{\theta}$ is

$$\begin{aligned} \text{cov}(\boldsymbol{\theta}) &= \text{cov}(E(\boldsymbol{\theta}|\mathbf{y})) + E(\text{cov}(\boldsymbol{\theta}|\mathbf{y})) \\ &= \text{cov}(\boldsymbol{\mu}_0 + \boldsymbol{\Sigma}_0\mathbf{t}) + E(\boldsymbol{\Sigma}_0) \\ &= \boldsymbol{\Sigma}_0 \text{cov}(\mathbf{t})\boldsymbol{\Sigma}_0' + \boldsymbol{\Sigma}_0. \end{aligned} \quad (6.62)$$

Equations (6.61) and (6.62) provide the relationship between the population mean and covariance matrix of the latent traits and the conditional mean and covariance matrix of the latent traits given the response patterns. In (6.61) and (6.62), $E(\mathbf{t})$ and $\text{cov}(\mathbf{t})$ can be calculated from the manifest probability $p(\mathbf{y})$. However, this involves evaluating $p(\mathbf{y})$ for all possible response patterns \mathbf{y} and it is computationally prohibitive when the number of items is large. Instead, in practice we can approximate $E(\mathbf{t})$ and $\text{cov}(\mathbf{t})$ by the sample mean and covariance matrix of \mathbf{t} calculated from the data.

Multidimensional LLLA Model With Item Covariates

Dichotomous model.

Assumptions. To derive the model for the manifest probability $p(\mathbf{y})$, we start from the following assumptions:

- (A) *The data follow the multidimensional Rasch model with item covariates.* First, local independence is satisfied. Conditional on the latent traits $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_D)$, the distribution of the manifest variables satisfies

$$p(\mathbf{y}|\boldsymbol{\theta}) = p(y_1, y_2, \dots, y_I | \theta_1, \dots, \theta_D) = \prod_{i=1}^I p(y_i | \theta_1, \dots, \theta_D). \quad (6.63)$$

The distribution $p(y_i|\boldsymbol{\theta})$ is a Bernoulli distribution

$$p(y_i|\boldsymbol{\theta}) = P(Y_i = y_i|\boldsymbol{\theta}) = P_i(\boldsymbol{\theta})^{y_i} Q_i(\boldsymbol{\theta})^{1-y_i}, \quad (6.64)$$

where $P_i(\boldsymbol{\theta}) = P(Y_i = 1|\boldsymbol{\theta})$ is the item response function (IRF) and $Q_i(\boldsymbol{\theta}) = 1 - P_i(\boldsymbol{\theta})$.

The IRF of the model is given by

$$P_i(\boldsymbol{\theta}) = P(Y_i = 1|\boldsymbol{\theta}) = \frac{\exp(\mathbf{a}'_i \boldsymbol{\theta} + b_i)}{1 + \exp(\mathbf{a}'_i \boldsymbol{\theta} + b_i)} = \frac{\exp(a_{i1}\theta_1 + \dots + a_{iD}\theta_D + b_i)}{1 + \exp(a_{i1}\theta_1 + \dots + a_{iD}\theta_D + b_i)} \quad (6.65)$$

where a_{id} , $d = 1, \dots, D$, is the item-trait adjacency value that indicates whether the i -th item Y_i measures the d -th latent trait θ_d ; or whether Y_i is connected to θ_d ($a_{id} = 1$, connected; $a_{id} = 0$, not connected) in the graphical model representation of the item-trait structure. We assume the item-trait structure is already known. Furthermore the item difficulty parameter b_i is modeled as a linear combination of item covariates,

$$b_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_K x_{iK} = \beta_0 + \sum_{k=1}^K \beta_k x_{ik}, \quad (6.66)$$

or, written in matrix form,

$$b_i = \mathbf{x}_i' \boldsymbol{\beta}, \quad (6.67)$$

where $\mathbf{x}_i = (1, x_{i1}, \dots, x_{iK})'$ and $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_K)'$. The IRF of the multidimensional Rasch model with item covariates is

$$P_i(\boldsymbol{\theta}) = P(Y_i = 1 | \boldsymbol{\theta}) = \frac{\exp(\mathbf{a}_i' \boldsymbol{\theta} + \mathbf{x}_i' \boldsymbol{\beta})}{1 + \exp(\mathbf{a}_i' \boldsymbol{\theta} + \mathbf{x}_i' \boldsymbol{\beta})}. \quad (6.68)$$

(B) *Conditional normality of $\boldsymbol{\theta}$.* Given that the responses to the items are the reference response pattern $\mathbf{Y} = \mathbf{y}_0 = \mathbf{0} = (0, \dots, 0)$, the distribution of the latent traits $\boldsymbol{\theta}$ follows a multivariate normal distribution

$$\boldsymbol{\theta} | \mathbf{Y} = \mathbf{y}_0 \sim N(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0), \quad (6.69)$$

with the $D \times 1$ mean vector $\boldsymbol{\mu}_0 = (\mu_{01}, \dots, \mu_{0D})'$ and the $D \times D$ covariance matrix

$$\boldsymbol{\Sigma}_0 = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1D} \\ \sigma_{12} & \sigma_{22} & \dots & \sigma_{2D} \\ \vdots & & & \vdots \\ \sigma_{1D} & \sigma_{2D} & \dots & \sigma_{DD} \end{bmatrix}.$$

The LLLAi model. The equivalence of the multidimensional Rasch model with item covariates and the LLLAi model is given in the following theorem.

Theorem 6.8. (*Multidimensional Rasch/LLLA equivalence with item covariates*) *If the assumptions (A) and (B) are satisfied, then for the manifest probability $p(\mathbf{y}) = p(y_1, \dots, y_I) = P(Y_1 = y_1, \dots, Y_I = y_I)$,*

$$\log p(\mathbf{y}) = \log p(\mathbf{0}) + \mathbf{y}' \mathbf{X} \boldsymbol{\beta} + \mathbf{y}' \mathbf{A} \boldsymbol{\mu}_0 + \frac{1}{2} \mathbf{y}' \mathbf{A} \boldsymbol{\Sigma}_0 \mathbf{A}' \mathbf{y}, \quad (6.70)$$

which can be written in the form of the LLLA model

$$\log p(\mathbf{y}) = \lambda + \sum_{i=1}^I \sum_{k=0}^K \beta_{i(y_i)} x_{ik} + \sum_{d=1}^D \sigma_{dd} \sum_{i=1}^I \sum_{i' > i} \nu_{i(y_i)d} \nu_{i'(y'_i)d} + \sum_{d=1}^D \sum_{d' > d} \sigma_{dd'} \sum_{i=1}^I \sum_{i' \neq i} \nu_{i(y_i)d} \nu_{i'(y'_i)d'} . \quad (6.71)$$

In (6.70), \mathbf{X} is an $I \times (K + 1)$ matrix of item covariates,

$$\mathbf{X} = \begin{pmatrix} \mathbf{1} & \mathbf{x}_1 & \dots & \mathbf{x}_K \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & \dots & x_{1K} \\ 1 & x_{21} & \dots & x_{2K} \\ \vdots & & & \vdots \\ 1 & x_{21} & \dots & x_{2K} \end{pmatrix} \quad (6.72)$$

Proof. The proof is the same as in Theorem 6.5. From assumptions (A) and (B), (6.70) is derived by application of the Dutch Identity theorem (Theorem 2.1). Then (6.71) can be derived by collecting and renaming the terms in (6.70). \square

Relationship between LLLA and Rasch parameters. By writing (6.70) in scalar form and rearranging the terms, we obtain

$$\begin{aligned} \log p(\mathbf{y}) &= \log p(\mathbf{0}) && \text{(Intercept)} \\ &+ \sum_{i=1}^I (\beta_0 + \sum_{k=1}^K \beta_k x_{ik}) y_i + \sum_{d=1}^D \mu_{0d} \sum_{i=1}^I a_{id} y_i \\ &+ \frac{1}{2} \sum_{d=1}^D \sigma_{dd} \sum_{i=1}^I a_{id}^2 y_i^2 + \sum_{d=1}^D \sum_{d' > d} \sigma_{dd'} \sum_{i=1}^I a_{id} a_{id'} y_i^2 && \text{(Main effects)} \\ &+ \sum_{d=1}^D \sigma_{dd} \sum_{i=1}^I \sum_{i' > i} a_{id} a_{i'd} y_i y_{i'} + \sum_{d=1}^D \sum_{d' > d} \sigma_{dd'} \sum_{i=1}^I \sum_{i' \neq i} a_{id} a_{i'd'} y_i y_{i'} . && \text{(Interactions)} \end{aligned} \quad (6.73)$$

We obtain the LLLA model (6.71) by renaming the terms in (6.73) as follows,

$$\lambda = \log p(\mathbf{0}) \quad (6.74)$$

$$\beta_{k(0)} = 0 \quad (6.75)$$

$$\beta_{k(1)} = \beta_k, \quad k = 1, \dots, K, \quad (6.76)$$

$$\beta_{0(1)} = \beta_0 + \sum_{d=1}^D \mu_{0d} a_{id} + \frac{1}{2} \sum_{d=1}^D \sigma_{dd} a_{id}^2 + \sum_{d=1}^D \sum_{d' > d} \sigma_{dd'} a_{id} a_{id'} \quad (6.77)$$

$$\nu_{i(0)d} = 0 \quad (6.78)$$

$$\nu_{i(1)d} = a_{id} \quad (6.79)$$

$$\sigma_{dd} = \sigma_{dd} \quad (6.80)$$

$$\sigma_{dd'} = \sigma_{dd'} , \quad (6.81)$$

and noting that $y_i = y_i^2$ since y_i can only take values 0 or 1; and the fact $\nu_{i(y_i)d} = \nu_{i(1)d} y_i$.

The above equations (6.74)-(6.81) present the exact relationship of the parameters in the LLLAi model and the multidimensional Rasch model with item covariates.

Polytomous model.

Assumptions. To derive the model for the manifest probability $p(\mathbf{y})$, we start from the following assumptions:

- (A) *The data follow the multidimensional polytomous Rasch model with item covariates.* First, local independence is satisfied. Conditional on the latent traits $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_D)$, the distribution of the manifest variables satisfies

$$p(\mathbf{y}|\boldsymbol{\theta}) = p(y_1, y_2, \dots, y_I | \theta_1, \dots, \theta_D) = \prod_{i=1}^I p(y_i | \theta_1, \dots, \theta_D). \quad (6.82)$$

The distribution $p(y_i|\boldsymbol{\theta})$ is a multinomial distribution

$$p(y_i|\boldsymbol{\theta}) = P(Y_i = y_i|\boldsymbol{\theta}) = \prod_{h=0}^m P_{ih}(\boldsymbol{\theta})^{I(y_i=h)}, \quad (6.83)$$

where $P_{ih}(\boldsymbol{\theta}) = P(Y_i = h|\boldsymbol{\theta})$ is the item category response function (ICRF) and $\sum_{h=0}^m P_{ih}(\boldsymbol{\theta}) = 1$; $I(\cdot)$ is the indicator function.

The ICRF of the multidimensional polytomous Rasch model is given by

$$P_{ih}(\boldsymbol{\theta}) = P(Y_i = h|\boldsymbol{\theta}) \quad (6.84)$$

$$= \frac{\exp(h\mathbf{a}_i'\boldsymbol{\theta} + b_{ih})}{\sum_{l=0}^m \exp(l\mathbf{a}_i'\boldsymbol{\theta} + b_{il})}, \quad (6.85)$$

where a_{id} , $d = 1, \dots, D$, is the item-trait adjacency value which indicates whether the i -th item Y_i measures the d -th latent trait θ_d ; or whether Y_i is connected to θ_d ($a_{id} = 1$, connected; $a_{id} = 0$, not connected) in the graphical model representation of the item-trait structure. We assume the item-trait structure is already known. Furthermore the item difficulty parameter b_{ih} associated with the i -th item and h -th response category is modeled as a linear combination of item covariates,

$$b_{ih} = \beta_{0h} + \beta_{1h}x_{i1} + \dots + \beta_{Kh}x_{iK} = \beta_{0h} + \sum_{k=1}^K \beta_{kh}x_{ik}, \quad (6.86)$$

or written in matrix form,

$$b_{ih} = \mathbf{x}_i'\boldsymbol{\beta}_h, \quad (6.87)$$

where $\mathbf{x}_i = (1, x_{i1}, \dots, x_{iK})'$ and $\boldsymbol{\beta}_h = (\beta_{0h}, \beta_{1h}, \dots, \beta_{Kh})'$. The ICRF of the multidimensional polytomous Rasch model with item covariates is

$$P_{ih}(\boldsymbol{\theta}) = P(Y_i = h|\boldsymbol{\theta}) = \frac{\exp(h\mathbf{a}_i'\boldsymbol{\theta} + \mathbf{x}_i'\boldsymbol{\beta}_h)}{\sum_{l=0}^m \exp(l\mathbf{a}_i'\boldsymbol{\theta} + \mathbf{x}_i'\boldsymbol{\beta}_l)}. \quad (6.88)$$

(B) *Conditional normality of $\boldsymbol{\theta}$.* Given that the responses to the items are the reference response pattern $\mathbf{Y} = \mathbf{y}_0 = \mathbf{0} = (0, \dots, 0)$, the distribution of the latent traits $\boldsymbol{\theta}$ follows a multivariate normal distribution

$$\boldsymbol{\theta} | \mathbf{Y} = \mathbf{y}_0 \sim N(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0), \quad (6.89)$$

with the $D \times 1$ mean vector $\boldsymbol{\mu}_0 = (\mu_{01}, \dots, \mu_{0D})'$ and the $D \times D$ covariance matrix

$$\boldsymbol{\Sigma}_0 = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1D} \\ \sigma_{12} & \sigma_{22} & \dots & \sigma_{2D} \\ \vdots & & & \vdots \\ \sigma_{1D} & \sigma_{2D} & \dots & \sigma_{DD} \end{bmatrix}.$$

The LLLAi model. The equivalence of the multidimensional polytomous Rasch model with item covariates and the LLLAi model is given in the following theorem.

Theorem 6.9. *(Multidimensional polytomous Rasch/LLLA equivalence with item covariates) If the assumptions (A) and (B) are satisfied, then for the manifest probability $p(\mathbf{y}) = p(y_1, \dots, y_I) = P(Y_1 = y_1, \dots, Y_I = y_I)$,*

$$\log p(\mathbf{y}) = \log p(\mathbf{0}) + \mathbf{u}' \mathbf{X} \boldsymbol{\beta} + \mathbf{y}' \mathbf{A} \boldsymbol{\mu}_0 + \frac{1}{2} \mathbf{y}' \mathbf{A} \boldsymbol{\Sigma}_0 \mathbf{A}' \mathbf{y}, \quad (6.90)$$

which can be written in the form of the LLLA model

$$\log p(\mathbf{y}) = \lambda + \sum_{i=1}^I \sum_{k=0}^K \beta_{i(y_i)} x_{ik} + \sum_{d=1}^D \sigma_{dd} \sum_{i=1}^I \sum_{i' > i} \nu_{i(y_i)d} \nu_{i'(y'_i)d} + \sum_{d=1}^D \sum_{d' > d} \sigma_{dd'} \sum_{i=1}^I \sum_{i' \neq i} \nu_{i(y_i)d} \nu_{i'(y'_i)d'} \cdot \quad (6.91)$$

In (6.90), \mathbf{u} is the $(m+1)I \times 1$ dummy coding vector for the response \mathbf{y} ; \mathbf{X} is an

$I \times (K + 1)$ matrix of item covariates,

$$\mathbf{X} = \begin{pmatrix} \mathbf{1} & \mathbf{x}_1 & \dots & \mathbf{x}_K \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & \dots & x_{1K} \\ 1 & x_{21} & \dots & x_{2K} \\ \vdots & & & \vdots \\ 1 & x_{21} & \dots & x_{2K} \end{pmatrix}. \quad (6.92)$$

Proof. The proof is the same as in Theorem 6.6. From assumptions (A) and (B), (6.90) is derived by application of the polytomous Dutch Identity theorem (Theorem 5.4). Then (6.91) can be derived by collecting and renaming the terms in (6.90). \square

Relationship between LLLA and Rasch parameters. By writing (6.90) in scalar form and rearranging the terms, we obtain

$$\begin{aligned} \log p(\mathbf{y}) &= \log p(\mathbf{0}) && \text{(Intercept)} \\ &+ \sum_{i=1}^I (\beta_{0(y_i)} + \sum_{k=1}^K \beta_{k(y_i)} x_{ik}) + \sum_{d=1}^D \mu_{0d} \sum_{i=1}^I a_{id} y_i \\ &+ \frac{1}{2} \sum_{d=1}^D \sigma_{dd} \sum_{i=1}^I a_{id}^2 y_i^2 + \sum_{d=1}^D \sum_{d' > d} \sigma_{dd'} \sum_{i=1}^I a_{id} a_{id'} y_i^2 && \text{(Main effects)} \\ &+ \sum_{d=1}^D \sigma_{dd} \sum_{i=1}^I \sum_{i' > i} a_{id} a_{i'd} y_i y_{i'} + \sum_{d=1}^D \sum_{d' > d} \sigma_{dd'} \sum_{i=1}^I \sum_{i' \neq i} a_{id} a_{i'd'} y_i y_{i'}. && \text{(Interactions)} \end{aligned} \quad (6.93)$$

We obtain the LLLA model (6.91) by renaming the terms in (6.93) as follows,

$$\lambda = \log p(\mathbf{0}), \quad (6.94)$$

$$\beta_{k(0)} = \beta_{k0} = 0, \quad (6.95)$$

$$\beta_{k(h)} = \beta_{kh}, \quad k = 1, \dots, K, \quad h = 1, \dots, m, \quad (6.96)$$

$$\beta_{0(h)} = \beta_{0h} + \sum_{d=1}^D \mu_{0d} a_{id} h + \frac{1}{2} \sum_{d=1}^D \sigma_{dd} a_{id}^2 h^2 + \sum_{d=1}^D \sum_{d' > d}^D \sigma_{dd'} a_{id} a_{id'} h^2, \quad (6.97)$$

$$\nu_{i(0)d} = 0, \quad (6.98)$$

$$\nu_{i(h)d} = a_{id} h, \quad h = 1 \dots, m, \quad (6.99)$$

$$\sigma_{dd} = \sigma_{dd}, \quad (6.100)$$

$$\sigma_{dd'} = \sigma_{dd'}. \quad (6.101)$$

The above equations (6.94)-(6.101) present the exact relationship of the parameters in the LLLAi model and the multidimensional polytomous Rasch model with item covariates.

Estimating Rasch parameters. The equivalence relationship between the LLLA model and the multidimensional dichotomous/polytomous Rasch model with item covariates provide one way to fit the multidimensional Rasch models through fitting the corresponding LLLA models. After fitting the LLLA model to the item response data, we transform the parameter estimates in the LLLA model into the parameters in the multidimensional Rasch model.

Item covariates effects. Unlike the LLLA models without covariates, the effects of item covariates in the LLLA model are the same as those in the multidimensional dichotomous/polytomous Rasch model (see Equation (6.96)). All those complex constant terms that we saw in the LLLA model without covariates are now absorbed into the intercept term $\beta_{0(h)}$ in (6.97).

Latent traits. For estimating the person parameters $\boldsymbol{\theta}$, there is no change from the LLLA model without covariates. The conditional distribution $p(\boldsymbol{\theta}|\mathbf{y})$ for any response pattern \mathbf{y} is

$$\boldsymbol{\theta}|\mathbf{y} \sim N(\boldsymbol{\mu}_0 + \boldsymbol{\Sigma}_0 \mathbf{t}, \boldsymbol{\Sigma}_0), \quad (6.102)$$

where $\mathbf{t} = \mathbf{A}'\mathbf{y}$ is the $D \times 1$ vector of total scores.

Based on the conditional normal distribution (6.102), we can estimate the latent traits by

$$\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\Sigma}}_0(\mathbf{t} - \bar{\mathbf{t}}). \quad (6.103)$$

where $\bar{\mathbf{t}} = \sum_{p=1}^N \mathbf{t}_p / N$ is the $D \times 1$ vector of the sample mean of the total scores over all the persons in the data set. The corresponding standard errors are obtained from the covariance matrix $\hat{\boldsymbol{\Sigma}}_0$.

Population latent trait distribution. Under the LLLA model with item covariates, the population distribution of the latent traits $p(\boldsymbol{\theta})$ is a mixture of the conditional distributions $p(\boldsymbol{\theta}|\mathbf{y})$, which are multivariate normal as given by (6.102),

$$p(\boldsymbol{\theta}) = \sum_{\text{All } \mathbf{y}} p(\mathbf{y}) p(\boldsymbol{\theta}|\mathbf{y}) = \sum_{\text{All } \mathbf{y}} p(\mathbf{y}) N(\boldsymbol{\theta}|\boldsymbol{\mu}_0 + \boldsymbol{\Sigma}_0 \mathbf{t}, \boldsymbol{\Sigma}_0). \quad (6.104)$$

where $p(\mathbf{y})$ is given by the LLLA model.

The mean of the population distribution of $\boldsymbol{\theta}$ is

$$E(\boldsymbol{\theta}) = \boldsymbol{\mu}_0 + \boldsymbol{\Sigma}_0 E(\mathbf{t}), \quad (6.105)$$

and the covariance matrix of $\boldsymbol{\theta}$ is

$$\text{cov}(\boldsymbol{\theta}) = \boldsymbol{\Sigma}_0 \text{cov}(\mathbf{t}) \boldsymbol{\Sigma}_0' + \boldsymbol{\Sigma}_0. \quad (6.106)$$

Equations (6.105) and (6.106) provide the relationship between the population mean and covariance matrix of the latent traits and the conditional mean and covariance matrix of the latent traits given the response patterns. In practice we can approximate $E(\mathbf{t})$ and $\text{cov}(\mathbf{t})$ by the sample mean and covariance matrix of \mathbf{t} calculated from the data.

Multidimensional LLLA Model With Person Covariates

Dichotomous model.

Assumptions. To derive the model for the manifest probability $p(\mathbf{y})$, we start from the following assumptions:

- (A) *The data follow the multidimensional Rasch model with person covariates.* First, local independence is satisfied. Conditional on the latent traits $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_D)$, the distribution of the manifest variables satisfies

$$p(\mathbf{y}|\boldsymbol{\theta}) = p(y_1, y_2, \dots, y_I | \theta_1, \dots, \theta_D) = \prod_{i=1}^I p(y_i | \theta_1, \dots, \theta_D). \quad (6.107)$$

The distribution $p(y_i|\boldsymbol{\theta})$ is a Bernoulli distribution

$$p(y_i|\boldsymbol{\theta}) = P(Y_i = y_i|\boldsymbol{\theta}) = P_i(\boldsymbol{\theta})^{y_i} Q_i(\boldsymbol{\theta})^{1-y_i}, \quad (6.108)$$

where $P_i(\boldsymbol{\theta}) = P(Y_i = 1|\boldsymbol{\theta})$ is the item response function (IRF) and $Q_i(\boldsymbol{\theta}) = 1 - P_i(\boldsymbol{\theta})$.

The IRF of the multidimensional Rasch model is given by

$$P_i(\boldsymbol{\theta}) = P(Y_i = 1|\boldsymbol{\theta}) = \frac{\exp(\mathbf{a}'_i \boldsymbol{\theta} + b_i)}{1 + \exp(\mathbf{a}'_i \boldsymbol{\theta} + b_i)} = \frac{\exp(a_{i1}\theta_1 + \dots + a_{iD}\theta_D + b_i)}{1 + \exp(a_{i1}\theta_1 + \dots + a_{iD}\theta_D + b_i)} \quad (6.109)$$

where a_{id} , $d = 1, \dots, D$, is the item-trait adjacency value which indicates whether the i -th item Y_i measures the d -th latent trait θ_d ; or whether Y_i is connected to θ_d ($a_{id} = 1$, connected; $a_{id} = 0$, not connected) in the graphical model representation of

the item-trait structure. We assume the item-trait structure is already known. The unknown item parameter b_i is the threshold or difficulty parameter, similar to that in the unidimensional Rasch model.

The latent traits $\boldsymbol{\theta}$ is modeled as a linear model on the person covariates $\mathbf{z} = (z_1, \dots, z_K)'$,

$$\boldsymbol{\theta} = \boldsymbol{\Gamma}\mathbf{z} + \boldsymbol{\epsilon}, \quad (6.110)$$

where $\boldsymbol{\Gamma}$ is a $D \times K$ matrix of person covariate effects,

$$\boldsymbol{\Gamma} = \begin{pmatrix} \gamma_{11} & \gamma_{12} & \cdots & \gamma_{1K} \\ \gamma_{21} & \gamma_{22} & \cdots & \gamma_{2K} \\ \vdots & & & \vdots \\ \gamma_{D1} & \gamma_{D2} & \cdots & \gamma_{DK} \end{pmatrix}, \quad (6.111)$$

and $\boldsymbol{\epsilon}$ is the part of the latent traits $\boldsymbol{\theta}$ that is not explained by the person covariates \mathbf{z} .

Combining (6.109) and (6.110), the IRF for the multidimensional Rasch model with person covariates is

$$P_i(\boldsymbol{\theta}) = P(Y_i = 1 | \boldsymbol{\theta}) = \frac{\exp(\mathbf{a}'_i \boldsymbol{\Gamma} \mathbf{z} + \mathbf{a}'_i \boldsymbol{\epsilon} + b_i)}{1 + \exp(\mathbf{a}'_i \boldsymbol{\Gamma} \mathbf{z} + \mathbf{a}'_i \boldsymbol{\epsilon} + b_i)} \quad (6.112)$$

(B) *Conditional normality of $\boldsymbol{\theta}$.* Given that the responses to the items are the reference response pattern $\mathbf{Y} = \mathbf{y}_0 = \mathbf{0} = (0, \dots, 0)$, the distribution of $\boldsymbol{\epsilon}$ follows a multivariate normal distribution

$$\boldsymbol{\epsilon} | \mathbf{Y} = \mathbf{y}_0 \sim N(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0), \quad (6.113)$$

with the $D \times 1$ mean vector $\boldsymbol{\mu}_0 = (\mu_{01}, \dots, \mu_{0D})'$ and the $D \times D$ covariance matrix

$$\boldsymbol{\Sigma}_0 = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1D} \\ \sigma_{12} & \sigma_{22} & \dots & \sigma_{2D} \\ \vdots & & & \vdots \\ \sigma_{1D} & \sigma_{2D} & \dots & \sigma_{DD} \end{bmatrix}.$$

The LLLAp model. The equivalence of the multidimensional Rasch model and the LLLA model given in the following theorem.

Theorem 6.10. (*Multidimensional Rasch/LLLA equivalence with person covariates*) If the assumptions (A) and (B) are satisfied, then for the manifest probability $p(\mathbf{y}) = p(y_1, \dots, y_I) = P(Y_1 = y_1, \dots, Y_I = y_I)$,

$$\log p(\mathbf{y}|\mathbf{z}) = \log p(\mathbf{0}) + \mathbf{y}'\mathbf{b} + \mathbf{y}'\mathbf{A}\boldsymbol{\mu}_0 + \mathbf{y}'\mathbf{A}\boldsymbol{\Gamma}\mathbf{z} + \frac{1}{2}\mathbf{y}'\mathbf{A}\boldsymbol{\Sigma}_0\mathbf{A}'\mathbf{y}, \quad (6.114)$$

which can be written in the form of LLLA the model

$$\begin{aligned} \log p(\mathbf{y}|\mathbf{z}) = & \lambda(\mathbf{z}) + \sum_{i=1}^I \lambda_{i(y_i)} + \sum_{d=1}^D \sum_{k=1}^K \gamma_{kd} z_k \sum_{i=1}^I \nu_{i(y_i)d} \\ & + \sum_{d=1}^D \sigma_{dd} \sum_{i=1}^I \sum_{i' > i} \nu_{i(y_i)d} \nu_{i'(y'_i)d} + \sum_{d=1}^D \sum_{d' > d} \sigma_{dd'} \sum_{i=1}^I \sum_{i' \neq i} \nu_{i(y_i)d} \nu_{i'(y'_i)d'} . \end{aligned} \quad (6.115)$$

Proof. The proof is very similar to the proof of Theorem 6.5. From assumptions (A) and (B), (6.114) is derived by application of the Dutch Identity theorem (Theorem 2.1). Then (6.115) can be derived by collecting and renaming the terms in (6.114). \square

Relationship between LLLA and Rasch parameters. By writing (6.114) in scalar form and rearranging the terms, we obtain

$$\begin{aligned}
\log p(\mathbf{y}|\mathbf{z}) &= \log p(\mathbf{0}|\mathbf{z}) && \text{(Intercept)} \\
&+ \sum_{i=1}^I b_i y_i + \sum_{d=1}^D \mu_{0d} \sum_{i=1}^I a_{id} y_i + \sum_{d=1}^D \sum_{k=1}^K \gamma_{kd} z_k \sum_{i=1}^I a_{id} y_i \\
&+ \frac{1}{2} \sum_{d=1}^D \sigma_{dd} \sum_{i=1}^I a_{id}^2 y_i^2 + \sum_{d=1}^D \sum_{d'>d} \sigma_{dd'} \sum_{i=1}^I a_{id} a_{id'} y_i^2 && \text{(Main effects)} \\
&+ \sum_{d=1}^D \sigma_{dd} \sum_{i=1}^I \sum_{i'>i} a_{id} a_{i'd} y_i y_{i'} + \sum_{d=1}^D \sum_{d'>d} \sigma_{dd'} \sum_{i=1}^I \sum_{i' \neq i} a_{id} a_{i'd'} y_i y_{i'} . && \text{(Interactions)}
\end{aligned} \tag{6.116}$$

We obtain the LLLA model (6.115) by renaming the terms in (6.116) as follows,

$$\lambda(\mathbf{z}) = \log p(\mathbf{0}|\mathbf{z}) \tag{6.117}$$

$$\lambda_{i(0)} = 0 \tag{6.118}$$

$$\lambda_{i(1)} = b_i + \sum_{d=1}^D \mu_{0d} a_{id} + \frac{1}{2} \sum_{d=1}^D \sigma_{dd} a_{id}^2 + \sum_{d=1}^D \sum_{d'>d} \sigma_{dd'} a_{id} a_{id'} \tag{6.119}$$

$$\gamma_{kd} = \gamma_{kd} \tag{6.120}$$

$$\nu_{i(0)d} = 0 \tag{6.121}$$

$$\nu_{i(1)d} = a_{id} \tag{6.122}$$

$$\sigma_{dd} = \sigma_{dd} \tag{6.123}$$

$$\sigma_{dd'} = \sigma_{dd'} , \tag{6.124}$$

and noting that $y_i = y_i^2$ since y_i can only take values 0 or 1; and the fact $\nu_{i(y_i)d} = \nu_{i(1)d} y_i$.

The above equations (6.117)-(6.124) present the exact relationship between the parameters in the LLLA model and those in the multidimensional Rasch model with person covariates. On the left-hand side are the parameters in the LLLA model, and on the right-

hand side are the parameters in the multidimensional Rasch model.

Polytomous model.

Assumptions. To derive the model for the manifest probability $p(\mathbf{y})$, we start from the following assumptions:

- (A) *The data follow the multidimensional polytomous Rasch model with person covariates.* First, local independence is satisfied. Conditional on the latent traits $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_D)$, the distribution of the manifest variables satisfies

$$p(\mathbf{y}|\boldsymbol{\theta}) = p(y_1, y_2, \dots, y_I | \theta_1, \dots, \theta_D) = \prod_{i=1}^I p(y_i | \theta_1, \dots, \theta_D). \quad (6.125)$$

The distribution $p(y_i|\boldsymbol{\theta})$ is a multinomial distribution

$$p(y_i|\boldsymbol{\theta}) = P(Y_i = y_i | \boldsymbol{\theta}) = \prod_{h=0}^m P_{ih}(\boldsymbol{\theta})^{I(y_i=h)}, \quad (6.126)$$

where $P_{ih}(\boldsymbol{\theta}) = P(Y_i = h | \boldsymbol{\theta})$ is the item category response function (ICRF) and $\sum_{h=0}^m P_{ih}(\boldsymbol{\theta}) = 1$; $I(\cdot)$ is the indicator function.

The ICRF of the multidimensional polytomous Rasch model is given by

$$\begin{aligned} P_{ih}(\boldsymbol{\theta}) &= P(Y_i = h | \boldsymbol{\theta}) \\ &= \frac{\exp(h\mathbf{a}'_i\boldsymbol{\theta} + b_{ih})}{\sum_{l=0}^m \exp(l\mathbf{a}'_i\boldsymbol{\theta} + b_{il})}, \end{aligned} \quad (6.127)$$

where a_{id} , $d = 1, \dots, D$, is the item-trait adjacency value which indicates whether the i -th item Y_i measures the d -th latent trait θ_d ; or whether Y_i is connected to θ_d ($a_{id} = 1$, connected; $a_{id} = 0$, not connected) in the graphical model representation of the item-trait structure. We assume the item-trait structure is already known. The unknown item parameter b_i is the threshold or difficulty parameter, similar to that in

the unidimensional Rasch model.

The latent traits $\boldsymbol{\theta}$ is modeled as a linear model on the person covariates $\mathbf{z} = (z_1, \dots, z_K)'$,

$$\boldsymbol{\theta} = \mathbf{\Gamma} \mathbf{z} + \boldsymbol{\epsilon}, \quad (6.128)$$

where $\mathbf{\Gamma}$ is the $D \times K$ matrix of person covariate effects,

$$\mathbf{\Gamma} = \begin{pmatrix} \gamma_{11} & \gamma_{12} & \cdots & \gamma_{1K} \\ \gamma_{21} & \gamma_{22} & \cdots & \gamma_{2K} \\ \vdots & & & \vdots \\ \gamma_{D1} & \gamma_{D2} & \cdots & \gamma_{DK} \end{pmatrix}, \quad (6.129)$$

and $\boldsymbol{\epsilon}$ is the part of the latent traits $\boldsymbol{\theta}$ that is not explained by the person covariates \mathbf{z} .

Combining (6.127) and (6.128), the ICRF for the multidimensional polytomous Rasch model with person covariates is

$$P_{ih}(\boldsymbol{\theta}) = P(Y_i = h | \boldsymbol{\theta}) = \frac{\exp(h \mathbf{a}_i' \mathbf{\Gamma} \mathbf{z} + h \mathbf{a}_i' \boldsymbol{\epsilon} + b_{ih})}{\sum_{l=0}^m \exp(l \mathbf{a}_i' \mathbf{\Gamma} \mathbf{z} + l \mathbf{a}_i' \boldsymbol{\epsilon} + b_{il})} \quad (6.130)$$

(B) *Conditional normality of $\boldsymbol{\theta}$.* Given that the responses to the items are the reference response pattern $\mathbf{Y} = \mathbf{y}_0 = \mathbf{0} = (0, \dots, 0)$, the distribution of $\boldsymbol{\epsilon}$ follows a multivariate normal distribution

$$\boldsymbol{\epsilon} | \mathbf{Y} = \mathbf{y}_0 \sim N(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0), \quad (6.131)$$

with the $D \times 1$ mean vector $\boldsymbol{\mu}_0 = (\mu_{01}, \dots, \mu_{0D})'$ and the $D \times D$ covariance matrix

$$\boldsymbol{\Sigma}_0 = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1D} \\ \sigma_{12} & \sigma_{22} & \dots & \sigma_{2D} \\ \vdots & & & \vdots \\ \sigma_{1D} & \sigma_{2D} & \dots & \sigma_{DD} \end{bmatrix}.$$

The LLLAp model. The equivalence of the multidimensional polytomous Rasch model with person covariates and the LLLAp model is given in the following theorem.

Theorem 6.11. *(Multidimensional polytomous Rasch/LLLA equivalence with person covariates) If the assumptions (A) and (B) are satisfied, then for the manifest probability*
 $p(\mathbf{y}) = p(y_1, \dots, y_I) = P(Y_1 = y_1, \dots, Y_I = y_I),$

$$\log p(\mathbf{y}|\mathbf{z}) = \log p(\mathbf{0}) + \mathbf{u}'\mathbf{b} + \mathbf{y}'\mathbf{A}\boldsymbol{\mu}_0 + \mathbf{y}'\mathbf{A}\boldsymbol{\Gamma}\mathbf{z} + \frac{1}{2}\mathbf{y}'\mathbf{A}\boldsymbol{\Sigma}_0\mathbf{A}'\mathbf{y}, \quad (6.132)$$

which can be written in the form of the LLLA model

$$\begin{aligned} \log p(\mathbf{y}|\mathbf{z}) = & \lambda(\mathbf{z}) + \sum_{i=1}^I \lambda_{i(y_i)} + \sum_{d=1}^D \sum_{k=1}^K \gamma_{kd} z_k \sum_{i=1}^I \nu_{i(y_i)d} \\ & + \sum_{d=1}^D \sigma_{dd} \sum_{i=1}^I \sum_{i' > i} \nu_{i(y_i)d} \nu_{i'(y'_i)d} + \sum_{d=1}^D \sum_{d' > d} \sigma_{dd'} \sum_{i=1}^I \sum_{i' \neq i} \nu_{i(y_i)d} \nu_{i'(y'_i)d'} . \end{aligned} \quad (6.133)$$

In (6.132), \mathbf{u} is the $(m+1)I \times 1$ dummy coding vector for the response \mathbf{y} ; \mathbf{b} is the $(m+1)I \times 1$ the item difficulty parameters.

Proof. The proof is very similar to the proof as Theorem 6.6. From assumptions (A) and (B), (6.132) is derived by application of the polytomous Dutch Identity theorem (Theorem 5.4). Then (6.133) can be derived by collecting and renaming the terms in (6.132). \square

Relationship between LLLA and Rasch parameters. By writing (6.132) in scalar form and rearranging the terms, we obtain

$$\begin{aligned}
\log p(\mathbf{y}|\mathbf{z}) &= \log p(\mathbf{0}|\mathbf{z}) && \text{(Intercept)} \\
&+ \sum_{i=1}^I b_{i(y_i)} + \sum_{d=1}^D \mu_{0d} \sum_{i=1}^I a_{id} y_i + \sum_{d=1}^D \sum_{k=1}^K \gamma_{kd} z_k \sum_{i=1}^I a_{id} y_i \\
&+ \frac{1}{2} \sum_{d=1}^D \sigma_{dd} \sum_{i=1}^I a_{id}^2 y_i^2 + \sum_{d=1}^D \sum_{d'>d} \sigma_{dd'} \sum_{i=1}^I a_{id} a_{id'} y_i^2 && \text{(Main effects)} \\
&+ \sum_{d=1}^D \sigma_{dd} \sum_{i=1}^I \sum_{i'>i} a_{id} a_{i'd} y_i y_{i'} + \sum_{d=1}^D \sum_{d'>d} \sigma_{dd'} \sum_{i=1}^I \sum_{i' \neq i} a_{id} a_{i'd'} y_i y_{i'} . && \text{(Interactions)}
\end{aligned} \tag{6.134}$$

We obtain the LLLA model (6.133) by renaming the terms in (6.134) as follows,

$$\lambda(\mathbf{z}) = \log p(\mathbf{0}|\mathbf{z}) \tag{6.135}$$

$$\lambda_{i(0)} = 0 \tag{6.136}$$

$$\lambda_{i(h)} = b_{ih} + \sum_{d=1}^D \mu_{0d} a_{id} h + \frac{1}{2} \sum_{d=1}^D \sigma_{dd} a_{id}^2 h^2 + \sum_{d=1}^D \sum_{d'>d} \sigma_{dd'} a_{id} a_{id'} h^2 \tag{6.137}$$

$$\gamma_{kd} = \gamma_{kd} \tag{6.138}$$

$$\nu_{i(0)d} = 0 \tag{6.139}$$

$$\nu_{i(h)d} = a_{id} \tag{6.140}$$

$$\sigma_{dd} = \sigma_{dd} \tag{6.141}$$

$$\sigma_{dd'} = \sigma_{dd'} . \tag{6.142}$$

The above equations (6.135)-(6.142) present the exact relationship between the parameters in the LLLA model and those in the multidimensional polytomous Rasch model with person covariates. On the left-hand side are the parameters in the LLLA model, and on the right-hand side are the parameters in the multidimensional polytomous Rasch model.

Estimating Rasch parameters. After fitting the LLLA model to the item response data, we would like to obtain the parameters in the multidimensional polytomous Rasch model. For the item difficulty parameters b_i , $i = 1, \dots, I$, one can solve (6.137) and obtain

$$b_{ih} = \lambda_{i(h)} - \sum_{d=1}^D \mu_{0d} \nu_{i(h)d} - \frac{1}{2} \sum_{d=1}^D \sigma_{dd} \nu_{i(h)d}^2 - \sum_{d=1}^D \sum_{d'>d} \sigma_{dd'} \nu_{i(h)d} \nu_{i(h)d'} . \quad (6.143)$$

In the above equation, μ_{0d} are not in the LLLA model equation. Actually μ_{0d} is the mean of d -th latent trait θ_d conditional on the reference response pattern $\mathbf{Y} = \mathbf{0}$. It is not identifiable unless constraints are applied. Usually one can constraint μ_{0d} such that the sample mean of the estimated person parameter $\hat{\theta}_d$ for the persons in the data set is 0 (this is called “anchoring” according to the person parameters). The formula for estimating μ_{0d} is given later in this section (see Equation (6.145)).

In order to estimate the person parameters $\boldsymbol{\theta}$, we need to derive the conditional distribution $p(\boldsymbol{\theta}|\mathbf{y})$. It is given by the following theorem.

Theorem 6.12. *Under the assumptions (A) and (B) in Theorem 6.11, for any response pattern \mathbf{y} , the conditional distribution of $\boldsymbol{\theta}|\mathbf{y}$ is*

$$\boldsymbol{\theta}|\mathbf{y}, \mathbf{z} \sim N(\boldsymbol{\mu}_0 + \boldsymbol{\Gamma}\mathbf{z} + \boldsymbol{\Sigma}_0\mathbf{t}, \boldsymbol{\Sigma}_0), \quad (6.144)$$

where $\mathbf{t} = \mathbf{A}'\mathbf{y}$ is the $D \times 1$ vector of total scores.

Proof. The proof is similar to the unidimensional case. □

Based on the conditional normal distribution (6.144), we can estimate the latent traits by $\hat{\boldsymbol{\theta}} = \boldsymbol{\mu}_0 + \hat{\boldsymbol{\Gamma}}\mathbf{z} + \hat{\boldsymbol{\Sigma}}_0\mathbf{t}$, and the corresponding standard errors are obtained from the covariance matrix $\hat{\boldsymbol{\Sigma}}_0$. Since $\boldsymbol{\mu}_0$ is unidentifiable, constraints are imposed such that the sample mean

of $\hat{\boldsymbol{\theta}}$ is $\mathbf{0}$. Under this constraint, it is derived that

$$\boldsymbol{\mu}_0 = -\hat{\boldsymbol{\Sigma}}_0 \bar{\mathbf{t}} - \hat{\boldsymbol{\Gamma}} \bar{\mathbf{z}}, \quad (6.145)$$

where $\bar{\mathbf{t}} = \sum_{p=1}^N \mathbf{t}_p / N$ is the $D \times 1$ vector of the sample means of the total scores over all the persons in the data set; $\bar{\mathbf{z}} = \sum_{p=1}^N \mathbf{z}_p / N = (\bar{z}_1, \bar{z}_2, \dots, \bar{z}_K)'$ is the vector of sample mean of the person covariates over all the persons in the data set. Therefore the estimate of the latent traits is given by

$$\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\Gamma}}(\mathbf{z} - \bar{\mathbf{z}}) + \hat{\boldsymbol{\Sigma}}_0(\mathbf{t} - \bar{\mathbf{t}}). \quad (6.146)$$

Population latent trait distribution. Under the LLLA model with person covariates, the population distribution of the latent traits conditional on the person covariates $p(\boldsymbol{\theta}|\mathbf{z})$ is a mixture of the conditional distributions $p(\boldsymbol{\theta}|\mathbf{y}, \mathbf{z})$, which are multivariate normal as given by (6.144),

$$p(\boldsymbol{\theta}|\mathbf{z}) = \sum_{\text{All } \mathbf{y}} p(\mathbf{y}|\mathbf{z}) p(\boldsymbol{\theta}|\mathbf{y}, \mathbf{z}) = \sum_{\text{All } \mathbf{y}} p(\mathbf{y}|\mathbf{z}) N(\boldsymbol{\theta}|\boldsymbol{\mu}_0 + \boldsymbol{\Gamma}\mathbf{z} + \boldsymbol{\Sigma}_0\mathbf{t}, \boldsymbol{\Sigma}_0). \quad (6.147)$$

where $p(\mathbf{y}|\mathbf{z})$ is given by the LLLA model.

The mean of the distribution of $\boldsymbol{\theta}|\mathbf{z}$ is

$$\begin{aligned} E(\boldsymbol{\theta}|\mathbf{z}) &= E(E(\boldsymbol{\theta}|\mathbf{y}, \mathbf{z})) \\ &= E(\boldsymbol{\mu}_0 + \boldsymbol{\Gamma}\mathbf{z} + \boldsymbol{\Sigma}_0\mathbf{t}) \\ &= \boldsymbol{\mu}_0 + \boldsymbol{\Gamma}\mathbf{z} + \boldsymbol{\Sigma}_0 E(\mathbf{t}|\mathbf{z}), \end{aligned} \quad (6.148)$$

and the covariance matrix of $\boldsymbol{\theta}|\mathbf{z}$ is

$$\begin{aligned}
\text{cov}(\boldsymbol{\theta}|\mathbf{z}) &= \text{cov}(E(\boldsymbol{\theta}|\mathbf{y}, \mathbf{z})) + E(\text{cov}(\boldsymbol{\theta}|\mathbf{y}, \mathbf{z})) \\
&= \text{cov}(\boldsymbol{\mu}_0 + \boldsymbol{\Gamma}\mathbf{z} + \boldsymbol{\Sigma}_0\mathbf{t}) + E(\boldsymbol{\Sigma}_0) \\
&= \boldsymbol{\Sigma}_0 \text{cov}(\mathbf{t}|\mathbf{z})\boldsymbol{\Sigma}_0' + \boldsymbol{\Sigma}_0.
\end{aligned} \tag{6.149}$$

Equations (6.151) and (6.152) provide the relationship between the population mean and covariance matrix of the latent traits given the person covariates, and the conditional mean and covariance matrix of the latent traits given the response patterns and the person covariates. In (6.151) and (6.152), $E(\mathbf{t}|\mathbf{z})$ and $\text{cov}(\mathbf{t}|\mathbf{z})$ can be calculated from the manifest probability $p(\mathbf{y}|\mathbf{z})$. However, this involves evaluating $p(\mathbf{y}|\mathbf{z})$ for all possible response patterns \mathbf{y} and it is computationally prohibitive when the number of items is large. we might approximate $E(\mathbf{t}|\mathbf{z})$ and $\text{cov}(\mathbf{t}|\mathbf{z})$ by the sample mean and covariance matrix of \mathbf{t} calculated from the data conditional on \mathbf{z} . However, we may not have enough data points under each configuration of \mathbf{z} , especially if \mathbf{z} is continuous.

We can obtain the unconditional latent trait population distribution $p(\boldsymbol{\theta})$ by integrating out \mathbf{z} ,

$$p(\boldsymbol{\theta}) = \int p(\boldsymbol{\theta}|\mathbf{z})p(\mathbf{z})d\mathbf{z} \tag{6.150}$$

where $p(\mathbf{z})$ is the population distribution of the person covariates.

The mean of the population distribution of $\boldsymbol{\theta}$ is

$$\begin{aligned}
E(\boldsymbol{\theta}) &= E(E(\boldsymbol{\theta}|\mathbf{z})) \\
&= \boldsymbol{\mu}_0 + \boldsymbol{\Gamma}E(\mathbf{z}) + \boldsymbol{\Sigma}_0E(\mathbf{t}),
\end{aligned} \tag{6.151}$$

and the covariance matrix of $\boldsymbol{\theta}$ is

$$\begin{aligned}
\text{cov}(\boldsymbol{\theta}) &= \text{cov}(E(\boldsymbol{\theta}|\mathbf{z})) + E(\text{cov}(\boldsymbol{\theta}|\mathbf{z})) \\
&= \text{Cov}[\boldsymbol{\mu}_0 + \boldsymbol{\Gamma}\mathbf{z} + \boldsymbol{\Sigma}_0 E(\mathbf{t}|\mathbf{z})] + E[\boldsymbol{\Sigma}_0 \text{cov}(\mathbf{t}|\mathbf{z})\boldsymbol{\Sigma}_0 + \boldsymbol{\Sigma}_0] \\
&= \boldsymbol{\Gamma} \text{cov}(\mathbf{z})\boldsymbol{\Gamma}' + \boldsymbol{\Gamma} \text{cov}(\mathbf{z}, E(\mathbf{t}|\mathbf{z}))\boldsymbol{\Sigma}_0' + \boldsymbol{\Sigma}_0 \text{cov}(E(\mathbf{t}|\mathbf{z}), \mathbf{z})\boldsymbol{\Gamma}' + \boldsymbol{\Sigma}_0 \text{cov}(\mathbf{t})\boldsymbol{\Sigma}_0' + \boldsymbol{\Sigma}_0 \\
&= \boldsymbol{\Gamma} \text{cov}(\mathbf{z})\boldsymbol{\Gamma}' + \boldsymbol{\Gamma} \text{cov}(\mathbf{z}, \mathbf{t})\boldsymbol{\Sigma}_0' + \boldsymbol{\Sigma}_0 \text{cov}(\mathbf{t}, \mathbf{z})\boldsymbol{\Gamma}' + \boldsymbol{\Sigma}_0 \text{cov}(\mathbf{t})\boldsymbol{\Sigma}_0' + \boldsymbol{\Sigma}_0. \tag{6.152}
\end{aligned}$$

The last equality is true because

$$\begin{aligned}
\text{cov}(\mathbf{z}, E(\mathbf{t}|\mathbf{z})) &= E(\mathbf{z}'E(\mathbf{t}|\mathbf{z})) - E(\mathbf{z})'E(E(\mathbf{t}|\mathbf{z})) \\
&= E(E(\mathbf{z}'\mathbf{t}|\mathbf{z})) - E(\mathbf{z})'E(\mathbf{t}) \\
&= E(\mathbf{z}'\mathbf{t}) - E(\mathbf{z})'E(\mathbf{t}) \\
&= \text{cov}(\mathbf{z}, \mathbf{t}), \tag{6.153}
\end{aligned}$$

and $\text{cov}(E(\mathbf{t}|\mathbf{z}), \mathbf{z}) = \text{cov}(\mathbf{z}, E(\mathbf{t}|\mathbf{z}))$.

Matching with parameters in MML Rasch models. In the LLLA model with person covariates, latent traits $\boldsymbol{\theta}$ are modeled as a linear regression model conditional on the response pattern \mathbf{y} and person covariates \mathbf{z} (see Equation (6.144)). The linear regression model can be written as

$$\boldsymbol{\theta}|\mathbf{y}, \mathbf{z} = \boldsymbol{\mu}_0 + \boldsymbol{\Gamma}\mathbf{z} + \boldsymbol{\Sigma}_0\mathbf{t} + \boldsymbol{\epsilon}, \tag{6.154}$$

$$\boldsymbol{\epsilon} \sim N(\mathbf{0}, \boldsymbol{\Sigma}_0). \tag{6.155}$$

The interpretation of the parameters $\boldsymbol{\Gamma}$ and $\boldsymbol{\Sigma}_0$ is based on the condition of the given \mathbf{y} . The parameter $\boldsymbol{\Gamma}$ is the effect of \mathbf{z} on $\boldsymbol{\theta}$ for all those individuals in the subpopulation who answer the items with response pattern \mathbf{y} . The parameter $\boldsymbol{\Sigma}_0$ is the covariance of $\boldsymbol{\theta}$ for all those individuals who answer the items with response pattern \mathbf{y} and whose person covariates

equal \mathbf{z} .

On the other hand, in the multidimensional polytomous Rasch model with person covariates under MML formulation, the linear regression model of the latent traits $\boldsymbol{\theta}$ on the covariates \mathbf{z} is

$$\boldsymbol{\theta}|\mathbf{z} = \boldsymbol{\mu}_M + \boldsymbol{\Gamma}_M \mathbf{z} + \boldsymbol{\epsilon}_M, \quad (6.156)$$

$$\boldsymbol{\epsilon}_M \sim N(\mathbf{0}, \boldsymbol{\Sigma}_M), \quad (6.157)$$

where the subscript M is attached to the parameters to indicate they are under the MML formulation of the multidimensional polytomous Rasch model. The interpretation of the parameters $\boldsymbol{\Gamma}_M$ and $\boldsymbol{\Sigma}_M$ applies to the individuals for the whole population after adjusting for person covariates \mathbf{z} .

Under the MML formulation the interpretation of the marginal effects applies to the whole population, where as under the LLLA model the interpretation of the conditional effects applies to the subpopulation with a specific response pattern. Therefore after fitting the LLLA model, it is desirable to estimate or at least approximate the MML parameters $\boldsymbol{\Gamma}_M$ and $\boldsymbol{\Sigma}_M$.

Equations (6.151) and (6.152) gives the mean and the covariance of $p(\boldsymbol{\theta}|\mathbf{z})$ under the LLLA model, by matching them with the mean and the covariance under the MML Rasch model, we can derive

$$\boldsymbol{\Sigma}_M \approx \boldsymbol{\Sigma}_0 \text{cov}(\mathbf{t}|\mathbf{z})\boldsymbol{\Sigma}_0' + \boldsymbol{\Sigma}_0 \quad (6.158)$$

that gives the relationship between $\boldsymbol{\Sigma}_M$ and $\boldsymbol{\Sigma}_0$; and

$$\boldsymbol{\mu}_M + \boldsymbol{\Gamma}_M \mathbf{z} \approx \boldsymbol{\mu}_0 + \boldsymbol{\Gamma} \mathbf{z} + \boldsymbol{\Sigma}_0 E(\mathbf{t}|\mathbf{z}). \quad (6.159)$$

By differentiating both sides of the above equation with respect to \mathbf{z} , we get

$$\begin{aligned}
\mathbf{\Gamma}_M &\approx \mathbf{\Gamma} + \mathbf{\Sigma}_0 \frac{d}{d\mathbf{z}} E(\mathbf{t}|\mathbf{z}) \\
&= \mathbf{\Gamma} + \mathbf{\Sigma}_0 \text{cov}(E(\mathbf{t}|\mathbf{z}), \mathbf{z}) \text{cov}(\mathbf{z})^{-1/2} \\
&= \mathbf{\Gamma} + \mathbf{\Sigma}_0 \text{cov}(\mathbf{t}, \mathbf{z}) \text{cov}(\mathbf{z})^{-1/2}.
\end{aligned} \tag{6.160}$$

The last equality in the above equation is true because of (6.153). This equation gives the relationship between $\mathbf{\Gamma}_M$ and $\mathbf{\Gamma}_0$.

Chapter 7

Simulation Studies

Simulation studies were conducted to evaluate the log-linear-as-IRT models and methods presented in this thesis. Table 7 shows the designs of the simulation studies that were conducted and will be reported in this chapter. There are 4 types of models according to the dimension of the latent traits and the number of categories of the items: unidimensional (latent traits) and binary (item categories), unidimensional and polytomous, multidimensional and binary, multidimensional and polytomous. The simulated data were generated from the equivalent Rasch models with covariates. For each type of model I considered tests with small number of items (5 or 6, 10) and large number of items (50, 100). For each item, the difficulty parameters were assigned with either reasonable values or values generated from the standard normal distribution. For models with item covariates, the item difficulty parameters were calculated from the linear combinations of the item covariates.

The response data were generated as follows. First, person's ability were generated from a population distribution. For models not involving person covariates, a normal distribution was used. If there were person covariates, a linear combination of the person covariates was added to drawn value from the normal distribution. After the each person's ability was generated, I substituted the generated ability and other assigned parameters into the IRF to calculate the probability of answering each response category for each item. Lastly the responses were generated from the Bernoulli distribution for dichotomous items, or the multinomial distribution for polytomous items.

The LLLA models with covariates were then fitted to the simulated data, and finally the results were analyzed. The first criterion of whether the model and the estimation procedure are working is to see whether we can recover the true parameters used to generate the simulated data. For simulated data with small number of items, it is possible to fit the LLLA-with-covariates models with two different estimation procedures: MLE and PLE. It

Table 7

Simulation Designs

Model	dim.	cat.	# items	# persons	equivalent IRT model
Uni-D, Binary					
LLLA item covariates	1	2	5	1000	LLTM
			10	1000	
			50	1000	
			100	1000	
LLLA person covariates	1	2	5	1000	Latent Regression Rasch
			10	1000	
			50	1000	
			100	1000	
Uni-D, Polytomous					
LLLA item covariates	1	3	5	1000	Polytomous Rasch + item covariates (i.e., PCMi)
			10	1000	
			50	1000	
			100	1000	
LLLA person covariates	1	3	5	1000	Polytomous Rasch + person covariates (i.e., PCMp)
			10	1000	
			50	1000	
			100	1000	
Multi-D, Binary					
Multi-D LLLA item covariates	2	2	6	1000	Multidimensional Rasch + item covariates
			10	1000	
			50	1000	
			100	1000	
Multi-D LLLA person covariates	2	2	6	1000	Multidimensional Rasch + person covariates
			10	1000	
			50	1000	
			100	1000	
Multi-D, Polytomous					
Multi-D LLLA item covariates	2	3	6	1000	Multidimensional Polytomous Rasch + item covariates
			10	1000	
			50	1000	
			100	1000	
Multi-D LLLA person covariates	2	3	6	1000	Multidimensional polytomous Rasch + person covariates
			10	1000	
			50	1000	
			100	1000	

is expected that the PLE will be close to the MLE, and they both will be close to the true parameters, because PLE and MLE are consistent estimators. Note that the data in the simulation studies were not generated exactly from LLLA models because the person's ability was not generated from a mixture of normal distributions with constant variances. For large number of items it will not cause problems because according to the central limit theorem for posterior distribution (Dawid, 1970) the conditional distribution of the latent variable given the manifest variables will be close to normal distribution, as assumed by LLLA models. For small number of items, it is desirable to see how close are the parameters estimates from the LLLA procedures to the true parameters in the Rasch models. Such comparisons have been done for unidimensional Rasch and 2PL models for dichotomous items (Anderson & Yu, 2007) and Rasch models (uni- and multidimensional) for binary and polytomous items (Anderson, Li, & Vermunt, 2007), and the parameter recovery was excellent even for small item tests. It is interesting to see how it will be when the covariates are present in the model. Similar results are expected for the models with covariates for both dichotomous and polytomous items.

It is important to accurately estimate the SE in pseudolikelihood estimation procedures because the construction of confidence intervals and statistical inference depend on correctly estimated standard errors. Theoretically, the SE for PLE will be larger than the SE for MLE, and I would like to determine the relative efficiency of PLE to MLE in the simulated data. The comparison of relative efficiency will only be conducted for tests with small numbers of items because it is not feasible to get MLE for tests with large numbers of items.

Unidimensional Dichotomous Models

LLLA with item covariates.

Simulating data from the LLTM. Tests of different lengths were simulated and there were $I = 5, 10, 50$ and 100 items in these tests. The response matrices were generated from the LLTM. The item difficulties were calculated as the linear combinations of 4 item covariates with prespecified coefficients,

$$b_i = \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i4}.$$

The item covariates X_{i1} , X_{i2} , X_{i3} , and X_{i4} were drawn from a normal distribution with mean 0 and sd 0.2. The item covariate effects were assigned values $\beta_1 = 1$, $\beta_2 = 2$, $\beta_3 = 3$ and $\beta_4 = 4$.

One thousand persons were simulated with their ability generated from the standard normal distribution. Then the dichotomous responses were generated from the Rasch model.

Fitting LLLA with item covariates model. The LLLA with item covariates (LLLAi) model is fitted to the simulated data sets. Pseudolikelihood estimation is used to obtain the estimates and the robust estimate of standard errors (“sandwich” estimator) is used. Maximum likelihood estimation is also conducted for the two short tests ($I = 5$ and 10).

Comparison of PLE to MLE. Table 8 shows the results of ML estimates and PL estimates for fitting the LLLAi model on the simulated tests with 5 items and 10 items. The table also reports relative efficiencies of PLE calculated by the ratio of SE_{ml}/SE_{pl} . The estimated parameters given by MLE and PLE are very close to each other. This closeness is also evident by plotting the PL vs ML estimates of the item covariate effects, as the resulting points are almost exactly on the 45 degree lines (Figure 13). The robust standard errors given by PLE are also seen to be close to the standard errors by MLE, as the relative

efficiencies are close to 1. Thus the loss of efficiency of PLE relative to the MLE is very low.

Table 8

MLE and PLE Obtained From Fitting the LLLAi Model on Simulated Data With 5 and 10 Items

(a) 5 items					
	MLE		PLE		Relative Efficiency
	Estimate	SE	Estimate	SE	
λ	3.962	0.109			
β_0	-1.152	0.098	-1.150	0.098	1.001
β_1	0.955	0.263	0.953	0.270	0.974
β_2	1.940	0.315	1.936	0.314	1.004
β_3	2.717	0.526	2.723	0.514	1.022
β_4	4.087	0.376	4.080	0.377	0.996
σ_0^2	0.557	0.039	0.556	0.039	1.008

(b) 10 items					
	MLE		PLE		Relative Efficiency
	Estimate	SE	Estimate	SE	
λ	1.288	0.199			
β_0	-2.081	0.080	-2.082	0.079	1.011
β_1	0.736	0.172	0.735	0.175	0.982
β_2	2.013	0.221	2.008	0.231	0.959
β_3	2.665	0.182	2.660	0.181	1.005
β_4	3.782	0.152	3.778	0.156	0.970
σ_0^2	0.342	0.012	0.342	0.012	0.988

Comparing LLLAi estimates to true parameters. In Figure 14, estimated item covariate effects are plotted against the corresponding true values of the parameters used in the simulations with 5, 10, 50 and 100 items. In each plot, the x-axis represents true values ($\beta = 1, 2, 3, 4$ for four item covariates), and the y-axis represents PL estimates. For each PL estimated effect, the 95% confidence interval (CI) ($\hat{\beta}_i \pm 1.96SE$) is plotted as vertical bars around the estimated value. In all four plots, the PL estimates are close to the 45 degree line, and the 95% CIs cover the true values, indicating that the true parameters are successfully recovered by the PLE. Note that the length of the 95% CIs decreases as the number of items increases. In LLTM, item difficulty is modeled as the linear combination

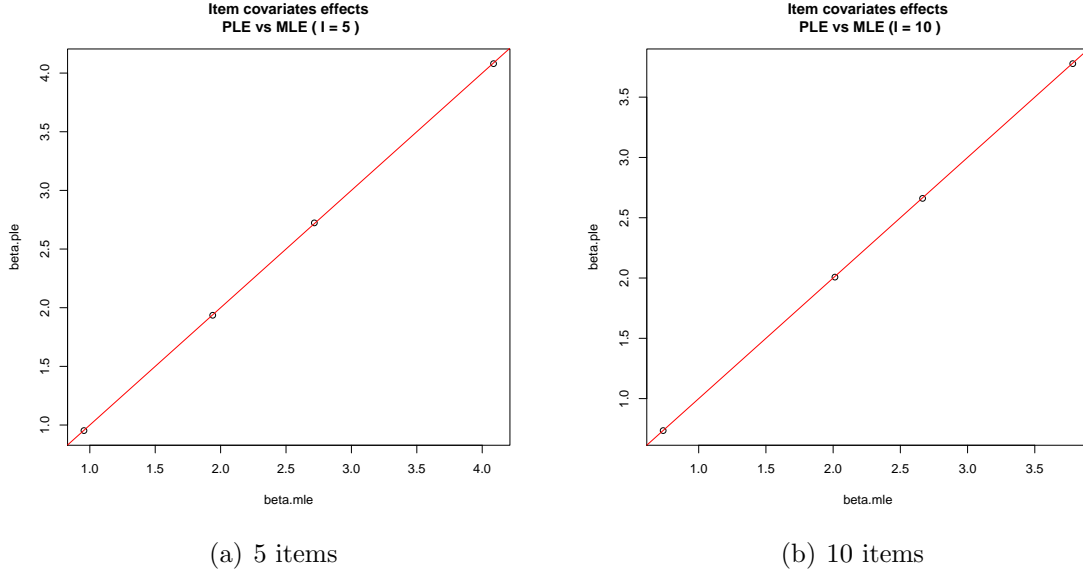
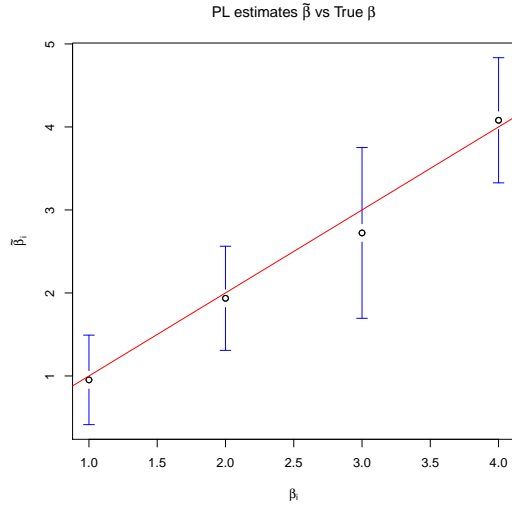


Figure 13. PLE vs MLE for item covariates effects. Test length $I = 5$ and 10 items.

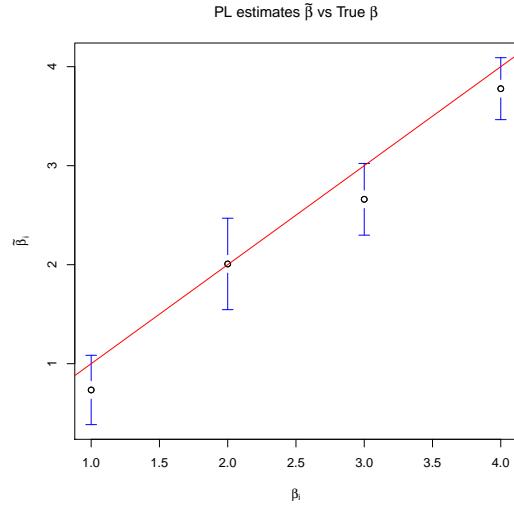
of item covariates; so it is similar to a linear regression model where the number of items is the sample size. As the number of items increases, there is more information for estimating item covariate effects; therefore, we see smaller SE as the number of items increases.

Comparing SE estimates to true SE by Monte Carlo simulations. To confirm that the SE's estimated by the robust or "sandwich" estimator is correct, they are compared with the true SE's through the Monte Carlo simulations. The true SE's are calculated by repeatedly simulating data sets from the same model, and applying the PLE procedure to the data sets. The standard deviation of the parameter estimates obtained from these replicated data sets will be the true SE's.

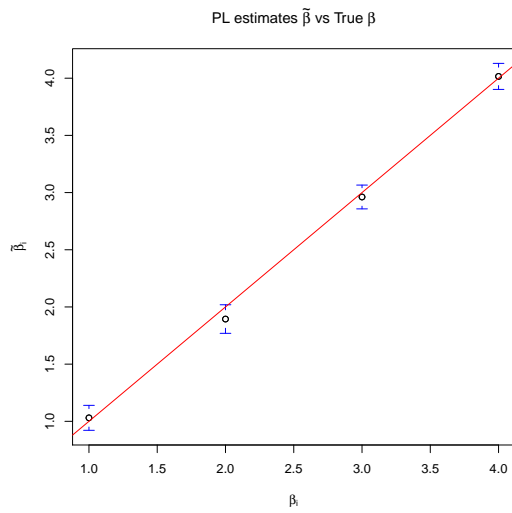
Table 9 presents the results for 10,000 replicated data sets for each of the models with test lengths $I = 5, 10, 50$, and 100 items. For each model, I started with the same set of items that were generated previously, so the item covariates and their effects were kept the same for the replicated data. For each replicate, a new set of 1000 examinees' θ 's were drawn from the standard normal distribution and a response data matrix was simulated from the Rasch model. By fitting the LLLAi model with PLE on each replicated data set, the PL



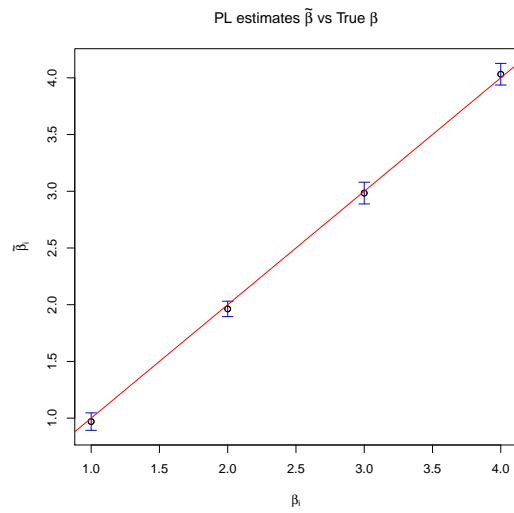
(a) 5 items



(b) 10 items



(c) 50 items



(d) 100 items

Figure 14. PLE vs true item covariate effects. Test length $I = 5, 10, 50$, and 100 items.

estimates and the robust estimates of the SE's were obtained. For each parameter, two sets of SE values are reported in the table. The first set is the average of the robust estimates of the SE's obtained from the 10,000 replicates (average PLE SE). The second set is the true SE given by the standard deviation of the PL estimates from the 10,000 replicates. We can see that the average PLE SE's are very close to true SE's. This means that the robust estimator is very effective in obtaining the correct SE's.

Table 9

Comparing the Robust Estimates (“Sandwich” Estimator) of the SE's to True SE's for the LLLAi models by Monte Carlo Simulation. Each Model was Replicated 10,000 Times

	5 items		10 items		50 items		100 items	
	PLE SE	true SE	PLE SE	true SE	PLE SE	true SE	PLE SE	true SE
β_0	0.099	0.10	0.081	0.082	0.022	0.022	0.016	0.016
β_1	0.26	0.26	0.17	0.17	0.056	0.056	0.038	0.038
β_2	0.31	0.31	0.23	0.23	0.061	0.062	0.034	0.035
β_3	0.53	0.52	0.19	0.19	0.053	0.052	0.048	0.048
β_4	0.37	0.37	0.15	0.16	0.055	0.055	0.047	0.047
σ_0^2	0.040	0.040	0.012	0.012	0.00099	0.00099	0.00033	0.00033

LLLA with person covariates.

Simulating data from the latent regression Rasch model. Four tests with different numbers of items were simulated, with two short tests of length $I = 5$ and 10, and two long tests of length 50 and 100. The item difficulty parameters $b_i, i = 1, \dots, I$ were drawn from the standard normal distribution. For each test, responses for 1000 persons were simulated. For each person, 4 person covariates Z_{p1}, Z_{p2}, Z_{p3} and Z_{p4} were drawn from the standard normal distribution and recorded. The linear combination

$$\mu_p = \gamma_1 Z_{p1} + \gamma_2 Z_{p2} + \gamma_3 Z_{p3} + \gamma_4 Z_{p4}$$

were calculated. The person covariate effects $(\gamma_1, \gamma_2, \gamma_3, \gamma_4)$ were assigned values (0.2, 0.4, 0.6, 0.8). Each person's ability was sampled from the normal distribution: $\theta_p \sim N(\mu_p, 1)$,

$p = 1, \dots, 1000$. Then the probability of correct answering the items was calculated from the Rasch IRF and the dichotomous responses are generated according to the probability.

Comparing PLE to MLE. It is not feasible to fit the LLLA model with continuous person covariates on the simulated data, even with small number of items of 5 or 10. Therefore here I will not compare PLE with MLE.

Comparing PLE to true parameters: person covariate effects. A major issue in simulation studies is how well the estimated parameters recover the true parameters. “Recovering” the true parameters means that the estimated parameters should be close to the true parameters, and the corresponding 95% confidence interval should cover the true value most of the time.

In this simulation study, we will look at the parameter estimates by PLE for item parameters, person covariate effects, and person parameters. Before we look into the results, I would like to remind that the model we used to fit the data (the LLLA with person covariates model, or for short the LLLAp model) is different from the model we used for simulating the data (the latent regression Rasch model). So when we do the comparison of the estimated parameters in the LLLA model with the true parameters in the latent regression Rasch model, what we will face is often a linear transform relationship between the two sets of values.

What is special in the LLLAp model, in comparison to the LLLA model, is the inclusion of the person covariate effects (γ), so we will first look at the PLE estimates ($\hat{\gamma}$)

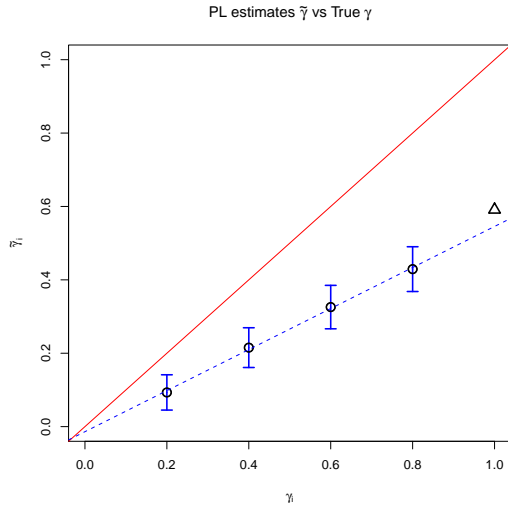
In Figure 15, PL estimated person covariate effects in the LLLAp model ($\hat{\gamma}_0$) are plotted against the corresponding true values of the parameters γ in the latent regression Rasch model used in the simulations with 5, 10, 50 and 100 items. In each plot, the x-axis represents true values ($\gamma = 0.2, 0.4, 0.6, 0.8$ for four person covariates) in the latent regression Rasch model, and the y-axis represents PL estimates in the LLLAp model. Each circle point represents the pair $(\gamma_k, \hat{\gamma}_{0k}), k = 1, \dots, 4$.

In all four plots, the circle points are located on a line (the dashed line), but not on the 45 degree line (the solid line). This suggests that there is a linear relationship between the γ_0 parameter in the LLLAp model and the γ parameter in the linear regression Rasch model. Another relationship revealed in the plots is that as the number of test items increases, the slope of the dashed line decreases. Actually, the slope of the dashed line is closely related to the ratio of the conditional variance parameter σ_0^2 in the LLLAp model to the population variance parameter σ^2 in the latent regression Rasch model. To demonstrate this, on each figure the point $(\sigma^2, \hat{\sigma}_0^2)$ is also plotted as a triangular point. All the triangular points are located closed to the dashed line. This suggests that the slope in the linear relationship between γ_0 and γ is strongly related to σ_0^2/σ^2 .

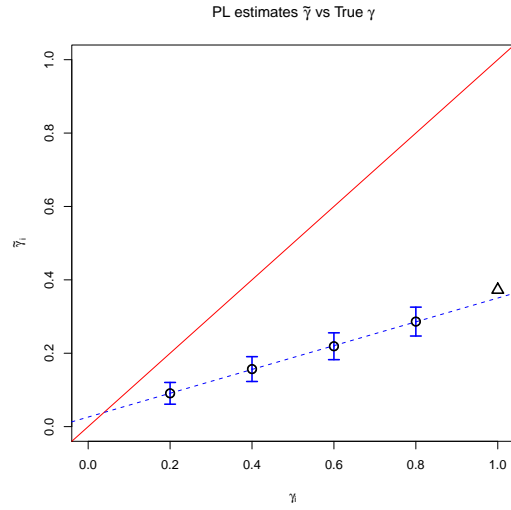
For each PL estimated effect, the 95% confidence interval ($\hat{\gamma}_k \pm 1.96SE$) is plotted as vertical bars around the estimated value. The length of the 95% CIs decreases as the number of items increases. As the number of items increases, there is more information to help obtain more precise person covariate effects estimates.

Make a three-way comparison. When studying the performance of PLE for person parameter and item parameters, in addition to comparing the PLE obtained by fitting the LLLAp model to the true parameters, I also fitted the LLLA model that ignores the person covariates to the same simulated data, and made a three-way comparison of three sets of values: (a) true parameter, (b) PLE by fitting the LLLAp model, and (c) PLE by fitting the LLLA model.

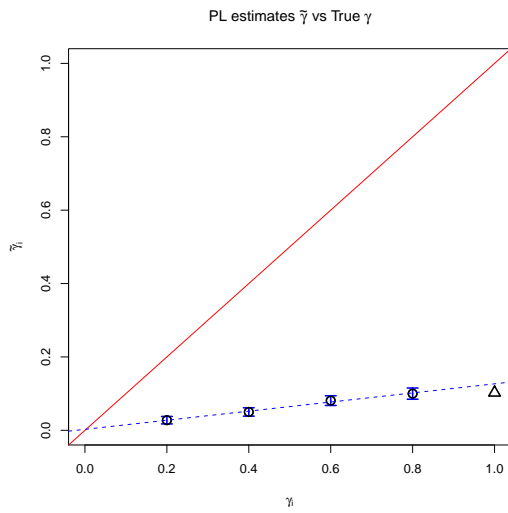
The reason to make the three-way comparison is that I would like to shift the focus from whether fitting the LLLAp model can recover the true values of item difficulty and person ability (we would expect successful recovery of true parameters, just as what we have previously seen in the LLLA/Rasch simulation study), to how the inclusion of person covariates in the model improves the performance of the estimates for these parameters (LLLAp vs LLLA problem).



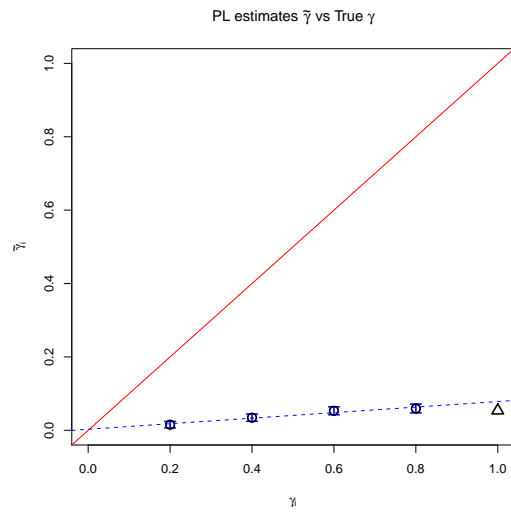
(a) 5 items



(b) 10 items



(c) 50 items



(d) 100 items

Figure 15. PLE vs true person covariate effects. Test length $I = 5, 10, 50$, and 100 items.

Another perspective gained from including the results of fitting the LLLA model on the simulated data, is the ability to study the robustness of the LLLA model. If we ignore the person covariates in the model, can we still get good estimates of the item difficulties and person's ability?

Table 10 shows the results of fitting the LLLAp model and the LLLA model along with the true parameters.

Table 10

Parameter Estimates of the LLLAp Model and the LLLA Model Along With the True Parameters for the 10-item Test

	LLLAp		LLLA		True Parameter
	Estimates	SE	Estimates	SE	
Item 1	-2.38	0.10	-2.74	0.09	-0.85
Item 2	-1.21	0.09	-1.57	0.08	0.27
Item 3	-1.60	0.09	-1.96	0.08	-0.18
Item 4	-1.65	0.09	-2.01	0.08	-0.26
Item 5	-0.21	0.08	-0.58	0.08	1.33
Item 6	-3.26	0.11	-3.62	0.11	-1.85
Item 7	-1.22	0.08	-1.58	0.08	0.19
Item 8	-0.99	0.09	-1.35	0.08	0.44
Item 9	-3.16	0.11	-3.52	0.10	-1.46
Item 10	-2.13	0.09	-2.49	0.09	-0.75
σ^2	0.37	0.01	0.46	0.01	1
γ_1	0.09	0.01			0.2
γ_2	0.16	0.02			0.4
γ_3	0.22	0.02			0.6
γ_4	0.29	0.02			0.8

Three-way comparison: item parameters. We compare the estimated item effects in the LLLAp model vs in the LLLA model (Figure 16, left) and the corresponding SEs (Figure 16, right). The points are not on the 45 degree lines, but obviously they are located on a line with slope 1. This indicates that the two sets of item parameter estimates are consistent, but differ by a constant.

Next I will compare how well the estimated parameters recover the true parameters

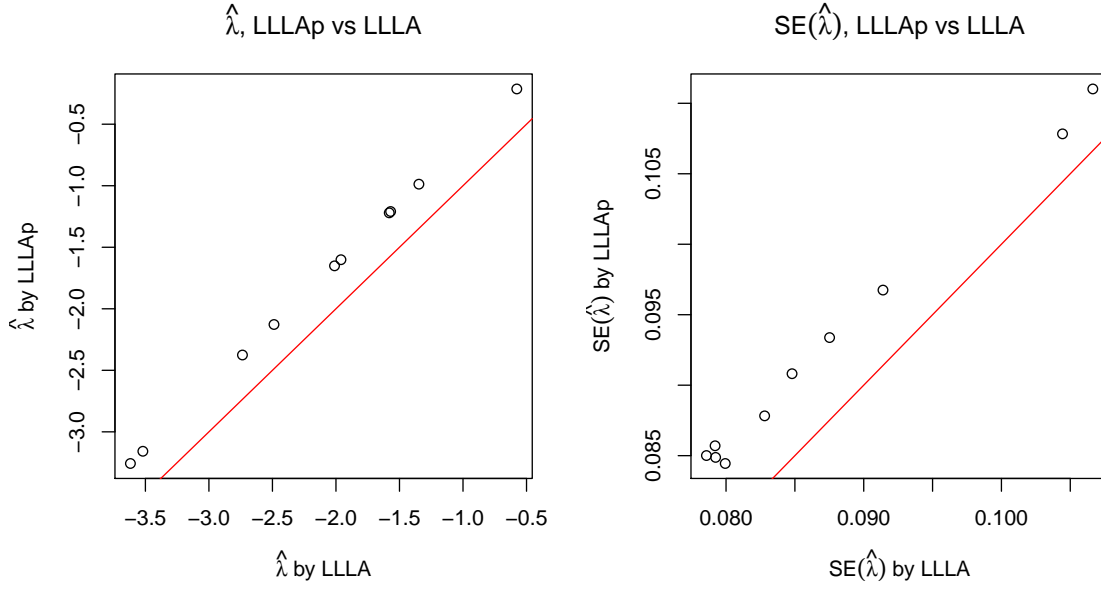


Figure 16. Parameter estimates and SEs for λ by the LLLAp and LLLA models.

for the item difficulty parameters.

As I have stated before, the item effects parameter λ_i in the LLLA model and the LLLAp model is not exactly equal to the item difficulty parameter in the Rasch model or the latent regression Rasch model, but is instead equal to the item difficulty parameter plus some constant ($\lambda_i = \beta_i + \mu_0 + 1/2\sigma_0^2$). Therefore to evaluate the performance of the item parameter estimates in the LLLA model and LLLAp model, we need to first transform the item parameter estimates $\hat{\lambda}_i$ into item difficulty parameter estimates $\hat{\beta}_i$, by

$$\hat{\beta}_i = \hat{\lambda}_i - \hat{\mu}_0 - \frac{1}{2}\hat{\sigma}_0^2$$

where $\hat{\mu}_0 = \hat{\sigma}_0^2 \bar{T}$ and \bar{T} is mean of the total scores from the persons in the data set.

Now consider the simulated data set by the latent regression Rasch model with 10 items. In this case we have three sets of values for the item difficulty parameters: (a) β_i , the true parameters used in the simulation; (b) LLLAp $\hat{\beta}_i$, estimated parameter from the LLLAp model; and (c) LLLA $\hat{\beta}_i$, estimated parameter from the LLLA model. Now we make a three way comparison of the three set of values, as shown in the scatter plots in Figure 17.

From the scatter plots we can see that both LLLA and LLLAp estimates have good recovery of the true item parameters; and the LLLA and LLLAp estimates are nearly identical to each other.

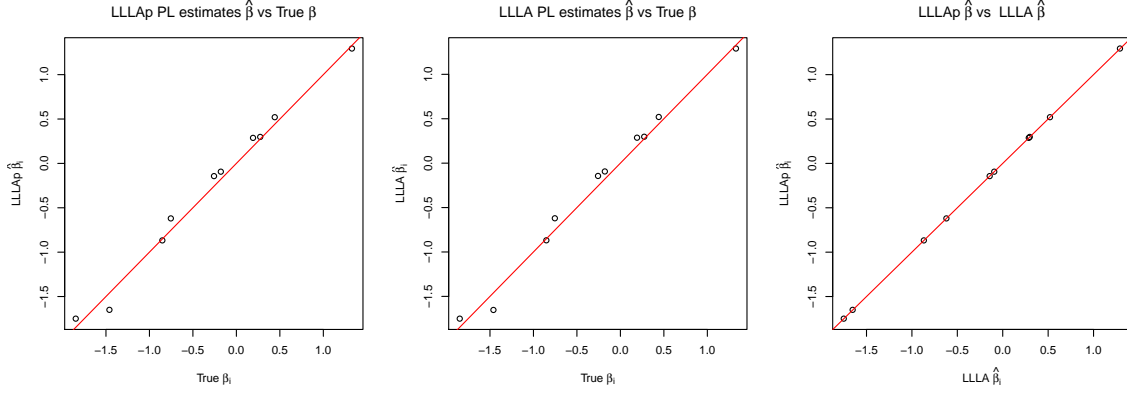


Figure 17. Three-way comparison of the true, LLLAp estimated, and LLLA estimated item difficulty parameters β_i .

The correlation between the true parameters and the LLLAp estimates for item difficulty is 0.99488, and the correlation between the true parameters and the LLLA estimates for item difficulty is 0.99479. The mean square error (MSE) of the LLLAp estimates is 0.01018, and the MSE of the LLLA estimates is 0.01020. The relative improvement in MSE by including person covariates is $(0.01020 - 0.01018)/0.01020 \times 100\% = 0.24\%$. For item parameter estimates, the LLLAp model has only a very slight improvement over the LLLA model. So in this case including person covariates into the LLLA model does not improve much on the item parameter estimates.

On the other hand, it also demonstrates the strength of the LLLA model: even in the situation where the model is misspecified by ignoring the person covariates, it will not overly influence the performance of the item parameter estimates.

Three-way comparison: person parameters. Under the two models, we estimated the person ability parameter θ for each person. In the LLLA model, the formula for estimating θ is

$$\hat{\theta} = \hat{\sigma}_0^2(T - \bar{T})$$

where T is the total score of the person.

In the LLLAp model, the formula for estimating θ is

$$\hat{\theta} = \hat{\sigma}_0^2(T - \bar{T}) + \sum_j \hat{\gamma}_j Z_j$$

where $\hat{\gamma}_j$ is the estimated effects of the j -th person covariate Z_j .

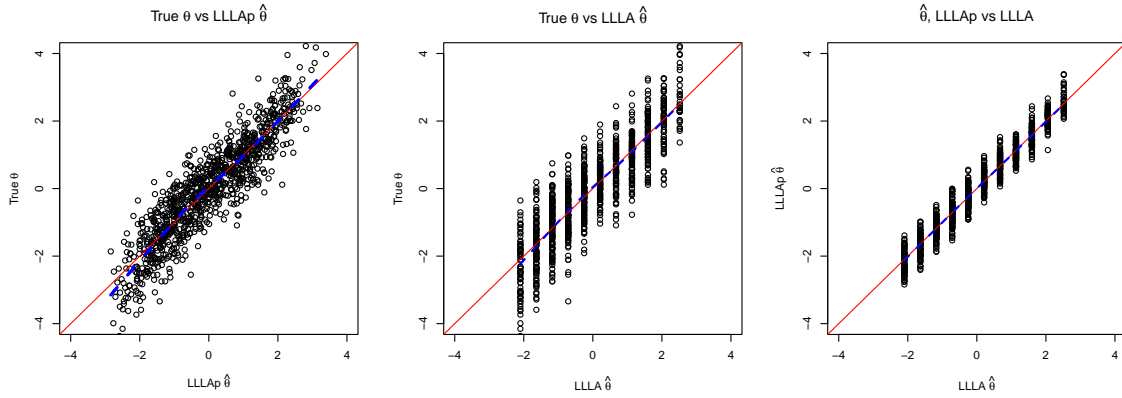


Figure 18. Three-way comparison of the true, LLLAp estimated, and LLLA estimated person ability parameters θ_p .

Figure 18 shows the results of the three-way comparison in regard to (a) true ability parameter θ used in the simulation by the latent regression Rasch model; (b) estimated ability parameter by the LLLAp model (i.e., with person covariates); and (c) estimated ability parameter by the LLLA model (i.e., without person covariates). The first two plots show the scatter plots of the true parameter vs the estimated parameters by LLLAp and LLLA, respectively.

The dashed lines in the plots are locally weighted scatterplot smoothing (lowess) curves calculated from the points in the scatter plot. The lowess curves are very close to the 45 degree lines (solid lines). This suggests that the estimated person parameters and the true parameters have the following relationship: $E(\theta|\hat{\theta}) = \hat{\theta}$.

The correlation between the true parameters and the LLLAp estimates is 0.90, and the correlation between the true parameters and the LLLA estimates is 0.87. The mean square

error (MSE) of the LLLAp estimates is 0.41, and the MSE of the LLLA estimates is 0.54. The relative improvement in MSE by including person covariates is $(0.54 - 0.41)/0.54 \times 100\% = 24\%$. It demonstrates that by correctly including person covariates into the LLLA model, better estimates of person parameters are obtained.

Comparing SE estimates to true SE by Monte Carlo simulations. To confirm that the SE's estimated by the robust or "sandwich" estimator is correct, they are compared with the true SE's through the Monte Carlo simulations. The true SE's are calculated by repeatedly simulating data sets from the same model, and applying the PLE procedure to the data sets. The standard deviation of the parameter estimates obtained from these replicated data sets will be the true SE's.

Table 11

Comparing the Robust Estimates ("Sandwich" Estimator) of the SE's to True SE's for the LLLAp models by Monte Carlo Simulation. Each Model was Replicated 10,000 Times. For the Tests With Length 10, 50 and 100 Items, the Item Effects for Only the First 5 Items

	5 items		10 items		50 items		100 items	
	PLE SE	true SE	PLE SE	true SE	PLE SE	true SE	PLE SE	true SE
$\lambda_{1(1)}$	0.12	0.12	0.097	0.096	0.083	0.083	0.081	0.081
$\lambda_{2(1)}$	0.1	0.099	0.086	0.087	0.079	0.079	0.077	0.079
$\lambda_{3(1)}$	0.11	0.11	0.089	0.090	0.080	0.079	0.078	0.078
$\lambda_{4(1)}$	0.11	0.11	0.090	0.090	0.080	0.079	0.078	0.078
$\lambda_{5(1)}$	0.095	0.095	0.086	0.085	0.083	0.084	0.081	0.081
\vdots			\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
γ_1	0.025	0.025	0.016	0.016	0.0052	0.0051	0.0042	0.0040
γ_2	0.028	0.028	0.017	0.018	0.0059	0.0058	0.0047	0.0045
γ_3	0.030	0.030	0.019	0.019	0.0065	0.0062	0.0052	0.0049
γ_4	0.032	0.032	0.021	0.021	0.0074	0.0071	0.0058	0.0054
σ_0^2	0.037	0.037	0.013	0.013	0.00099	0.00099	0.00037	0.00036

Table 11 presents the results for 10,000 replicated data sets for each of the models with test lengths $I = 5, 10, 50$, and 100 items. For each LLLAp model, The same set of items, person covariates, and person covariate effects were kept the same for all the replicated data.

For each replicate, a new set of 1000 examinees' θ 's were drawn from the standard normal distribution and a response data matrix was simulated from the Rasch model. By fitting the LLLAp model with PLE on each replicated data set, the PL estimates and the robust estimates of the SE's were obtained. For each parameter, two sets of SE values are reported in the table. The first set is the average of the robust estimates of the SE's obtained from the 10,000 replicates (average PLE SE). The second set is the true SE given by the standard deviation of the PL estimates from the 10,000 replicates. We can see that the average PLE SE's are very close to true SE's. This means that the robust estimator is very effective in obtaining the correct SE's.

Unidimensional Polytomous Models

Polytomous LLLA with item covariates.

Simulating data from partial credit model with item covariates. The response matrices were simulated from the partial credit model with item parameters modeled as linear combinations of item covariates. Tests with different lengths $I = 5, 10, 50$ and 100 were simulated. Each item had 3 possible outcomes, recorded as 0, 1, and 2. For each item, the item parameters b_{i1} and b_{i2} were calculated as the linear combination of item covariates:

$$b_{i1} = \beta_{11}X_{i1} + \beta_{21}X_{i2} + \beta_{31}X_{i3} + \beta_{41}X_{i4},$$

$$b_{i2} = \beta_{12}X_{i1} + \beta_{22}X_{i2} + \beta_{32}X_{i3} + \beta_{42}X_{i4}.$$

The item covariates X_{i1} , X_{i2} , X_{i3} and X_{i4} took values from a normal distribution with mean 0 and sd 0.2. The item covariate effects β_{jh} , $j = 1, 2, 3, 4$, $h = 1, 2$, were assigned values $\{\beta_{j1}\} = (\beta_{11}, \beta_{21}, \beta_{31}, \beta_{41}) = (1, 2, 3, 4)$, and $\{\beta_{j2}\} = (\beta_{12}, \beta_{22}, \beta_{32}, \beta_{42}) = (5, 6, 7, 8)$. The abilities of 1000 persons were drawn from the standard normal distribution: $\theta_p \sim N(0, 1)$, $p = 1, \dots, 1000$. With the values of the person parameters θ_p and item parameters b_{i1} and

b_{i2} , the probability of choosing one of the three responses $(P_{pi0}, P_{pi1}, P_{pi2})$ were calculated from the item category response functions of the partial credit model

$$P_{pih} = P(Y_{pi} = h | \theta_p, b_{i1}, b_{i2}) = \frac{\exp(h\theta_p + b_{ih})}{\sum_{l=0}^2 \exp(l\theta_p + b_{il})}, \quad h = 0, 1, 2,$$

where b_{i0} always takes value 0. The response Y_{pi} , which takes on one of the possible values 0, 1 and 2, was generated with the corresponding probabilities $(P_{pi0}, P_{pi1}, P_{pi2})$.

Comparing PLE to MLE. Table 12 shows the results of ML estimates and PL estimates for fitting the LLLAi model on the simulated data with test lengths of 5 items and 10 items. The table also reports estimated relative efficiencies of PLE calculated by the ratio of SE_{ml}/SE_{pl} . The estimated parameters given by MLE and PLE are very close to each other. This closeness is also evident by plotting the PL vs ML estimates of the item covariate effects, as the resulting points are almost exactly on the 45 degree lines (Figure). The robust standard errors given by PLE are close to the standard errors by MLE, as the relative efficiencies are close to 1. Thus the loss of efficiency of PLE relative to the MLE is very low.

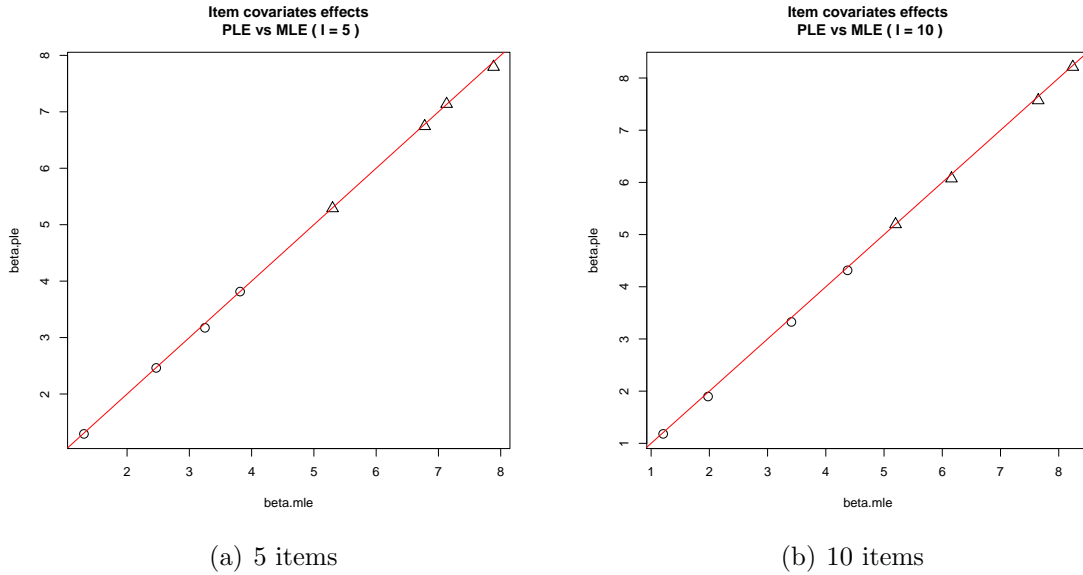


Figure 19. PLE vs MLE for item covariates effects. Test length $I = 5$ and 10 items.

Table 12

MLE and PLE Obtained From Fitting the LLLAi Model on Simulated (Polytomous) Data With 5 and 10 Items

(a) 5 items					
	MLE		PLE		Relative Efficiency
	Estimate	SE	Estimate	SE	
λ	2.840	0.166			
β_{01}	-1.664	0.090	-1.654	0.087	1.041
β_{02}	-2.976	0.166	-2.961	0.163	1.017
β_{11}	1.308	0.355	1.293	0.355	1.001
β_{12}	5.300	0.383	5.289	0.394	0.973
β_{21}	2.468	0.422	2.462	0.415	1.016
β_{22}	6.779	0.514	6.745	0.502	1.025
β_{31}	3.251	0.936	3.173	0.943	0.992
β_{32}	7.885	0.985	7.795	0.994	0.991
β_{41}	3.815	0.423	3.815	0.402	1.051
β_{42}	7.131	0.755	7.139	0.749	1.009
σ_0^2	0.375	0.017	0.373	0.017	1.014
(b) 10 items					
	MLE		PLE		Relative Efficiency
	Estimate	SE	Estimate	SE	
λ	0.142	0.247			
β_{01}	-2.647	0.072	-2.648	0.071	1.006
β_{02}	-5.096	0.126	-5.118	0.125	1.003
β_{11}	1.207	0.261	1.181	0.272	0.961
β_{12}	5.198	0.258	5.197	0.266	0.970
β_{21}	1.978	0.382	1.897	0.378	1.011
β_{22}	6.158	0.372	6.074	0.383	0.971
β_{31}	3.409	0.326	3.326	0.324	1.007
β_{32}	7.647	0.329	7.571	0.334	0.985
β_{41}	4.374	0.241	4.316	0.240	1.004
β_{42}	8.241	0.254	8.212	0.249	1.019
σ_0^2	0.197	0.004	0.199	0.004	0.988

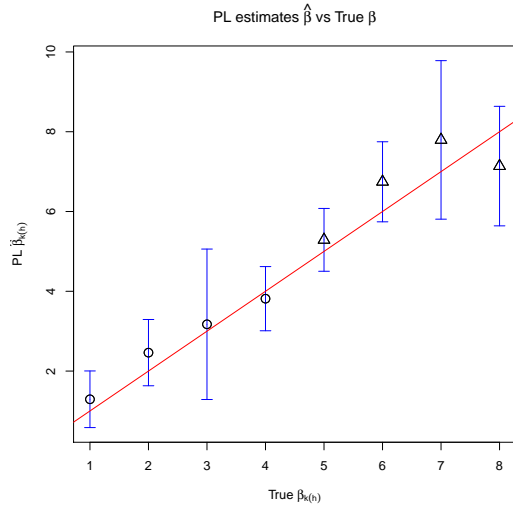
Comparing LLLA estimates to PCMi true parameters. In Figure 20, estimated item covariate effects are plotted against the corresponding true values of the parameters used in the simulations with 5, 10, 50 and 100 items. In each plot, the x-axis represents true values and the y-axis represents PL estimates. There are 8 item parameters, and for each item there are two parameter corresponding to responses 1 (circular points) and 2 (triangular points). For item covariate effect parameters, the 95% confidence intervals ($\hat{\beta}_k^* \pm 1.96SE$) are plotted as vertical bars around each estimated value. In all four plots, the PL estimates are close to the 45 degree line, and the 95% CIs cover the true values. This indicates that true parameters are successfully recovered by the PLE. Note that the length of the 95% CIs decreases as the number of items increases. As the number of items increases, there is more information for estimating item covariate effects. Therefore we see smaller SE as the number of items increases.

Polytomous LLLA with person covariates .

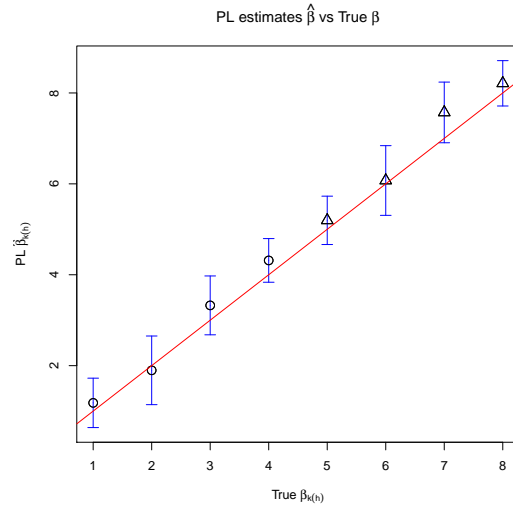
Simulating data from partial credit model with person covariates. The response matrices were simulated from the partial credit model modified to incorporate person covariates (PCMp model). Tests with different lengths $I = 5, 10, 50$ and 100 were simulated. Each item had 3 possible outcomes, recorded as 0, 1, and 2. For each item, the item parameters β_{i1} and β_{i2} were assigned values generated from the standard normal distribution. For each test, responses for 1000 persons were simulated. For each person, 4 person covariates Z_{p1} , Z_{p2} , Z_{p3} and Z_{p4} were drawn from the standard normal distribution and recorded. The linear combination

$$\mu_p = \gamma_1 Z_{p1} + \gamma_2 Z_{p2} + \gamma_3 Z_{p3} + \gamma_4 Z_{p4}$$

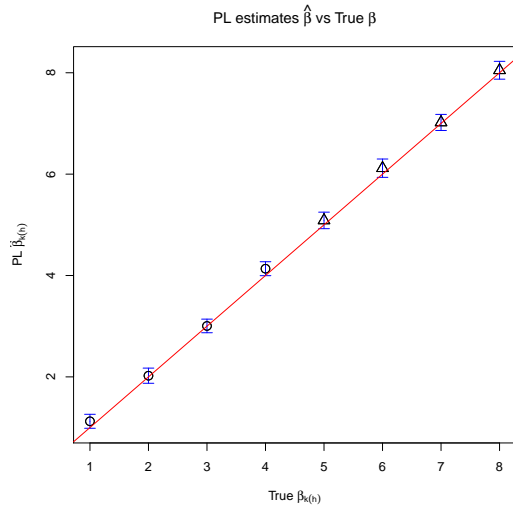
were calculated. The person covariate effects $(\gamma_1, \gamma_2, \gamma_3, \gamma_4)$ were assigned values $(0.2, 0.4, 0.6, 0.8)$. Each person's ability was sampled from the normal distribution: $\theta_p \sim N(\mu_p, 1)$, $p =$



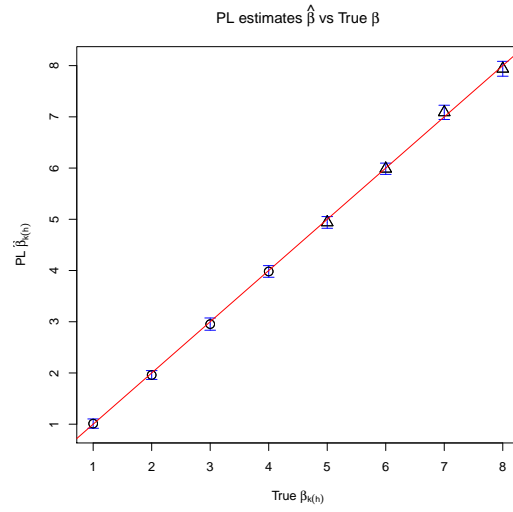
(a) 5 items



(b) 10 items



(c) 50 items



(d) 100 items

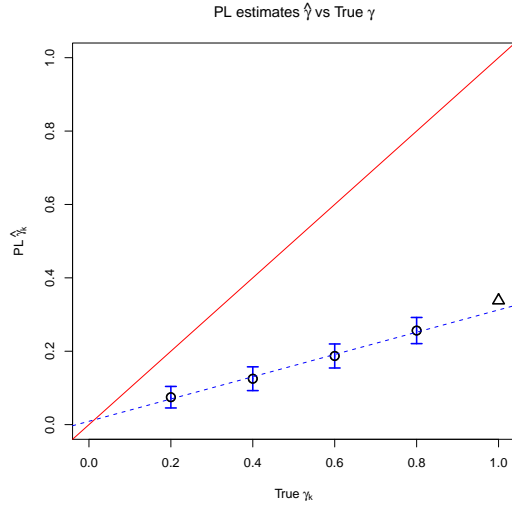
Figure 20. PLE vs true item covariate effects. Test length $I = 5, 10, 50$, and 100 (polytomous) items.

$1, \dots, 1000$. For every combination of persons and items, the probability of choosing one of the three responses ($P_{pi0}, P_{pi1}, P_{pi2}$) were calculated from the partial credit model item category response functions, and the response Y_{pi} , which takes values 0, 1, or 2, was generated according to these probabilities.

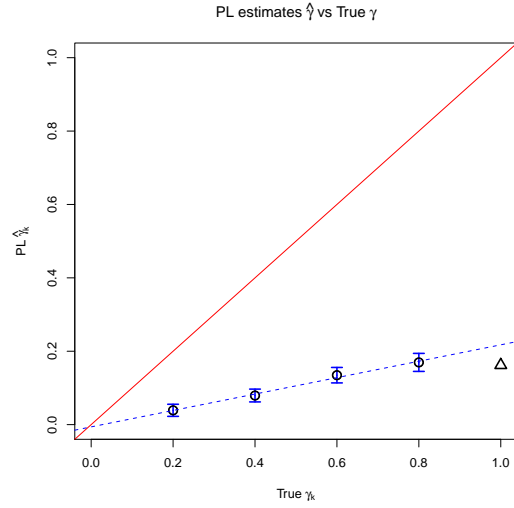
Comparing PLE to true parameters: person covariate effects. In Figure 21, PL estimated person covariate effects in the LLLAp model ($\hat{\gamma}_0$) are plotted against the corresponding true values of the parameters γ in the latent regression Rasch model used in the simulations with 5, 10, 50 and 100 items. In each plot, the x-axis represents true values ($\gamma = 0.2, 0.4, 0.6, 0.8$ for four person covariates) in the latent regression Rasch model, and the y-axis represents PL estimates in the LLLAp model. Each circular point represents the pair $(\gamma_k, \hat{\gamma}_{0k}), k = 1, \dots, 4$.

In all four plots, the circle points are located on a line (the dashed line), but not on the 45 degree line (the solid line). This suggests that there is a linear relationship between the γ_0 parameter in the LLLAp model and the γ parameter in the linear regression Rasch model. The plots also revealed that as the number of test items increases, the slope of the dashed line decreases. Actually, the slope of the dashed line is closely related to the ratio of the conditional variance parameter σ_0^2 in the LLLAp model to the population variance parameter σ^2 in the PCMp model. To demonstrate this, on each figure the point $(\sigma^2, \hat{\sigma}_0^2)$ is also plotted as a triangular point. All the triangular points are located closed to the dashed line. This suggests that the slope in the linear relationship between γ_0 and γ is strongly related to σ_0^2/σ^2 .

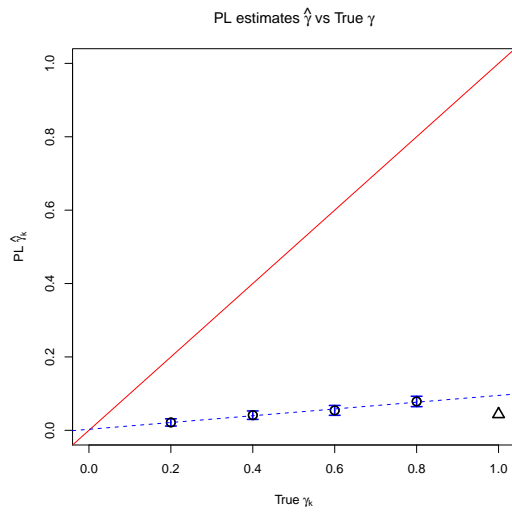
For each PL estimated effect, the 95% confidence interval ($\hat{\gamma} \pm 1.96SE$) is plotted as vertical bars around the estimated value. The length of the 95% CIs decreases as the number of items increases. As the number of items increases, there is more information to help obtain more precise person covariate effects estimates.



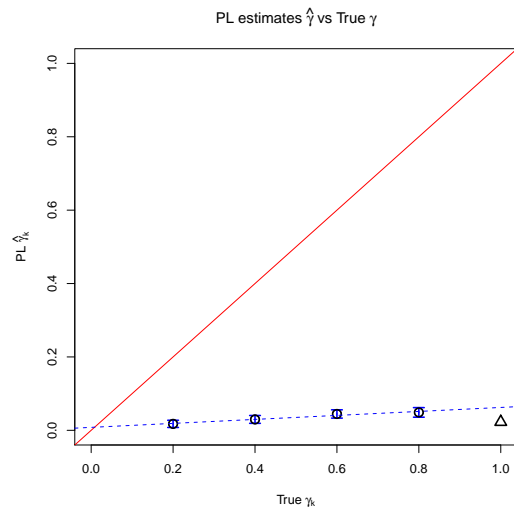
(a) 5 items



(b) 10 items



(c) 50 items



(d) 100 items

Figure 21. PLE vs true person covariate effects. Test length $I = 5, 10, 50$, and 100 items.

Make a three-way comparison. When studying the performance of PLE for person parameter and item parameters, in addition to comparing the PLE obtained by fitting the LLLAp model to the true parameters, I also fitted the LLLA model that ignores the person covariates to the same simulated data, and made a three-way comparison of three sets of values: (a) true parameter, (b) PLE by fitting the LLLAp model, and (c) PLE by fitting the LLLA model.

Table 13 shows the results of fitting the LLLAp model and the LLLA model along with the true parameters.

Three-way comparison: item parameters. We compare the estimated item effects in the LLLAp model vs in the LLLA model (Figure 22, left) and the corresponding SEs (Figure 22, right). The points are not on the 45 degree lines, but obviously they are located on a line with slope 1. This indicates that the two sets of item parameter estimates are consistent, but differ by a constant.

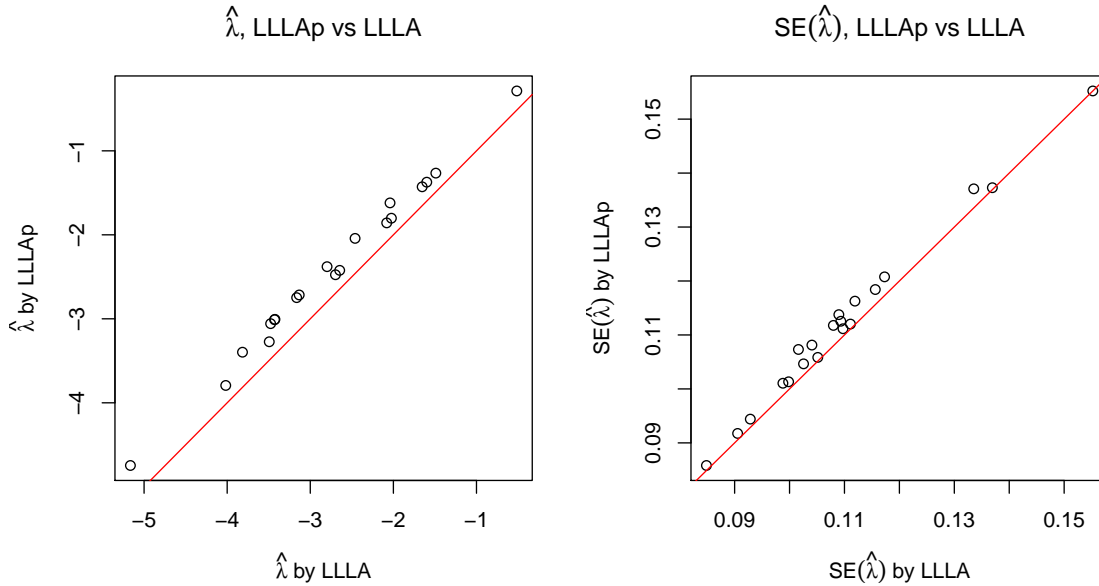


Figure 22. Parameter estimates and SEs for λ by the LLLAp and LLLA models.

Next I will compare how well the estimated parameters recover the true parameters for the item difficulty parameters.

Table 13

Parameter Estimates of the LLLAp Model and the LLLA Model Along With the True Parameters for the 10-item (Polytomous) Test

	LLLAp		LLLA		True Parameter
	Estimates	SE	Estimates	SE	
Item 1.1	-2.48	0.11	-2.70	0.11	-0.85
Item 2.1	-1.43	0.10	-1.65	0.10	0.27
Item 3.1	-1.80	0.09	-2.02	0.09	-0.18
Item 4.1	-1.86	0.10	-2.08	0.10	-0.26
Item 5.1	-0.29	0.09	-0.51	0.08	1.33
Item 6.1	-3.79	0.14	-4.02	0.14	-1.85
Item 7.1	-1.37	0.11	-1.60	0.11	0.19
Item 8.1	-1.27	0.09	-1.49	0.09	0.44
Item 9.1	-3.27	0.16	-3.49	0.16	-1.46
Item 10.1	-2.42	0.11	-2.65	0.11	-0.75
Item 1.2	-3.01	0.12	-3.43	0.11	0.18
Item 2.2	-2.04	0.11	-2.46	0.10	1.07
Item 3.2	-3.40	0.12	-3.82	0.12	-0.15
Item 4.2	-2.72	0.11	-3.13	0.11	0.35
Item 5.2	-3.01	0.12	-3.43	0.12	-0.12
Item 6.2	-4.75	0.14	-5.16	0.13	-1.61
Item 7.2	-1.62	0.10	-2.04	0.10	1.34
Item 8.2	-2.38	0.11	-2.80	0.11	0.66
Item 9.2	-2.75	0.11	-3.16	0.10	0.44
Item 10.2	-3.06	0.11	-3.48	0.11	0.07
σ^2	0.16	0.00	0.18	0.00	1.00
γ_1	0.04	0.01			0.20
γ_2	0.08	0.01			0.40
γ_3	0.13	0.01			0.60
γ_4	0.17	0.01			0.80

The item effects parameter $\lambda_{i(h)}$ in the LLLA model and the LLLAp model is not exactly equal to the item difficulty parameter in the PCM model or the PCMp model, but is instead equal to the item difficulty parameter plus some constant ($\lambda_{i(h)} = \beta_{i(h)} + h\mu_0 + 1/2h^2\sigma_0^2$). Therefore to evaluate the performance of the item parameter estimates in the LLLA model and LLLAp model, we need to first transform the item parameter estimates $\hat{\lambda}_{i(h)}$ into item difficulty parameter estimates $\hat{\beta}_{i(h)}$, by

$$\hat{\beta}_{i(h)} = \hat{\lambda}_i - h\hat{\mu}_0 - \frac{1}{2}h^2\hat{\sigma}_0^2,$$

where $\hat{\mu}_0 = \hat{\sigma}_0^2\bar{T}$ and \bar{T} is mean of the total scores from the persons in the data set.

Now consider the simulated data set by the PCMp model with 10 items. In this case there are three sets of values for the item difficulty parameters: (a) $\beta_{i(h)}$, the true parameters used in the simulation; (b) LLLAp $\hat{\beta}_{i(h)}$, estimated parameter from the LLLAp model; and (c) LLLA $\hat{\beta}_{i(h)}$, estimated parameter from the LLLA model. A three way comparison of the threes set of values is shown in the scatter plots in Figure 23. From the scatter plots we can see that both LLLA and LLLAp estimates have good recovery of the true item parameters; and the LLLA and LLLAp estimates are nearly identical to each other.

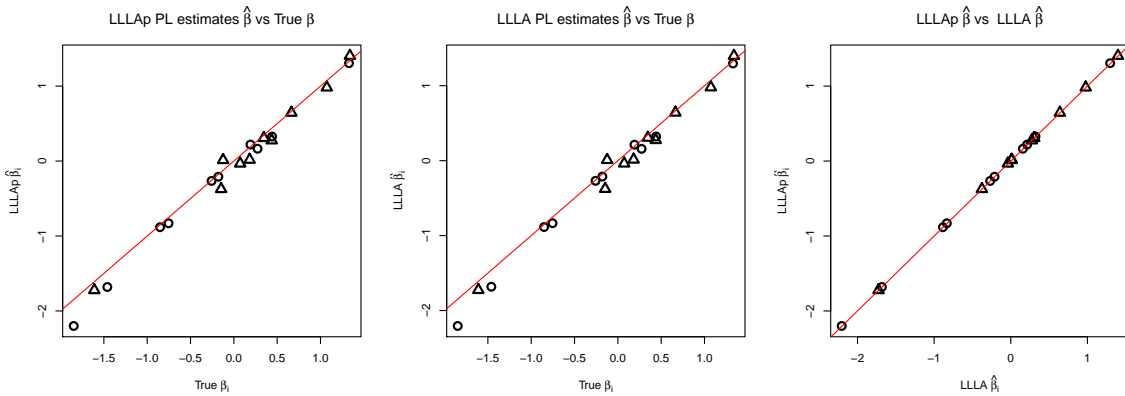


Figure 23. Three-way comparison of the true, LLLAp estimated, and LLLA estimated item difficulty parameters $\beta_{i(h)}$.

The correlation between the true parameters and the LLLAp estimates for item diffi-

culty is 0.994993, and the correlation between the true parameters and the LLLA estimates for item difficulty is 0.994998. The mean square error (MSE) of the LLLAp estimates is 0.0184, and the MSE of the LLLA estimates is 0.0190. The relative improvement in MSE by including person covariates is $(0.0184 - 0.0184)/0.0190 \times 100\% = 0.33\%$. For item parameter estimates, the LLLAp model has only a very slight improvement over the LLLA model. So in this case including person covariates into the LLLA model does not improve much on the item parameter estimates.

Three-way comparison: person parameters. Under the two models, we estimated the person ability parameter θ for each person. In the LLLA model, the formula for estimating θ is

$$\hat{\theta} = \hat{\sigma}_0^2(T - \bar{T})$$

where T is the total score of the person.

In the LLLAp model, the formula for estimating θ is

$$\hat{\theta} = \hat{\sigma}_0^2(T - \bar{T}) + \sum_j \hat{\gamma}_j Z_j$$

where $\hat{\gamma}_j$ is the estimated effects of the j -th person covariate Z_j .

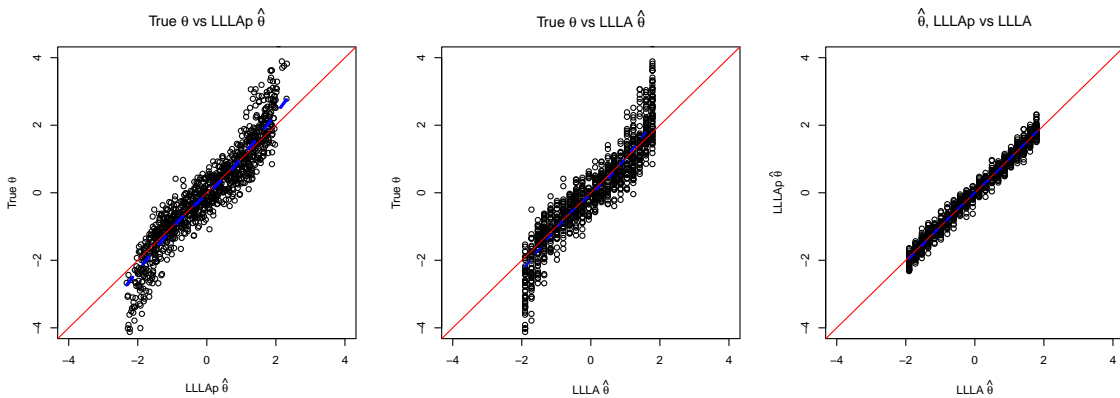


Figure 24. Three-way comparison of the true, LLLAp estimated, and LLLA estimated person ability parameters θ_p .

Figure 24 shows the results of the three-way comparison of the ability parameter θ in

(a) true parameter used in the simulation by the PCMp model; (b) estimated parameter by the LLLAp model (i.e., with person covariates); and (c) estimated parameter by the LLLA model (i.e., without person covariates). The first two plots show the scatter plots of the true parameter vs the estimated parameters by LLLAp and LLLA, respectively.

The dashed lines in the plots are locally weighted scatterplot smoothing (lowess) curves calculated from the points in the scatter plot. The lowess curves are very close to the 45 degree lines (solid lines). This suggests that the estimated person parameters and the true parameters have the following relationship: $E(\theta|\hat{\theta}) = \hat{\theta}$. Also note that in the left and middle panels, at both ends of the extreme values for θ , the lowess curves start to go off the 45 degree lines. This suggests that the measurement for θ may not be good at extreme low or high values.

The correlation between the true parameters and the LLLAp estimates is 0.935, and the correlation between the true parameters and the LLLA estimates is 0.924. The mean square error (MSE) of the LLLAp estimates is 0.305, and the MSE of the LLLA estimates is 0.354. The relative improvement in MSE by including person covariates is $(0.354 - 0.305)/0.354 \times 100\% = 14\%$. It demonstrates that by correctly including person covariates into the LLLA model, better estimates of person parameters are obtained.

Multidimensional Dichotomous Models

Multidimensional LLLA with Item covariates.

Simulating data from 2-dimensional Rasch with item covariates. Four tests with length $I = 6, 10, 50$ and 100 items were simulated. The item difficulty parameters $b_i, i = 1, \dots, I$ were calculated from linear combinations of item covariates. There were 4 item covariates for each item, and they took values from a normal distribution with mean 0

and standard deviation 0.2.

$$b_i = \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i4}.$$

The values of item effects coefficients were $(\beta_1, \beta_2, \beta_3, \beta_4) = (1, 2, 3, 4)$.

For each test, 1000 persons were simulated. For each person, the latent traits were two-dimensional: $\theta_p = (\theta_{p1}, \theta_{p2})$. The latent traits were sampled from a bivariate normal distribution with mean $(0, 0)'$ and variances 1 and correlation $\rho = 0.5$. Then the compensatory model was used to calculate the probability of getting a correct answer for each item:

$$P(Y_{pi} = 1 | \theta_{p1}, \theta_{p2}) = \frac{\exp(a_{i1}\theta_{p1} + a_{i2}\theta_{p2} - b_i)}{1 + \exp(a_{i1}\theta_{p1} + a_{i2}\theta_{p2} - b_i)}$$

The $I \times 2$ matrix $\{a_{id}\}$ is the item-trait adjacency matrix that represents the associations between the items and the latent traits. It was assumed that in the test, the first half of the items only measure the first trait θ_1 and the second half of the items only measure the second trait θ_2 . So $(a_{i1}, a_{i2}) = (1, 0)$ for the first half of the items, and $(a_{i1}, a_{i2}) = (0, 1)$ for the second half of the items.

Fitting LLLA with item covariates model. An LLLA with item covariates (LLLAi) model is fitted to the simulated data sets. The model assumes the same structure as used in the simulation (i.e., 2-dimensional latent traits, with first half of the items loading on the first latent trait, and second half of the items loading on the second latent trait). Pseudolikelihood estimation is used to obtain the estimates and the robust estimate of standard errors (“sandwich” estimator) is used. Maximum likelihood estimation is also conducted for the two short tests ($I = 6$ and 10).

Comparison of PLE to MLE. Table 14 shows the results of ML estimates and PL estimates for fitting the LLLAi model on the simulated data with test lengths of 6 items and 10 items. The table also reports estimated relative efficiencies of PLE calculated by the

ratio of SE_{ml}/SE_{pl} . The estimated parameters given by MLE and PLE are very close to each other. This closeness is also evident by plotting the PL vs ML estimates of the item covariate effects, as the resulting points are almost exactly on the 45 degree lines (Figure 25). The robust standard errors given by PLE are close to the standard errors by MLE, as the relative efficiencies are close to 1. Thus the loss of efficiency of PLE relative to the MLE is very low.

Table 14

MLE and PLE Obtained From Fitting the LLLAi 2D Model on Simulated (Polytomous) Data. Test Length $I = 6$ and 10 Items

(a) 6 items					
	MLE		PLE		Relative Efficiency
	Estimate	SE	Estimate	SE	
λ	3.312	0.129			
β_0	-1.021	0.083	-1.020	0.082	1.018
β_1	0.863	0.233	0.878	0.235	0.994
β_2	2.646	0.592	2.589	0.580	1.021
β_3	2.096	0.493	2.170	0.497	0.992
β_4	3.748	0.322	3.724	0.321	1.003
σ_{11}	0.595	0.071	0.605	0.072	0.981
σ_{12}	0.206	0.037	0.206	0.037	0.993
σ_{22}	0.674	0.071	0.665	0.071	0.993

(b) 10 items					
	MLE		PLE		Relative Efficiency
	Estimate	SE	Estimate	SE	
λ	0.254	0.248			
β_0	-1.948	0.092	-1.945	0.092	0.998
β_1	1.449	0.185	1.442	0.184	1.008
β_2	1.884	0.230	1.880	0.224	1.028
β_3	3.376	0.194	3.374	0.194	0.999
β_4	4.459	0.182	4.461	0.191	0.951
σ_{11}	0.540	0.032	0.539	0.032	1.000
σ_{12}	0.149	0.020	0.150	0.020	1.009
σ_{22}	0.459	0.025	0.455	0.025	1.002

Comparing LLLAi estimates to true parameters. In Figure 26, estimated item covariate effects are plotted against the corresponding true values of the parameters

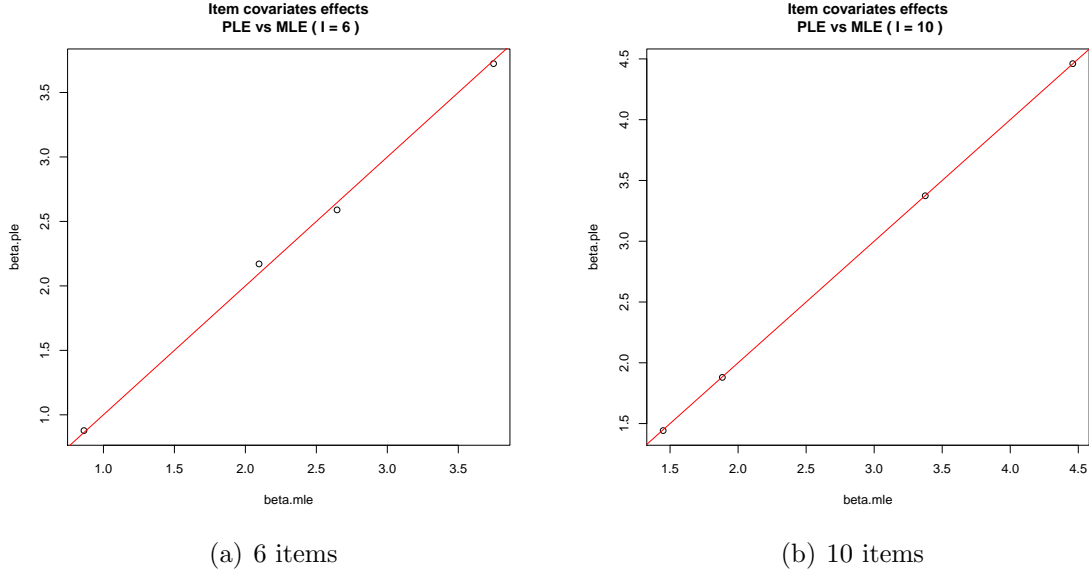


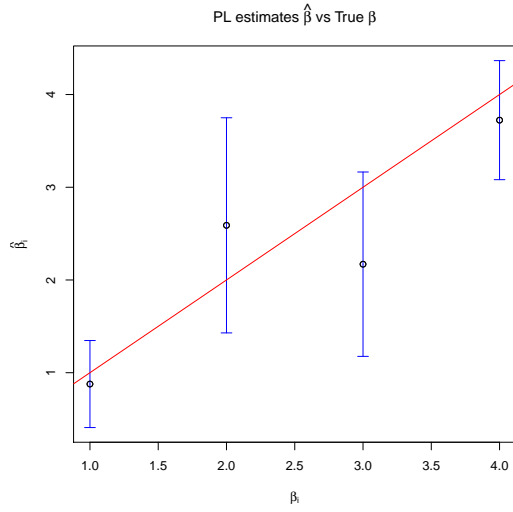
Figure 25. PLE vs MLE for item covariate effects. 2-D latent traits, Test length $I = 6$ and 10 items.

used in the simulations with 6, 10, 50 and 100 items. In each plot, the x-axis represents true values and the y-axis represents PL estimates. For each of the 4 item covariate effect parameters, the 95% confidence interval ($\hat{\beta}^* \pm 1.96SE$) is plotted as vertical bars around the estimated value. In all four plots, the PL estimates are close to the 45 degree line, and the 95% CIs cover the true values. It indicates that true parameters are successfully recovered by the PLE. Note that the lengths of the 95% CIs decrease as the number of items increases. As the number of items increases, there is more information for estimating item covariate effects. Therefore we see smaller SE as the number of items increases.

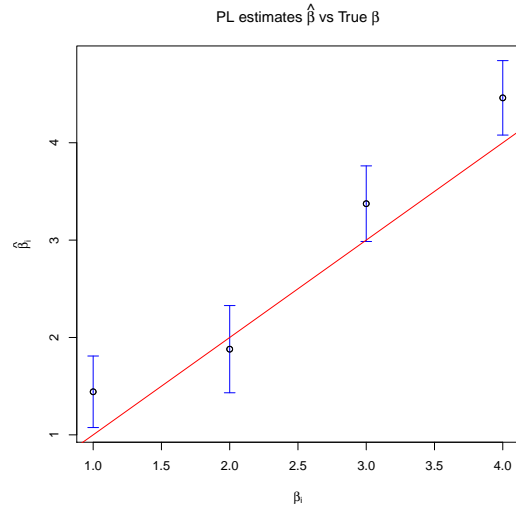
Multidimensional LLLA with person covariates.

Simulating data from 2-dimensional Rasch with person covariates. Four tests with length $I = 6, 10, 50$ and 100 items were simulated. The item difficulty parameters $b_i, i = 1, \dots, I$ were drawn from the standard normal distribution.

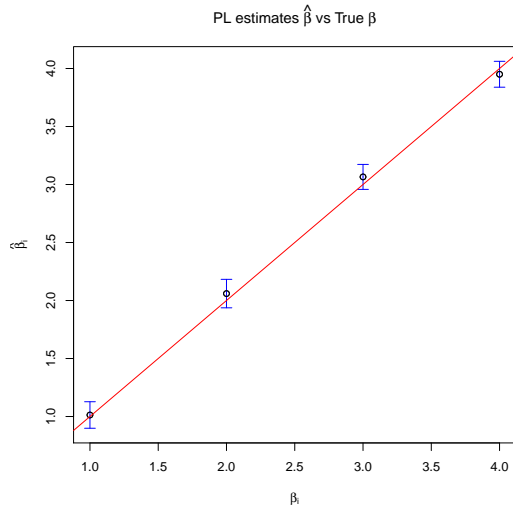
For each test, responses for 1000 persons were simulated. For each person, the two-



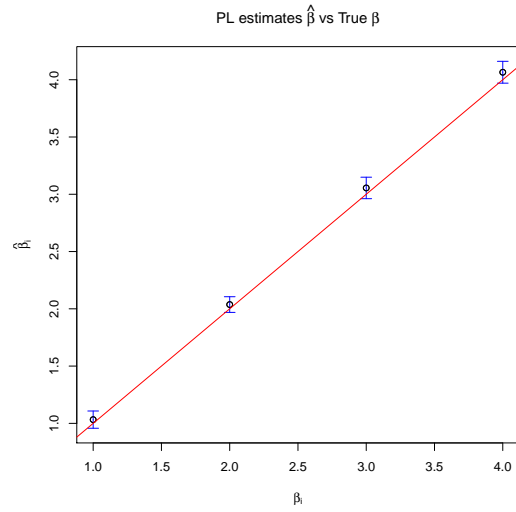
(a) 6 items



(b) 10 items



(c) 50 items



(d) 100 items

Figure 26. PLE vs true item covariate effects. Test length $I = 6, 10, 50$, and 100 items. 2-D latent traits and dichotomous response.

dimensional latent traits $\theta_p = (\theta_{p1}, \theta_{p2})$ were calculated from

$$\theta_p = Z_p \gamma + \epsilon_p.$$

The person covariate vector has 4 variables $Z_p = (Z_{p1}, Z_{p2}, Z_{p3}, Z_{p4})$ and they were all drawn from the standard normal distribution. The person covariate effects form a 4×2 matrix $\{\gamma_{jd}\}$ and were assigned the following values:

$$\begin{pmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \\ \gamma_{31} & \gamma_{32} \\ \gamma_{41} & \gamma_{42} \end{pmatrix} = \begin{pmatrix} 0.2 & 0.7 \\ 0.4 & 0.5 \\ 0.6 & 0.3 \\ 0.8 & 0.1 \end{pmatrix}$$

The error term ϵ_p was sampled from a bivariate normal distribution with mean $(0, 0)'$ and variances 1 and correlation $\rho = 0.5$. Then the compensatory model was used to calculate the probability of getting a correct answer for each item:

$$P(Y_{pi} = 1 | \theta_{p1}, \theta_{p2}) = \frac{\exp(a_{i1}\theta_{p1} + a_{i2}\theta_{p2} - b_i)}{1 + \exp(a_{i1}\theta_{p1} + a_{i2}\theta_{p2} - b_i)}$$

The $I \times 2$ matrix $\{a_{id}\}$ is the item-trait adjacency matrix that represents the associations between the items and the latent traits. It was assumed that in the test, the first half of the items only measure the first trait θ_1 and the second half of the items only measure the second trait θ_2 . So $(a_{i1}, a_{i2}) = (1, 0)$ for the first half of the items, and $(a_{i1}, a_{i2}) = (0, 1)$ for the second half of the items.

Comparing PLE to MLE. It is not possible to run the Poisson regression to fit the LLLAp model on the data set to get MLE.

Comparing PLE to true parameters: person covariate effects. In Figure 27, PL estimated person covariate effects in the LLLAp model ($\hat{\gamma}_0$) are plotted against the

corresponding true values of the parameters γ in the latent regression Rasch model used in the simulations with 6, 10, 50 and 100 items. In each plot, the x-axis represents true values ($\gamma = \{(0.2, 0.7), (0.4, 0.5), (0.6, 0.3), (0.8, 0.1)\}$ for four person covariates) in the latent regression Rasch model, and the y-axis represents PL estimates in the LLLAp model.

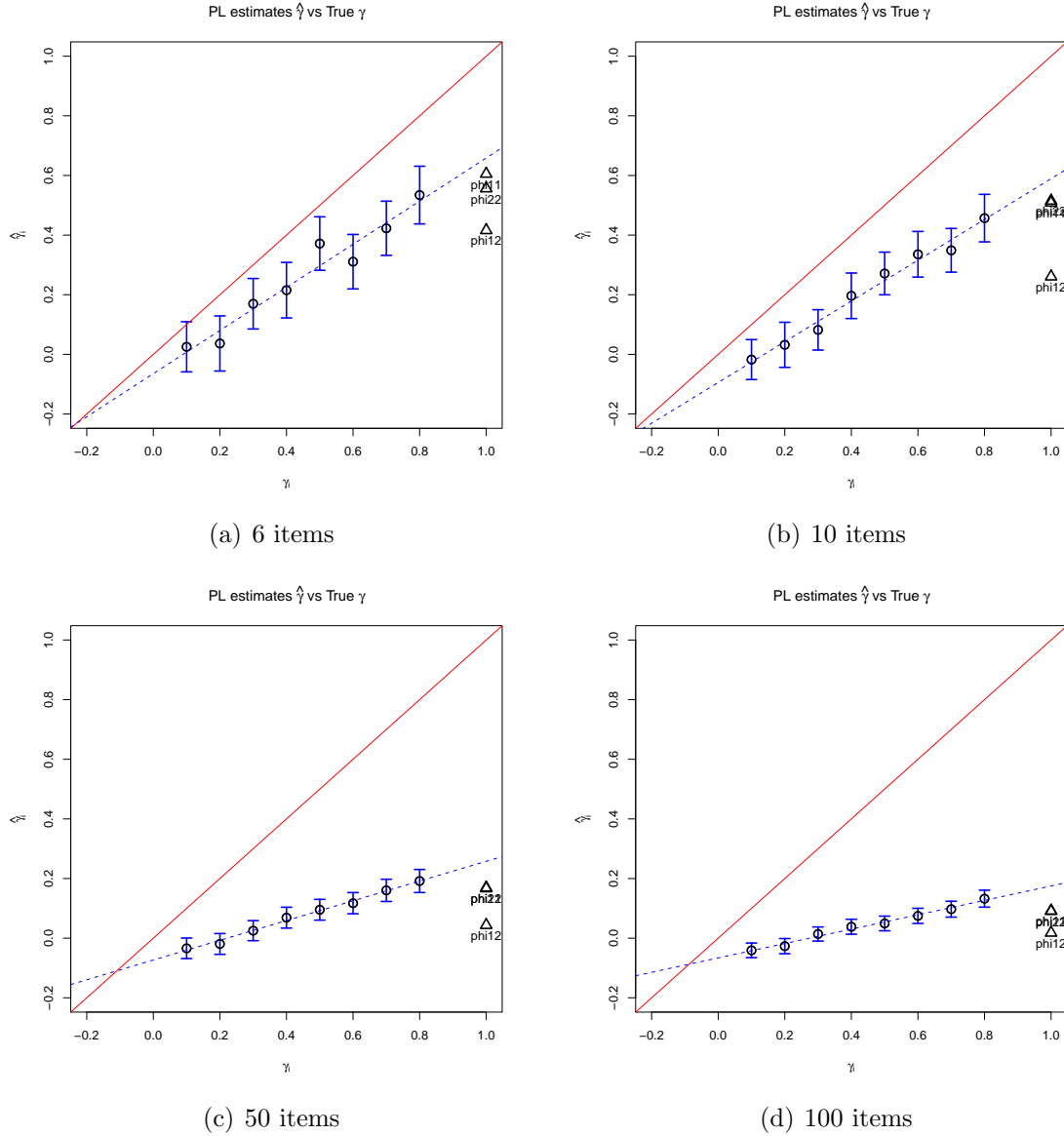


Figure 27. PLE vs true person covariate effects. Test length $I = 6, 10, 50$, and 100 items; 2-D latent traits and dichotomous responses.

In all four plots in Figure 27, the circle points are located on a line (the dashed line), but not on the 45 degree line (the solid line). This suggests that there is a linear relationship

between the γ_0 parameter in the LLLAp model and the γ parameter in the linear regression Rasch model. As the number of test items increases, the slope of the dashed line decreases. Similar to the unidimensional model with person covariates, the slope of the dashed line is closely related to the ratio of the conditional variance parameter (σ_{01}^2 for dimension 1, and σ_{02}^2 for dimension 2 of the latent traits) in the LLLAp model to the population variance parameter (σ_1^2 and σ_2^2) in the latent regression Rasch model. To demonstrate this, on each figure the three points $(\sigma_1^2, \hat{\sigma}_{01}^2)$, $(\sigma_2^2, \hat{\sigma}_{02}^2)$, and $(\sigma_{12}, \hat{\sigma}_{012})$ are also plotted as triangular points (labeled as “phi11”, “phi22”, and “phi12”, respectively). Note that in all four plots, two of the triangular points for the conditional variance parameters are close to the dashed line. This suggest that slope in the linear relationship between γ_0 and γ is strongly related to σ_{01}^2/σ_1^2 and σ_{02}^2/σ_2^2 .

For each PL estimated effect, the 95% confidence interval ($\hat{\gamma} \pm 1.96SE$) is plotted as vertical bars around the estimated value. The length of the 95% CIs decreases as the number of items increases. As the number of items increases, there is more information to help obtain more precise person covariate effects estimates.

Note is that in all four plots in Figure 27, the two points $(\sigma_1^2, \hat{\sigma}_{01}^2)$ and $(\sigma_2^2, \hat{\sigma}_{02}^2)$ are located very close to each other. This can be explained in the following way. First of all, in the simulation setup, $\sigma_1^2 = \sigma_2^2 = 1$. Second, the same number of items load on the two dimensions of the latent traits, and their item difficulties are drawn from the same distribution. Therefore the contribution of the information for estimating the two latent traits are the same, and the uncertainty for the latent traits estimates $\hat{\theta}_1$ and $\hat{\theta}_2$, as measured by σ_{01}^2 and σ_{02}^2 , respectively, are the same. Therefore we see the estimates $\hat{\sigma}_{01}^2$ and $\hat{\sigma}_{02}^2$ very close to each other.

To further demonstrate the relationship among (a) the slope of the estimated $\hat{\gamma}_0$ to true γ , (b) the conditional variance estimates σ_{01}^2 and σ_{02}^2 , and (c) the information for estimating θ_1 and θ_2 , an “unbalanced” test is simulated, in which 3/4 of the items load on the first trait, and 1/4 of the items load on the second trait. In this case we would expect

smaller estimates of $\hat{\sigma}_{01}^2$ than $\hat{\sigma}_{02}^2$. Figure 28 shows the results of the “unbalanced” tests with total number of items 50 and 100. It is obvious that the points for the γ_0 estimates now reside on two lines, one for the first dimension of the latent traits (circular points), and the other for the second dimension (cross points). The slope of the line for the first dimension is lower than the slope for the second dimension, and the $\hat{\sigma}_{01}^2$ (triangular point labeled as “phi11”) is less than $\hat{\sigma}_{02}^2$ (triangular point labeled as “phi22”).

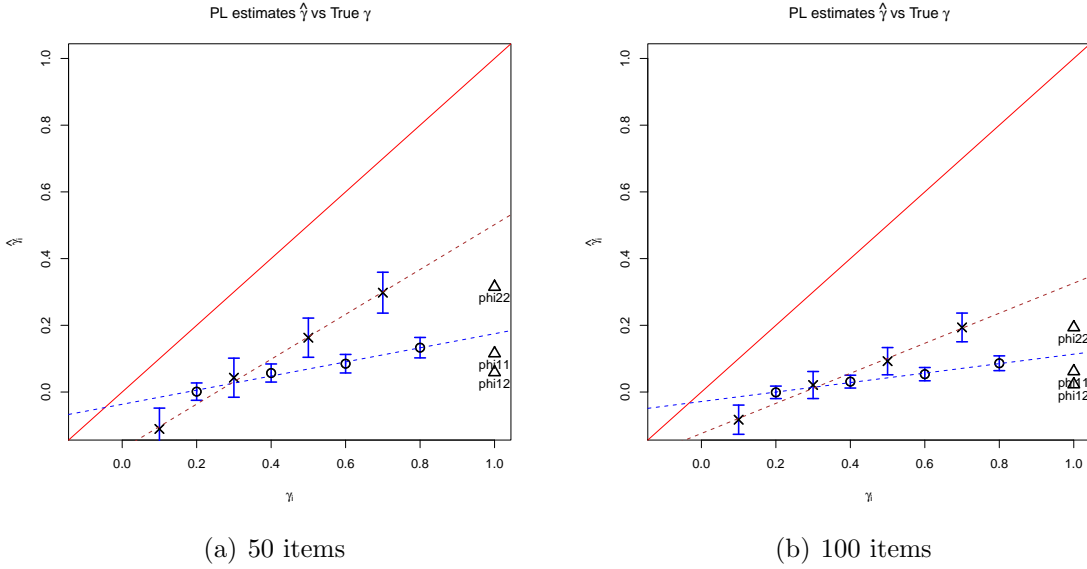


Figure 28. Fitted person covariates vs true parameters with “unbalanced” item loadings. Test length $I = 50$ and 100 items.

Multidimensional Polytomous Models

Multidimensional polytomous LLLA with item covariates.

Simulating data from 2-dimensional polytomous Rasch model with item covariates. Multicategorical response data were simulated from a multidimensional polytomous Rasch model modified to incorporate item covariates. For each item, there were 3 possible response categories, recorded as 0, 1, and 2. Tests of four different lengths, $I = 6$, 10, 50, and 100, were simulated.

For each item, the item parameters β_{i1} and β_{i2} were calculated as the linear combination of item covariates:

$$\begin{aligned}\beta_{i1} &= \beta_{11}^* X_{i1} + \beta_{21}^* X_{i2} + \beta_{31}^* X_{i3} + \beta_{41}^* X_{i4}, \\ \beta_{i2} &= \beta_{12}^* X_{i1} + \beta_{22}^* X_{i2} + \beta_{32}^* X_{i3} + \beta_{42}^* X_{i4}.\end{aligned}$$

The response 0 was set as reference, so the corresponding item parameter is always zero: $\beta_{i0} = 0$. The item covariates X_1, X_2, X_3 and X_4 took values from the standard normal distribution. The item covariate effects β_{jh}^* , $j = 1, 2, 3, 4$, $h = 1, 2$, were assigned values $\{\beta_{j1}^*\} = (\beta_{11}^*, \beta_{21}^*, \beta_{31}^*, \beta_{41}^*) = (0.2, 0.4, 0.6, 0.8)$, and $\{\beta_{j2}^*\} = (\beta_{12}^*, \beta_{22}^*, \beta_{32}^*, \beta_{42}^*) = (0.7, 0.5, 0.3, 0.1)$.

For each test, 1000 persons were simulated. For each person, the latent traits were two-dimensional: $\theta_p = (\theta_{p1}, \theta_{p2})$. The latent traits were sampled from a bivariate normal distribution with mean $(0, 0)'$ and variances 1 and correlation $\rho = 0.5$.

For every combination of persons and items, the probabilities of choosing one of the three responses $(P_{pi0}, P_{pi1}, P_{pi2})$ were calculated from the item category response functions

$$P_{pih} = P(Y_{pi} = h | \theta_p, \beta_{i1}, \beta_{i2}) = \frac{\exp[h(a_{i1}\theta_{p1} + a_{i2}\theta_{p2}) + \beta_{ih}]}{\sum_{l=0}^2 \exp[l(a_{i1}\theta_{p1} + a_{i2}\theta_{p2}) + \beta_{il}]}, \quad h = 0, 1, 2.$$

The response Y_{pi} , which takes on one of the possible values 0, 1 and 2, was generated with the corresponding probabilities $(P_{pi0}, P_{pi1}, P_{pi2})$. The $I \times 2$ matrix $\{a_{id}\}$ is the item-trait adjacency matrix that represents the associations between the items and the latent traits. It was assumed that in the test, the first half of the items only measure the first trait θ_1 and the second half of the items only measure the second trait θ_2 . So $(a_{i1}, a_{i2}) = (1, 0)$ for the first half of the items, and $(a_{i1}, a_{i2}) = (0, 1)$ for the second half of the items.

Comparison of PLE to MLE. Table 15 shows the results of ML estimates and PL estimates for fitting the LLLAi model on the simulated data with test lengths of 6 items and 10 items. The table also reports estimated relative efficiencies of PLE calculated by the

ratio of SE_{ml}/SE_{pl} . The estimated parameters given by MLE and PLE are very close to each other. This closeness is also evident by plotting the PL vs ML estimates of the item covariate effects, as the resulting points are almost exactly on the 45 degree lines (Figure). The robust standard errors given by PLE are close to the standard errors by MLE, as the relative efficiencies are close to 1. Thus the loss of efficiency of PLE relative to the MLE is very low.

Table 15

MLE and PLE Obtained From Fitting the LLLAi (2D) Model on Simulated Data. Test Length $I = 10$ Items. Polytomous items With 3 Categories

	MLE		PLE		Relative Efficiency
	Estimate	SE	Estimate	SE	
λ	-0.265	0.197			
β_{01}^*	-1.780	0.054	-1.783	0.056	0.961
β_{02}^*	-3.246	0.097	-3.252	0.101	0.962
β_{11}^*	0.093	0.043	0.090	0.041	1.031
β_{12}^*	0.614	0.049	0.613	0.048	1.026
β_{21}^*	0.467	0.052	0.469	0.052	0.997
β_{22}^*	0.538	0.060	0.538	0.059	1.017
β_{31}^*	0.562	0.040	0.561	0.041	0.969
β_{32}^*	0.219	0.051	0.217	0.052	0.984
β_{41}^*	0.723	0.035	0.723	0.036	0.993
β_{42}^*	0.012	0.043	0.012	0.042	1.025
σ_{11}	0.280	0.010	0.280	0.010	0.981
σ_{12}	0.074	0.007	0.074	0.007	0.982
σ_{22}	0.323	0.013	0.323	0.013	0.986

Comparing LLLAi estimates to true parameters. In Figure 30, estimated item covariate effects are plotted against the corresponding true values of the parameters used in the simulations with 6, 10, 50 and 100 items. In each plot, the x-axis represents true values and the y-axis represents PL estimates. There are 8 item parameters, and for each item there are two parameter corresponding to responses 1 (circular points) and 2 (triangular points). For item covariate effect parameters, the 95% confidence intervals ($\hat{\beta}^* \pm 1.96SE$) are plotted as vertical bars around the estimated values. In all four plots, the PL estimates

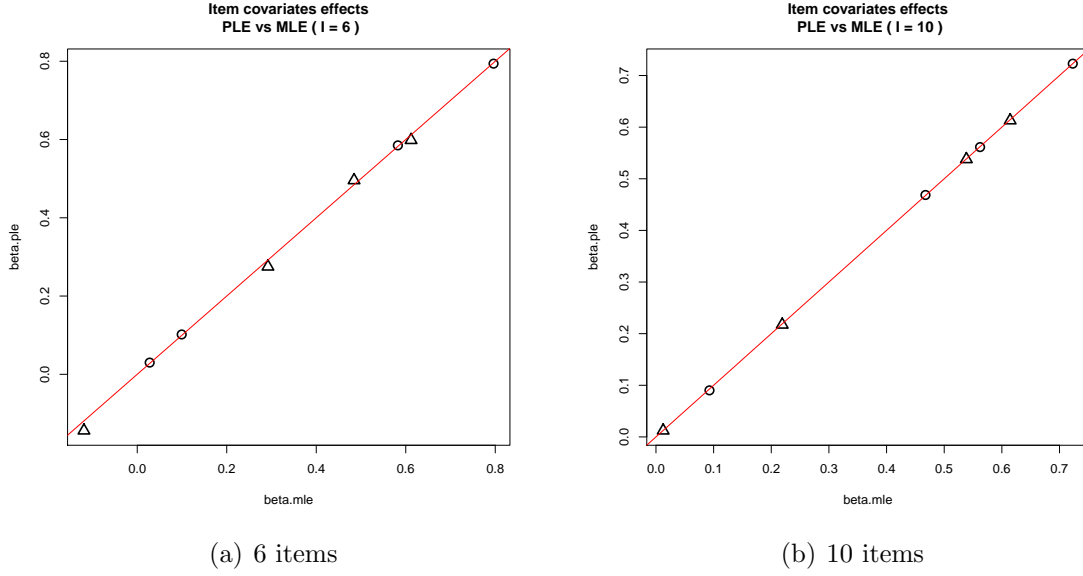


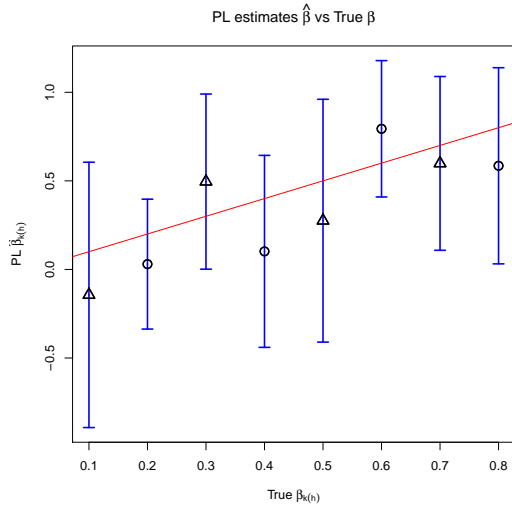
Figure 29. PLE vs MLE for item covariate effects. Test length $I = 6$ and 10 items; 2-dimensional latent traits, 3 category responses.

are close to the 45 degree line, and the 95% CIs cover the true values. This indicates that the true parameters are successfully recovered by the PLE. Note that the lengths of the 95% CIs decrease as the number of items increases. As the number of items increases, there is more information for estimating item covariate effects. Therefore we see smaller SE as the number of items increases.

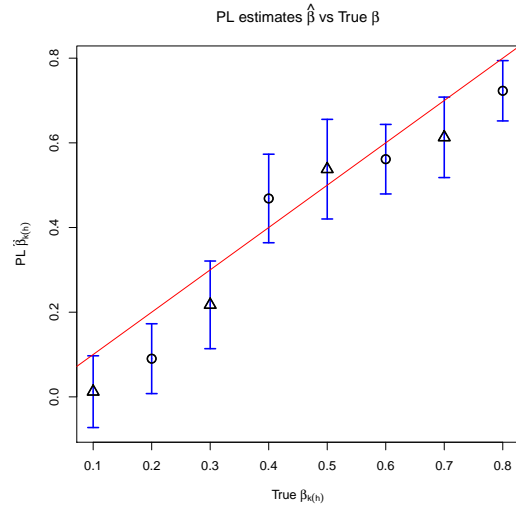
Multidimensional polytomous LLLA with person covariates.

Simulating data from multi-D polytomous Rasch model with person covariates. Multicategorical response data were simulated from a multidimensional polytomous Rasch model, which is a multidimensional version of the partial credit Model. For each item, there were 3 possible response categories, recorded as 0, 1, and 2. Tests of four different lengths, $I = 6, 10, 50$, and 100, were simulated.

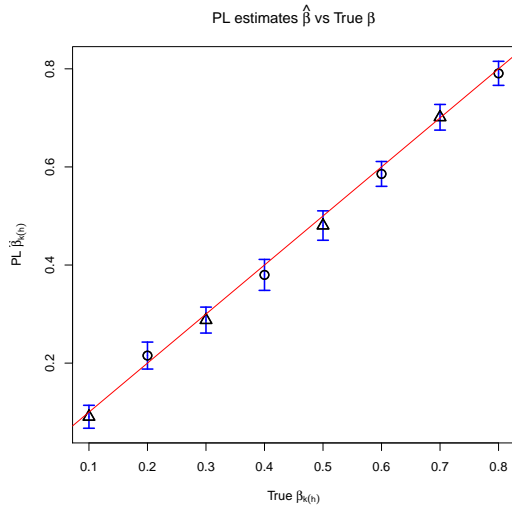
For each item, two item parameters β_{i1} and β_{i2} , which correspond to responses 1 and 2, respectively, were assigned values generated from the standard normal distribution. The response 0 was set as reference, so the corresponding item parameter is always zero: $\beta_{i0} = 0$.



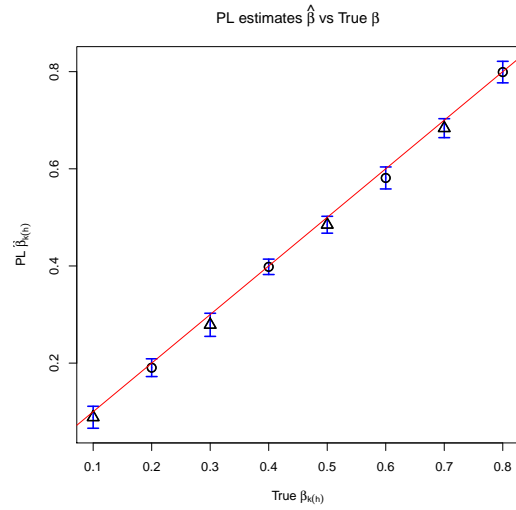
(a) 6 items



(b) 10 items



(c) 50 items



(d) 100 items

Figure 30. PLE vs true item covariate effects. Test length $I = 6, 10, 50$, and 100 items, 2-dimensional latent traits, 3 category response.

For each test, responses for 1000 persons were simulated. For each person, the two-dimensional latent traits $\theta_p = (\theta_{p1}, \theta_{p2})$ were calculated from

$$\theta_p = Z_p \gamma + \epsilon_p.$$

The person covariate vector had 4 variables $Z_p = (Z_{p1}, Z_{p2}, Z_{p3}, Z_{p4})$ and they were all generated from the standard normal distribution. The person covariate effects form a 4×2 matrix $\{\gamma_{jd}\}$ and were assigned the following values:

$$\begin{pmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \\ \gamma_{31} & \gamma_{32} \\ \gamma_{41} & \gamma_{42} \end{pmatrix} = \begin{pmatrix} 0.2 & 0.7 \\ 0.4 & 0.5 \\ 0.6 & 0.3 \\ 0.8 & 0.1 \end{pmatrix}$$

The error term ϵ_p were sampled from a bivariate normal distribution with mean $(0, 0)'$ and variances 1 and correlation $\rho = 0.5$.

For every combination of persons and items, the probabilities of choosing one of the three responses ($P_{pi0}, P_{pi1}, P_{pi2}$) were calculated from the item category response functions

$$P_{pih} = P(Y_{pi} = h | \theta_p, \beta_{i1}, \beta_{i2}) = \frac{\exp[h(a_{i1}\theta_{p1} + a_{i2}\theta_{p2}) + \beta_{ih}]}{\sum_{l=0}^2 \exp[l(a_{i1}\theta_{p1} + a_{i2}\theta_{p2}) + \beta_{il}]}, \quad h = 0, 1, 2.$$

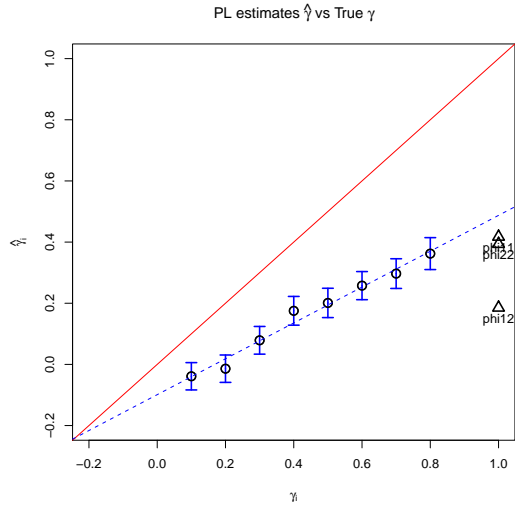
The response Y_{pi} , which took on one of the possible values 0, 1 and 2, was generated with the corresponding probabilities ($P_{pi0}, P_{pi1}, P_{pi2}$). The $I \times 2$ matrix $\{a_{id}\}$ is the item-trait adjacency matrix that represents the associations between the items and the latent traits. It was assumed that in the test, the first half of the items only measure the first trait θ_1 and the second half of the items only measure the second trait θ_2 . So $(a_{i1}, a_{i2}) = (1, 0)$ for the first half of the items, and $(a_{i1}, a_{i2}) = (0, 1)$ for the second half of the items.

Comparing PLE to MLE. It was not possible to run the Poisson regression to fit the LLLAp model on the data set to get MLE due to high computational cost.

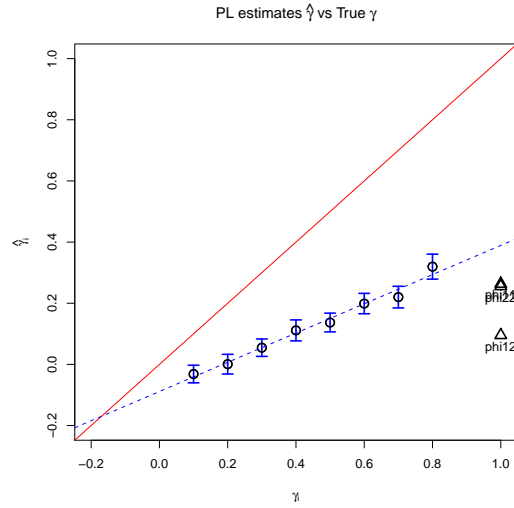
Comparing PLE to true parameters: person covariate effects. In Figure 31, PL estimated person covariate effects in the LLLAp model ($\hat{\gamma}_0$) are plotted against the corresponding true values of the parameters γ in the latent regression Rasch model used in the simulations with 6, 10, 50 and 100 items. In each plot, the x-axis represents true values ($\gamma = \{(0.2, 0.7), (0.4, 0.5), (0.6, 0.3), (0.8, 0.1)\}$ for four person covariates) in the latent regression Rasch model, and the y-axis represents PL estimates in the LLLAp model.

In all four plots in Figure 31, the circle points are located on a line (the dashed line), but not on the 45 degree line (the solid line). This suggests that there is a linear relationship between the γ_0 parameter in the LLLAp model and the γ parameter in the linear regression Rasch model. As the number of test items increases, the slope of the dashed line decreases. Similar to the unidimensional model with person covariates, the slope of the dashed line is closely related to the ratio of the conditional variance parameter (σ_{01}^2 for dimension 1, and σ_{02}^2 for dimension 2 of the latent traits) in the LLLAp model to the population variance parameter (σ_1^2 and σ_2^2) in the latent regression Rasch model. To demonstrate this, on each figure the three points $(\sigma_1^2, \hat{\sigma}_{01}^2)$, $(\sigma_2^2, \hat{\sigma}_{02}^2)$, and $(\sigma_{12}, \hat{\sigma}_{012})$ are also plotted as triangular points (labeled as “phi11”, “phi22”, and “phi12”, respectively). Note that in all four plots, two of the triangular points for the conditional variance parameters are close to the dashed line. This suggests that the slope in the linear relationship between γ_0 and γ is strongly related to σ_{01}^2/σ_1^2 and σ_{02}^2/σ_2^2 .

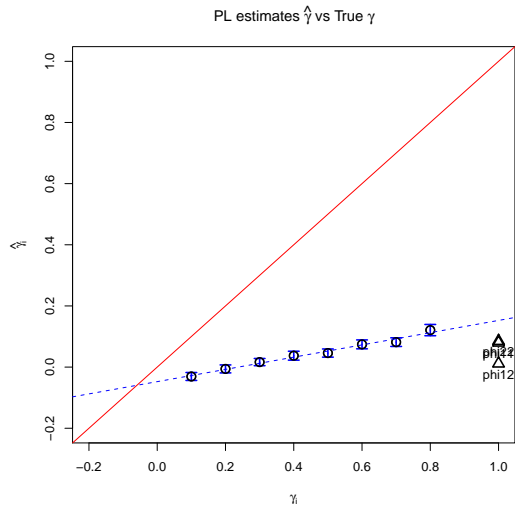
For each PL estimated effect, the 95% confidence interval ($\hat{\gamma} \pm 1.96SE$) is plotted as vertical bars around the estimated value. The length of the 95% CIs decreases as the number of items increases. As the number of items increases, there is more information to help obtain more precise person covariate effects estimates.



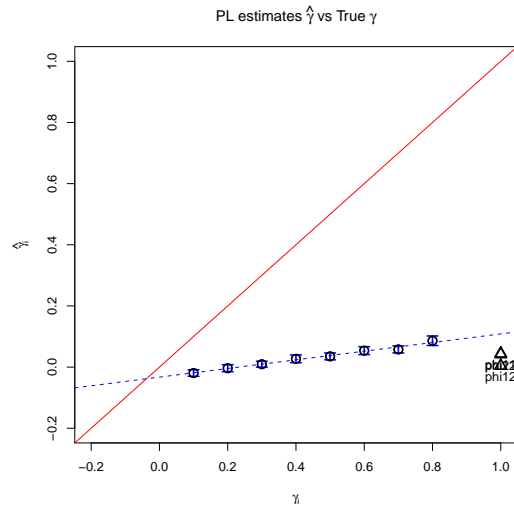
(a) 6 items



(b) 10 items



(c) 50 items



(d) 100 items

Figure 31. PLE vs true person covariate effects. Test length $I = 6, 10, 50$, and 100 items; 2 dimensional traits, 3-category responses.

Chapter 8

Real Data Analysis

In this chapter, I will apply the log-linear-models-as-IRT methods to a verbal aggression study data (De Boeck & Wilson, 2004). Several dichotomous and polytomous IRT models were fitted to the data (Table 16) by using their equivalent LLLA models through pseudo-likelihood estimation method.

Table 16

IRT Models for the Real Data Set

Dataset	Models fitted to data
Aggression	Rasch
24 items, 316 persons,	LLTM
binary or ternary responses	Latent Regression Rasch
	Partial credit model
	Partial credit model with item covariates
	Partial credit model with person covariates

The same data set was used extensively in De Boeck and Wilson (2004) to demonstrate the application of different IRT models under the unified framework of Generalized Linear/Nonlinear Mixed Models (GLMM). Under that framework, IRT models are treated as GLMM. For example, the Rasch model is written as

$$\eta_{pi} = \theta_p - b_i, \quad (8.1)$$

where $\eta_{pi} = \log[P(Y_{pi} = 1|\theta_p)/P(Y_{pi} = 0|\theta_p)]$ is the logit link function, and item parameters b_i are treated as fixed effects, and person parameters θ_p are treated as random effects and it is assumed that $\theta_p \sim N(0, \sigma^2)$. It is essentially the MML formulation of the Rasch model. By formulating IRT models as GLMM, these IRT models can be fitted by statistical packages for GLMM. In De Boeck and Wilson (2004), the SAS procedure PROC NLMIXED was used

to fit many IRT models. For each IRT model, I also include the GLMM results by PROC NLMIXED as a comparison to the LLLA results. The data set was downloaded from the De Boeck and Wilson (2004) book website.³ The SAS code was copied from the book and slightly modified when necessary.

The aggression data set consists of survey responses collected from a study on verbal aggression behavior. In the study, 316 college students answered a questionnaire with 24 items. Each item describes an unpleasant situation, and a behavioral response to the situation. Some of the items are:

- A bus fails to stop for me. I would want to curse.
- I miss a train because a clerk gave me faulty information. I would curse.
- The grocery store closes just as I am about to enter. I would want to scold.
- The operator disconnects me when I had used up my last 10 cents for a call.

I would shout.

The students were asked to respond each item with “yes”, “perhaps” or “no”. Detailed description of the data set and the complete list of the 24 items are given in De Boeck and Wilson (2004), pages 7-10.

There are 3 design factors in the items. The first factor is **situation type**. Four situations are described in the items, with two situations under the type “other-to-blame”, and two situations under the type “self-to-blame”. The “other-to-blame” situations are: “A bus fails to stop for me”, and “I miss a train because a clerk gave me faulty information”. The “self-to-blame” situations are: “the grocery store closes just as I am about to enter”, and “The operator disconnects me when I had used up my last 10 cents for a call”. The second design factor is **behavior type** that includes “curse”, “scold” and “shout”. The third design factor is **behavior mode** that has two levels: want (“I would want to”) vs do (“I would”). A complete factorial design produces the $4 \times 3 \times 2 = 24$ items in the questionnaire.

³<http://bearcenter.berkeley.edu/EIRM/>

These three factors are used as the item covariates in the analysis. The college student's gender (243 females and 73 males) and Trait Anger scores ($M = 20$, $SD=4.85$) derived from a personality inventory are also recorded in the data and they will be used as the person covariates in the analysis.

When the data set is analyzed by dichotomous models, the responses are coded by “yes or perhaps”=1 and “no”=0. When the data set is analyzed by polytomous models, the responses are coded by “yes”=2, “perhaps”=1, and “no”=0. The code for variable Gender is “female”=1 and “male”=0.

Computation Time

Before getting into the detailed results of fitting the IRT models, we will first look at a summary table of the time cost of fitting these models. Table 17 summarizes the computer time by the pseudolikelihood estimation of LLLA models as I implemented in R package ‘plRasch’, and by the GLMM approach with SAS PROC NLMIXED. All the analyses are done on the same personal computer with a 2.16GHz CPU and 1GB memory. We can see that the pseudolikelihood estimation method was very fast and obtained the results in less than 10 seconds, while it took PROC NLMIXED method longer time, ranging from 20 seconds to 6 minutes, to complete the computations. The reason it took PROC NLMIXED longer to fit the models is that it involves numerical integration with Gaussian quadratures to integrate out the random effect parameters (e.g., θ_p in the Rasch model). On the other hand, when fitting the LLLA models with pseudolikelihood estimation, no numerical integration is involved, so the time to get a solution is much lower.

Rasch Model

The first IRT model fitted to the aggression data is the Rasch model. The 316×24 response data matrix is dichotomized by letting responses “yes” or “perhaps”=1, and

Table 17

Time Cost of Fitting IRT Models on Verbal Aggression Data

Model	LLLA (plRasch) Time	NLMIXED Time
Rasch	<1 sec	1 min 48 sec
LLTM	<1 sec	20 sec
Latent regression Rasch	<1 sec	1 min 50 sec
Partial credit model	8 sec	5 min 43 sec
PCM with item covariates	2 sec	43 sec
PCM with person covariates	8 sec	6 min 19 sec

“no”=0. The item response function (IRF) of the Rasch model is

$$P(Y_{pi} = 1|\theta_p) = \frac{\exp(\theta_p - b_i)}{1 + \exp(\theta_p - b_i)}. \quad (8.2)$$

In the LLLA model, it is assumed $\theta_p|(y_1, \dots, y_I) \sim N(\mu_y, \sigma_0^2)$. Fitting the LLLA model produced the estimates of item effects λ_i , which were then transformed into the estimates of the Rasch difficulty parameters b_i ; and the estimate for the conditional variance σ_0^2 . On the other hand, in GLMM, it is assumed $\theta_p \sim N(0, \sigma^2)$. GLMM approach (PROC NLMIXED) gives the estimates for the item difficulties b_i and the variance σ^2 .

Table 18 presents the parameter estimates and the standard errors by the LLLA model (item parameters transformed into Rasch parameters) and GLMM. The two methods give similar estimates and SEs for the item difficulty parameters. The closeness of the parameter estimates from the two methods is also evident in the scatter plot of the two sets of item parameter estimates (Figure 32), where all the points are located very close to the 45 degree line.

The conditional variance parameter estimated by the LLLA model is $\hat{\sigma}_0^2 = 0.218$ ($SE = 0.006$). We calculate the population latent trait variance estimate under the LLLA model by $\hat{\sigma}_M^2 = \hat{\sigma}_0^2 + \hat{\sigma}_0^4 \text{var}(T)$ and the result is $\hat{\sigma}_M^2 = 1.748$. The estimate of the population latent trait variance parameter given by GLMM (PROC NLMIXED) is $\hat{\sigma}^2 = 1.976$ ($SE = 0.211$).

Table 18

Parameter Estimates and Standard Errors for the Rasch Model (Verbal Aggression Data)

	LLLA		GLMM	
	Estimate	SE	Estimate	SE
b_1	-1.22	0.14	-1.23	0.16
b_2	-0.57	0.13	-0.57	0.15
b_3	-0.08	0.15	-0.09	0.15
b_4	-1.74	0.15	-1.75	0.17
b_5	-0.71	0.13	-0.71	0.15
b_6	-0.01	0.15	-0.02	0.15
b_7	-0.54	0.14	-0.53	0.15
b_8	0.69	0.15	0.68	0.15
b_9	1.53	0.19	1.52	0.17
b_{10}	-1.09	0.14	-1.09	0.16
b_{11}	0.35	0.14	0.34	0.15
b_{12}	1.05	0.17	1.04	0.16
b_{13}	-1.23	0.14	-1.23	0.16
b_{14}	-0.39	0.13	-0.40	0.15
b_{15}	0.87	0.16	0.87	0.16
b_{16}	-0.88	0.14	-0.88	0.15
b_{17}	0.05	0.14	0.05	0.15
b_{18}	1.48	0.17	1.48	0.17
b_{19}	0.21	0.15	0.21	0.15
b_{20}	1.50	0.18	1.50	0.17
b_{21}	2.93	0.23	2.98	0.23
b_{22}	-0.71	0.14	-0.71	0.15
b_{23}	0.38	0.15	0.38	0.15
b_{24}	1.99	0.20	2.00	0.18
σ_0^2	0.218	0.006	—	—
σ^2	—	—	1.976	0.211

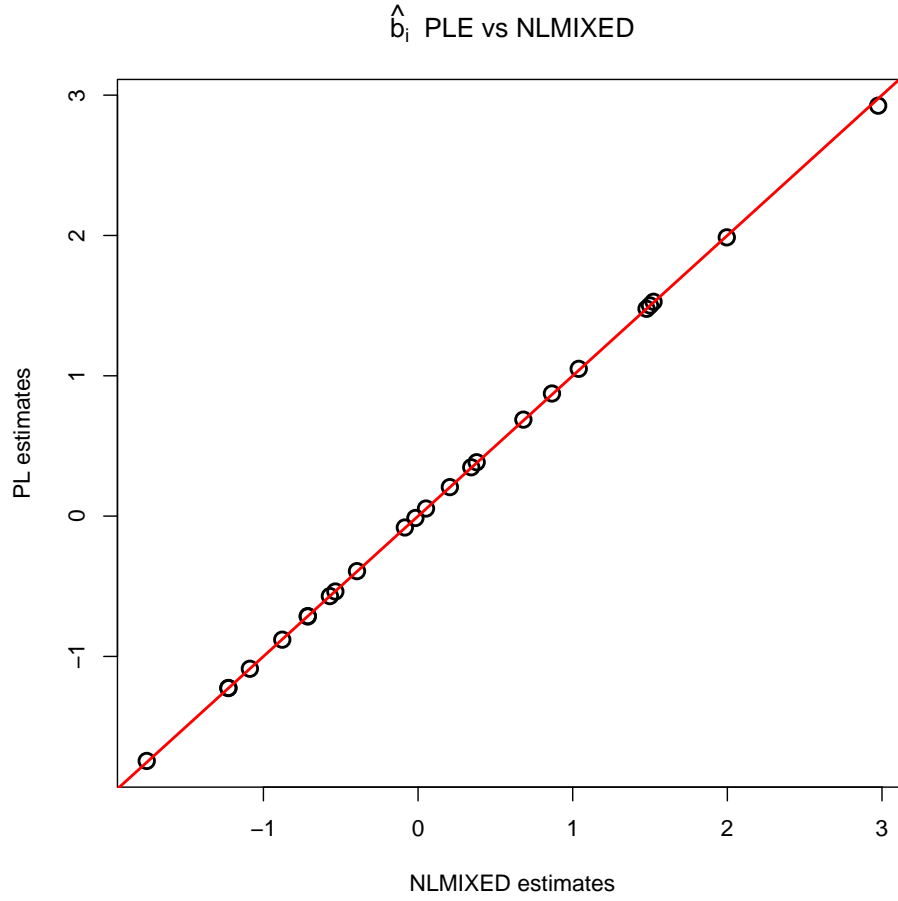


Figure 32. Item difficulty estimates for the Rasch model, LLLA (PL) vs GLMM (NLMIXED).

LLTM

One interesting question for the aggression data is how the three item design factors (situation type, behavior type, and behavior mode) relate to people's reactions with verbally aggressive behavior. To study the question, a linear logistic test model (LLTM) is used, where the item design factors are coded as 4 predictors. Situation type is coded by predictor X_{other} where other-to-blame=1 and self-to-blame=0. Behavior type is dummy coded by a pair of predictors (X_{scold}, X_{shout}) , where curse=(0,0), scold=(1,0), and shout=(0,1). Behavior mode is coded by predictor X_{want} where want=1 and do=0. Together with the intercept, these 4

predictors predict the item difficulty parameters with the model

$$b_i = \beta_0 + \beta_1 X_{other,i} + \beta_2 X_{scold,i} + \beta_3 X_{shout,i} + \beta_4 X_{want,i}. \quad (8.3)$$

So the item response function (IRF) for the LLTM is

$$P(Y_{pi} = 1|\theta_p) = \frac{\exp[\theta_p - (\beta_0 + \beta_1 X_{other,i} + \beta_2 X_{scold,i} + \beta_3 X_{shout,i} + \beta_4 X_{want,i})]}{1 + \exp[\theta_p - (\beta_0 + \beta_1 X_{other,i} + \beta_2 X_{scold,i} + \beta_3 X_{shout,i} + \beta_4 X_{want,i})]}. \quad (8.4)$$

In the LLLA model, it is assumed $\theta_p|(y_1, \dots, y_I) \sim N(\mu_y, \sigma_0^2)$. Fitting the LLLA model produces the estimates for item covariate effects β_0, \dots, β_4 , and the conditional variance σ_0^2 . On the other hand, in GLMM, it is assumed $\theta_p \sim N(0, \sigma^2)$. GLMM approach (PROC NLIMIXED) gives the estimates for the item covariate effects β_0, \dots, β_4 , and the variance σ^2 .

Table 19 presents the parameter estimates and the standard errors by the LLLA model and GLMM. The two methods give similar estimates and SEs for the item difficulty parameters for the four predictors. The closeness of the parameter estimates for the effects of the four predictors from the two methods is also evident in the scatter plot of the two sets of parameters (Figure 33), where the points for the 4 predictors are located very close to the 45 degree line. The estimates for the intercepts differ in the two models because in the LLLA model the intercept parameter is actually equal to the intercept parameter in the LLTM plus some constant (See Equation (4.7) on page 53). The conditional variance parameter estimated by the LLLA model is $\hat{\sigma}_0^2 = 0.212$ ($SE = 0.006$). We calculate the population latent trait variance estimate under the LLLA model by $\hat{\sigma}_M^2 = \hat{\sigma}_0^2 + \hat{\sigma}_0^4 \text{var}(T)$ and the result is $\hat{\sigma}_M^2 = 1.662$. The estimate of the population latent trait variance parameter given by GLMM (PROC NLIMIXED) is $\hat{\sigma}^2 = 1.841$ ($SE = 0.193$).

The estimated effect of the situation type is -1.03 ($SE = 0.08$), and it is highly significant. When others are to blame, compared to when oneself is to blame, while keeping

Table 19

Parameter Estimates and Standard Errors for the LLTM (Verbal Aggression Data)

	LLLA		GLMM	
	Estimate	SE	Estimate	SE
Intercept	2.26	0.10	-0.05	0.09
other-to-blame vs self-to-blame	-1.03	0.08	-1.03	0.06
scold vs curse	1.06	0.09	1.06	0.07
shout vs curse	2.04	0.12	2.04	0.07
want vs do	-0.67	0.08	-0.67	0.06
σ_0^2	0.212	0.006	—	—
σ^2	—	—	1.859	0.198

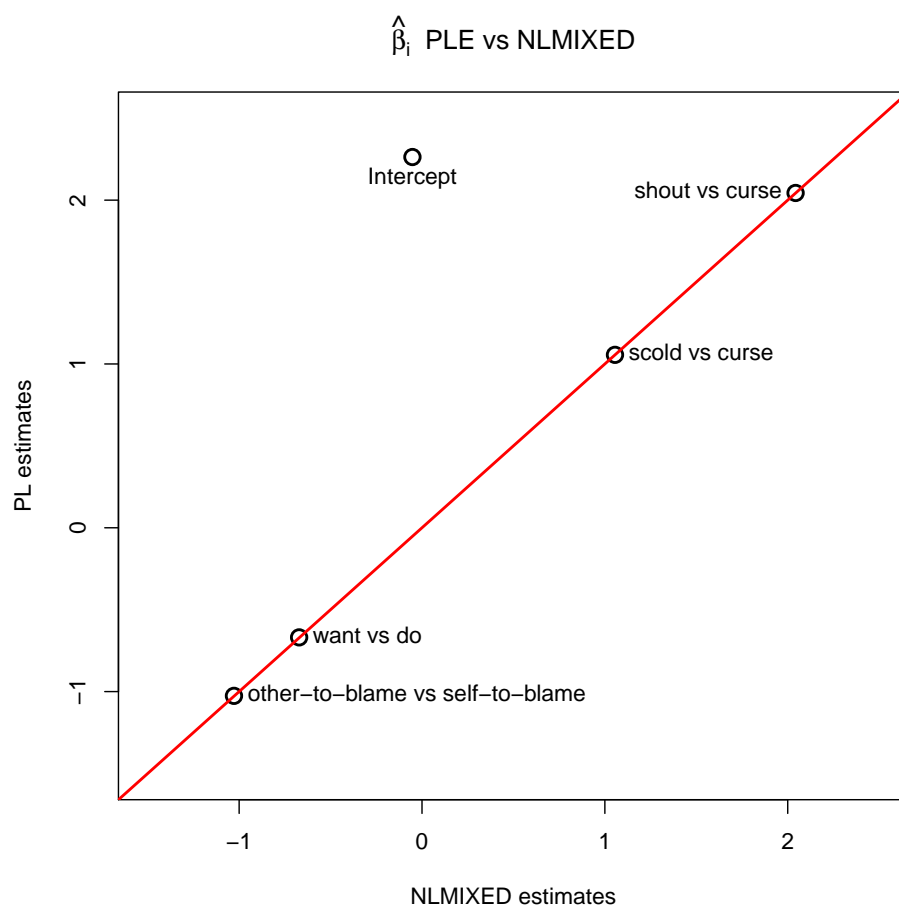


Figure 33. Item covariate effects estimates for the LLTM, LLLA (PL) vs GLMM (NLMIXED).

other factors the same (including the other item covariates behavior type and behavior mode, and the person parameter), the person would more likely to answer “yes” or “perhaps” rather than “no” to the item because the threshold of the item decreases by 1.03, so it will be easier to answer “yes” to the item; in other words, verbal aggression is more common in an other-to-blame situation than a self-to-blame situation. Quantitatively, in an other-to-blame situation as compared to a self-to-blame situation, the odds of answering “yes” or “perhaps” rather than “no” to the item increase by a factor $\exp(1.03) = 2.80$. For example, the effect on a probability of 0.50 (odds = 1) in a self-to-blame situation would increase the probability to 0.74 (odds = 2.80) in an other-to-blame situation.

Two predictors are used for the effect of the behavior type. The estimated effect of the first predictor (Scold vs Curse) is 1.06 ($SE = 0.09$); and the estimated effect of the second predictor (Shout vs Curse) is 2.04 ($SE = 0.12$). Both effects are highly significant. Therefore in terms of the “difficulty” (or the threshold) of the three behavior types, Curse < Scold < Shout; in other words, under an unpleasant situation, it is easier to curse than to scold, and easier to scold than to shout. The odds of cursing are $\exp(1.06) = 2.89$ times higher than those of scolding, and $\exp(2.04) = 7.69$ times higher than those of shouting. For example, in a situation where a specific person has probability 0.5 (odds = 1) to curse, this person would have probability 0.26 (odds = $1/2.89$) to scold, and probability 0.12 (odds = $1/7.69$) to shout.

The estimated effect of the behavior mode is -0.67 ($SE = .08$), and this effect is highly significant. To answer “yes” or “perhaps” rather than “no” to an item with “want” is less difficult than the corresponding item with “do”. For example, in an unpleasant situation, it is easier to “want to curse” than to “do curse”. The odds of doing are $\exp(-0.67) = 0.51$ times lower than the odds of wanting. For example, If the probability of “wanting to curse” were 0.50 (odds = 1), then the probability of “to curse” would be 0.34 (odds = 0.51).

Latent Regression Rasch Model

There are two person properties: Gender and the Trait Anger scores. Denote the variable for Gender as Z_{gender} (male=1 and female=0), and the centered Trait Anger Scores as Z_{anger} (so $M = 0$). The latent regression models the latent trait as

$$\theta_p = \gamma_1 Z_{gender,p} + \gamma_2 Z_{anger,p} + \epsilon. \quad (8.5)$$

So the response function (IRF) for the latent regression Rasch model is

$$P(Y_{pi} = 1|\theta_p) = \frac{\exp(\epsilon_p + \gamma_1 Z_{gender,p} + \gamma_2 Z_{anger,p} - b_i)}{1 + \exp(\epsilon_p + \gamma_1 Z_{gender,p} + \gamma_2 Z_{anger,p} - b_i)}. \quad (8.6)$$

In the LLLA model, it is assumed $\epsilon_p|(y_1, \dots, y_I) \sim N(\mu_y, \sigma_0^2)$. Fitting the LLLA model produces the estimates of item effects, which are transformed into the estimates of the Rasch difficulty parameters b_i ; and the estimate for the conditional variance σ_0^2 . On the other hand, in GLMM, it is assumed $\theta_p \sim N(0, \sigma^2)$. GLMM approach (PROC NLIMIXED) gives the estimates for the item difficulties b_i and the variance σ^2 .

Table 20 presents the parameter estimates and the standard errors for the person covariate effects and the variance by the LLLA model and GLMM. The two methods give similar estimates and SEs for the item difficulty parameters (not reported in the table). The person covariate effects estimates in the two models are different, because they have different interpretation; in the LLLA model, the effects of the person covariates are conditional, applicable to the subpopulation with a specific response pattern; in GLMM, the person covariate effects are marginal, applicable to the whole population. There is a linear relationship between the two sets of parameters (Figure 34). In Figure 34, the solid line is the 45 degree line. The conditional variance parameter estimated by the LLLA model is $\hat{\sigma}_0^2 = 0.216$ ($SE = 0.006$). We calculate the population latent trait variance estimate after adjusting the person covariates under the LLLA model by $\hat{\sigma}_M^2 = \hat{\sigma}_0^2 + \hat{\sigma}_0^4 \text{var}(T)$ and the

result is $\hat{\sigma}_M^2 = 1.721$. The estimate of the population latent trait variance parameter given by GLMM (PROC NLMIXED) is $\hat{\sigma}^2 = 1.841$ ($SE = 0.193$).

Table 20

Parameter Estimates and Standard Errors for the Latent Regression Rasch Model (Verbal Aggression Data)

LLLA			GLMM		
	Estimate	SE		Estimate	SE
Anger	0.00847	0.00291	Anger	0.0565	0.0158
Gender	0.0411	0.025	Gender	0.292	0.202
σ_0^2	0.216	0.006	σ^2	1.841	0.193

Partial Credit Model

The first IRT model fitted to the polytomous aggression data is the partial credit model (PCM). The entries in the 316×24 response data matrix are coded “yes”=2, “perhaps”=1, and “no”=0. The item category response function (ICRF) of the PCM is

$$P(Y_{pi} = h|\theta_p) = \frac{\exp(h\theta_p + b_{ih})}{1 + \exp(\theta_p + b_{i1}) + \exp(2\theta_p + b_{i2})}, \quad (8.7)$$

where b_{ih} , $i = 1, \dots, 24$, $h = 0, 1, 2$, is the item difficulty parameter for the i -th item and associated with the h -th response, and $b_{i0} = 0$. For each item, there are two item difficulty parameters b_{i1} and b_{i2} .

In the LLLA model, it is assumed $\theta_p|(y_1, \dots, y_I) \sim N(\mu_y, \sigma_0^2)$. Fitting the LLLA model produces the estimates of item effects, which are transformed into the estimates of the PCM difficulty parameters b_{i1} and b_{i2} ; and the estimate for the conditional variance σ_0^2 . On the other hand, in GLMM, it is assumed $\theta_p \sim N(0, \sigma^2)$. GLMM approach (PROC NLMIXED) gives the estimates for the item difficulties b_{i1} and b_{i2} and the variance σ^2 .

Table 21 presents the parameter estimates and the standard errors by the LLLA model (item parameters transformed into PCM parameters) and GLMM. The two methods

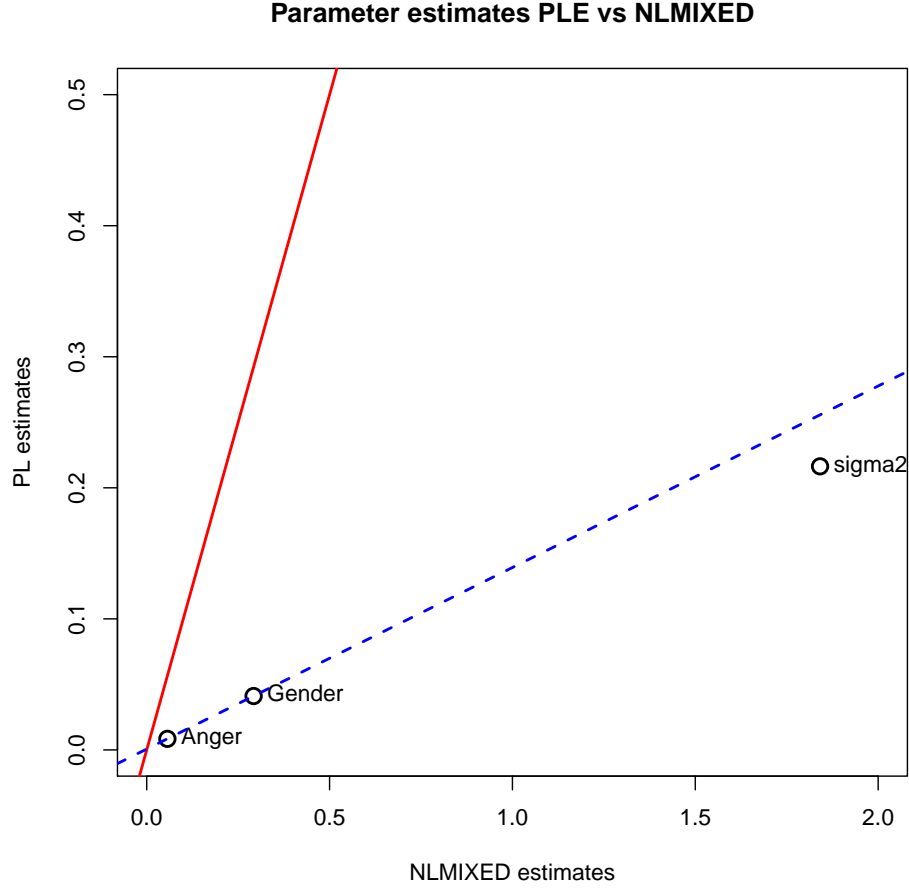


Figure 34. Person covariate effects estimates for the latent regression Rasch model, LLLA (PL) vs GLMM (NLMIXED).

give similar estimates and SEs for the item difficulty parameters. The closeness of the parameter estimates from the two methods is also evident in the scatter plot of the 48 item parameter estimates (Figure 35). All the points are located very close to the 45 degree line.

The conditional variance parameter estimated by the LLLA model is $\hat{\sigma}_0^2 = 0.0915$ ($SE = 0.003$). We calculate the population latent trait variance estimate under the LLLA model by $\hat{\sigma}_M^2 = \hat{\sigma}_0^2 + \hat{\sigma}_0^4 \text{var}(T)$ and the result is $\hat{\sigma}_M^2 = 0.806$. The estimate of the population latent trait variance parameter given by GLMM (PROC NLMIXED) $\hat{\sigma}^2 = 0.912$ ($SE = 0.090$).

Table 21

Parameter Estimates and Standard Errors for the Partial Credit Model (Verbal Aggression Data)

	LLLA		GLMM		(Contd.)	LLLA		GLMM	
	Est.	SE	Est.	SE		Est.	SE	Est.	SE
$b_{1(1)}$	0.42	0.14	0.41	0.17	$b_{1(2)}$	0.55	0.16	0.49	0.20
$b_{2(1)}$	-0.11	0.14	-0.14	0.16	$b_{2(2)}$	-0.22	0.17	-0.30	0.20
$b_{3(1)}$	-0.29	0.13	-0.32	0.15	$b_{3(2)}$	-1.20	0.21	-1.27	0.21
$b_{4(1)}$	0.94	0.15	0.96	0.17	$b_{4(2)}$	1.02	0.17	0.98	0.21
$b_{5(1)}$	0.05	0.14	0.03	0.16	$b_{5(2)}$	-0.11	0.17	-0.19	0.20
$b_{6(1)}$	-0.47	0.14	-0.50	0.15	$b_{6(2)}$	-1.03	0.21	-1.10	0.20
$b_{7(1)}$	0.15	0.13	0.12	0.15	$b_{7(2)}$	-0.83	0.21	-0.89	0.21
$b_{8(1)}$	-0.79	0.14	-0.83	0.15	$b_{8(2)}$	-2.65	0.27	-2.70	0.26
$b_{9(1)}$	-1.47	0.17	-1.50	0.16	$b_{9(2)}$	-3.95	0.34	-3.99	0.34
$b_{10(1)}$	0.56	0.14	0.55	0.15	$b_{10(2)}$	-0.06	0.19	-0.12	0.20
$b_{11(1)}$	-0.63	0.13	-0.67	0.15	$b_{11(2)}$	-1.77	0.22	-1.84	0.22
$b_{12(1)}$	-1.25	0.16	-1.28	0.16	$b_{12(2)}$	-2.56	0.27	-2.60	0.24
$b_{13(1)}$	0.53	0.14	0.52	0.16	$b_{13(2)}$	0.40	0.18	0.34	0.20
$b_{14(1)}$	-0.12	0.13	-0.15	0.15	$b_{14(2)}$	-0.64	0.19	-0.72	0.20
$b_{15(1)}$	-1.11	0.15	-1.15	0.16	$b_{15(2)}$	-2.30	0.25	-2.36	0.23
$b_{16(1)}$	0.20	0.14	0.18	0.16	$b_{16(2)}$	0.07	0.19	-0.01	0.20
$b_{17(1)}$	-0.43	0.13	-0.47	0.15	$b_{17(2)}$	-1.30	0.21	-1.38	0.21
$b_{18(1)}$	-1.60	0.16	-1.63	0.17	$b_{18(2)}$	-3.15	0.30	-3.19	0.27
$b_{19(1)}$	-0.38	0.13	-0.42	0.14	$b_{19(2)}$	-2.05	0.26	-2.11	0.24
$b_{20(1)}$	-1.49	0.16	-1.52	0.16	$b_{20(2)}$	-3.69	0.33	-3.73	0.31
$b_{21(1)}$	-2.74	0.23	-2.76	0.22	$b_{21(2)}$	-5.76	0.50	-5.84	0.57
$b_{22(1)}$	0.24	0.13	0.22	0.15	$b_{22(2)}$	-0.47	0.21	-0.54	0.20
$b_{23(1)}$	-0.62	0.14	-0.66	0.15	$b_{23(2)}$	-1.93	0.24	-1.99	0.23
$b_{24(1)}$	-1.98	0.18	-2.00	0.18	$b_{24(2)}$	-4.05	0.35	-4.07	0.33
σ_0^2	0.0915	0.0030	—	—	σ^2	—	—	0.912	0.090

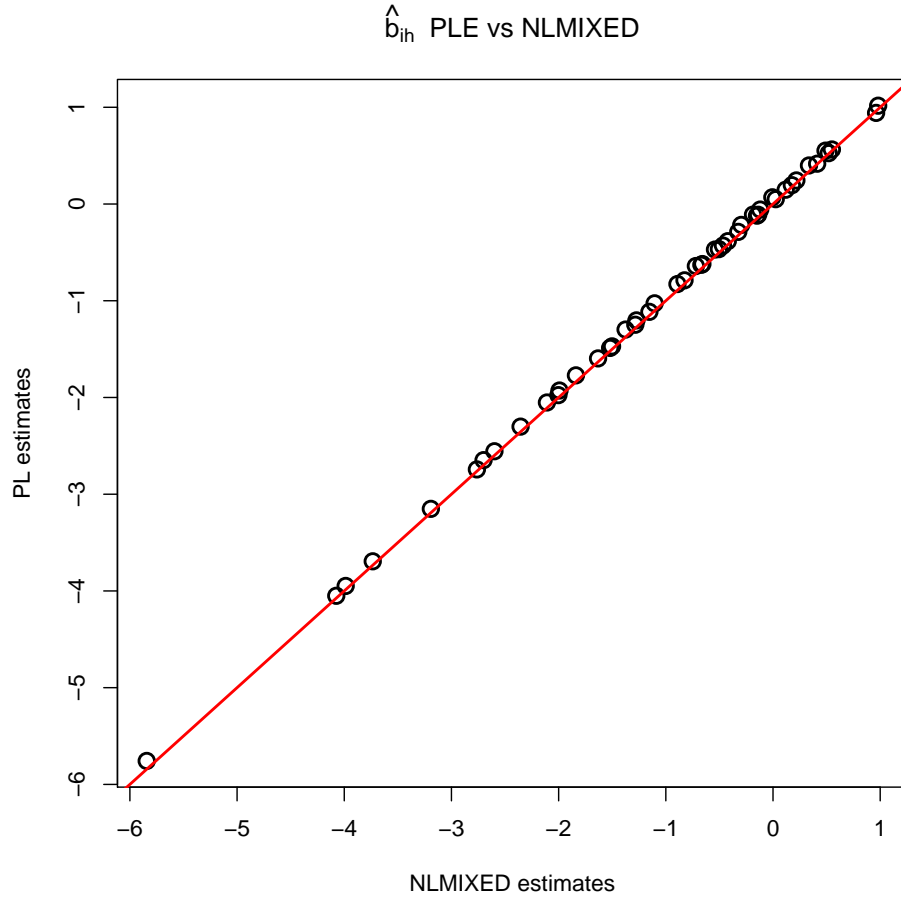


Figure 35. Item difficulty estimates for the partial credit model, LLA (PL) vs GLMM (NLMIXED).

Partial Credit Model With Item Covariates

The partial credit model with item covariates (PCMi) is the polytomous version of the LLTM. The item properties (situation type, behavior type, and behavior mode) are modeled in the PCMi as 4 predictors. Together with the intercept, these 4 predictors are to predict the item difficulty parameters with the model

$$b_{i1} = \beta_{01} + \beta_{11}X_{other,i} + \beta_{21}X_{scold,i} + \beta_{31}X_{shout,i} + \beta_{41}X_{want,i} , \quad (8.8)$$

$$b_{i2} = \beta_{02} + \beta_{12}X_{other,i} + \beta_{22}X_{scold,i} + \beta_{32}X_{shout,i} + \beta_{42}X_{want,i} . \quad (8.9)$$

So the item category response function (ICRF) for the PCMi is

$$P(Y_{pi} = h|\theta_p) = \frac{\exp[h\theta_p + (\beta_{0h} + \beta_{1h}X_{other,i} + \beta_{2h}X_{scold,i} + \beta_{3h}X_{shout,i} + \beta_{4h}X_{want,i})]}{1 + \exp[\theta_p + (\beta_{01} + \dots + \beta_{41}X_{want,i})] + \exp[2\theta_p + (\beta_{02} + \dots + \beta_{42}X_{want,i})]}. \quad (8.10)$$

In the LLLA model, it is assumed $\theta_p|(y_1, \dots, y_I) \sim N(\mu_y, \sigma_0^2)$. Fitting the LLLA model produces the estimates for item covariate effects $\beta_{01}, \dots, \beta_{41}, \beta_{02}, \dots, \beta_{42}$, and the conditional variance σ_0^2 . On the other hand, in GLMM, it is assumed $\theta_p \sim N(0, \sigma^2)$. GLMM approach (PROC NLIMIXED) gives the estimates for item covariate effects $\beta_{01}, \dots, \beta_{41}, \beta_{02}, \dots, \beta_{42}$, and the variance σ^2 .

Table 22 presents the parameter estimates and the standard errors by the LLLA model and GLMM. The two methods give similar estimates and SEs for the item difficulty parameters for the four predictors. The closeness of the parameter estimates for the effects of the four predictors in the two methods is also evident by scatter plot of the two sets of parameters (Figure 36). All the points for the 8 item covariate effects associated with the 4 predictors are located very close to the 45 degree line. The estimates for the intercept differ in the two models because in the LLLA model the intercept parameter is actually equal to the intercept parameter in the PCMi plus some constant (See Equation (5.49) on page 73). The conditional variance parameter estimated by the LLLA model is $\hat{\sigma}_0^2 = 0.088$ ($SE = 0.003$). We calculate the population latent trait variance estimate under the LLLA model by $\hat{\sigma}_M^2 = \hat{\sigma}_0^2 + \hat{\sigma}_0^4 \text{var}(T)$ and the result is $\hat{\sigma}_M^2 = 0.752$. The estimate of the population latent trait variance parameter given by GLMM (PROC NLIMIXED) is $\hat{\sigma}^2 = 0.864$ ($SE = 0.085$).

Partial Credit Model With Person Covariates

The partial credit model with person covariates (PCMp) is the polytomous version of the latent regression Rasch model. There are two person properties: Gender and the Trait Anger scores. Denote the variable for Gender as Z_{gender} (male=1 and female=0), and the centered Trait Anger Scores as Z_{anger} (so $M = 0$). In the PCMp the latent trait is modeled

Table 22

Parameter Estimates and Standard Errors for the Partial Credit Model With Item Covariates (Verbal Aggression Data)

	LLLA		GLMM	
	Estimate	SE	Estimate	SE
Intercept (1)	-1.66	0.08	-0.29	0.08
other-to-blame vs self-to-blame (1)	0.68	0.07	0.69	0.06
scold vs curse (1)	-0.86	0.08	-0.88	0.07
shout vs curse (1)	-1.65	0.10	-1.68	0.08
want vs do (1)	0.53	0.08	0.54	0.06
Intercept (2)	-4.13	0.18	-1.49	0.14
other-to-blame vs self-to-blame (2)	1.67	0.11	1.66	0.08
scold vs curse (2)	-1.30	0.11	-1.31	0.09
shout vs curse (2)	-2.61	0.16	-2.60	0.10
want vs do (2)	0.89	0.11	0.89	0.08
σ_0^2	0.088	0.003	—	—
σ^2	—	—	0.864	0.085

as

$$\theta_p = \gamma_1 Z_{gender,p} + \gamma_2 Z_{anger,p} + \epsilon. \quad (8.11)$$

So the response function (IRF) for the latent regression Rasch model is

$$P(Y_{pi} = h | \theta_p, b_{i1}, b_{i2}) = \frac{\exp(h\epsilon_p + h\gamma_1 Z_{gender,p} + h\gamma_2 Z_{anger,p} + b_{ih})}{1 + \exp(\epsilon_p + \gamma_1 Z_{gender,p} + \gamma_2 Z_{anger,p} + b_{i1}) + \exp(2\epsilon_p + 2\gamma_1 Z_{gender,p} + 2\gamma_2 Z_{anger,p} + b_{i2})}. \quad (8.12)$$

In the LLLA model, it is assumed $\epsilon_p | (y_1, \dots, y_I) \sim N(\mu_y, \sigma_0^2)$. Fitting the LLLA model produces the estimates of item effects, which are transformed into the estimates of the PCM difficulty parameters b_{i1} and b_{i2} , and the estimate for the conditional variance σ_0^2 . On the other hand, in GLMM, it is assumed $\theta_p \sim N(0, \sigma^2)$. GLMM approach (PROC NLIMIXED) gives the estimates for the item difficulties b_{i1} and b_{i2} and the variance σ^2 .

Table 23 presents the parameter estimates and the standard errors for the person

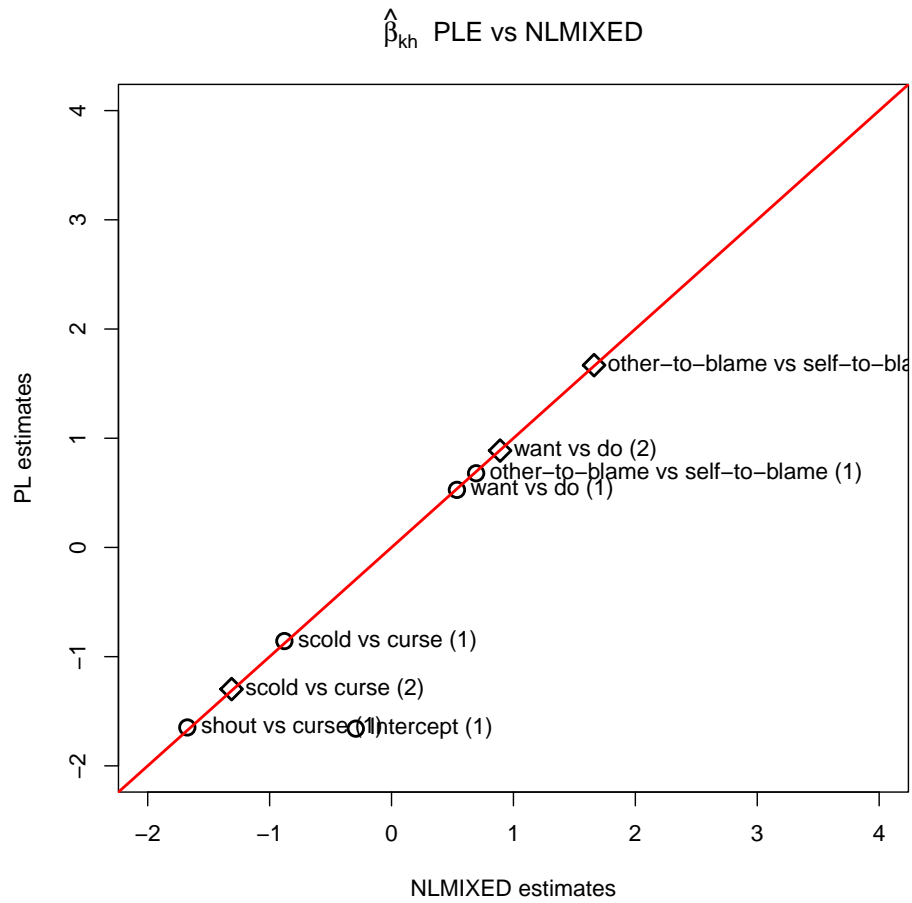


Figure 36. Item covariate effects estimates for the partial credit model with item covariates, LLLA (PL) vs GLMM (NLMIXED).

Table 23

Parameter Estimates and Standard Errors for Partial Credit Model With Person Covariates (Verbal Aggression Data)

	LLLA			GLMM	
	Estimate	SE		Estimate	SE
Anger	0.00742	0.00188	Anger	0.0566	0.0102
Gender	0.0416	0.0162	Gender	0.28	0.122
σ_0^2	0.0905	0.00294	σ^2	0.851	0.0862

covariate effects and the variance by the LLLA model and GLMM. The two methods give similar estimates and SEs for the item difficulty parameters (not reported in the table). The person covariate effects estimates in the two models are different, because they have different interpretation; in the LLLA model, the effects of the person covariates are conditional, applied to the subpopulation with specific response pattern; in GLMM, the person covariate effects are marginal, applied to the whole population. There is a linear relationship between the two sets of parameters (Figure 37). In Figure 37, the solid line is the 45 degree line. The conditional variance parameter estimated by the LLLA model is $\hat{\sigma}_0^2 = 0.091$ ($SE = 0.003$). We calculate the population latent trait variance estimate after adjusting the person covariates under the LLLA model by $\hat{\sigma}_M^2 = \hat{\sigma}_0^2 + \hat{\sigma}_0^4 \text{var}(T)$ and the result is $\hat{\sigma}_M^2 = 0.788$. The estimate of the population latent trait variance parameter given by GLMM (PROC NLMIXED) ($\hat{\sigma}^2 = 0.851$, $SE = 0.086$).

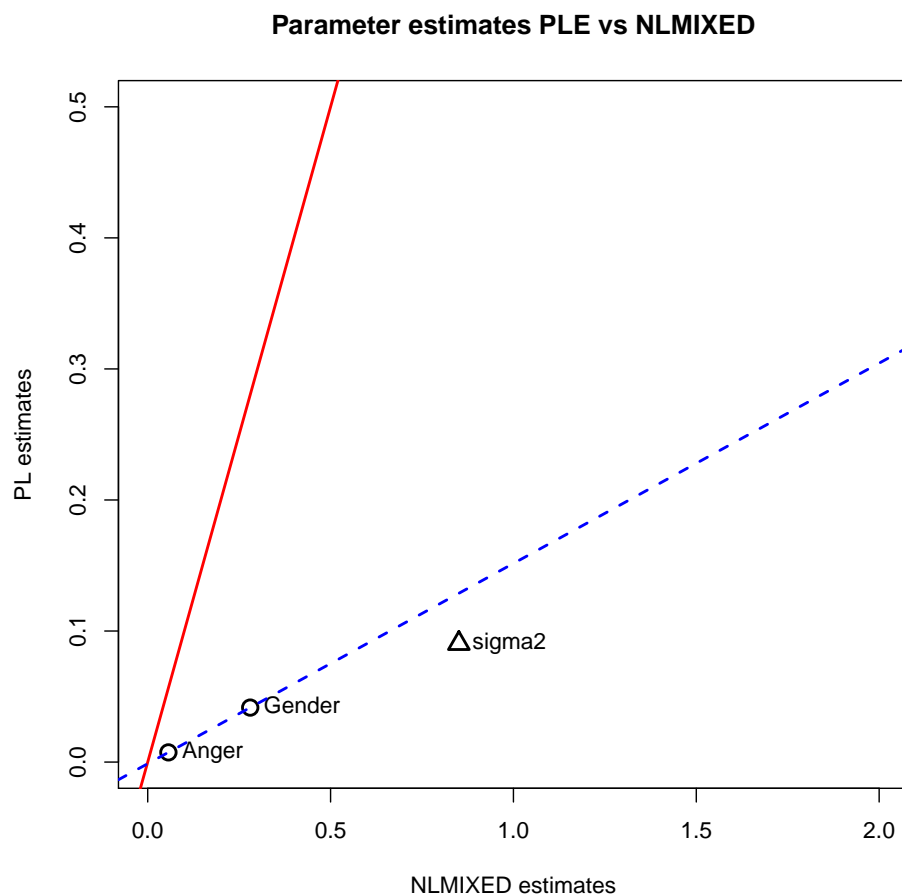


Figure 37. Person covariate effects estimates for the partial credit model with person covariates, LLLA (PL) vs GLMM (NLMIXED).

Chapter 9

Conclusion

In this thesis, we have seen the development of the models under the log-linear-as-IRT model framework. Under this framework, we started from an IRT model, as specified by the item response function, in the form of conditional probabilities of the response to the items given the latent trait. Then we utilized the Dutch Identity theorem, including the polytomous Dutch Identity theorem proved in this thesis, to derive the manifest probability of the responses. The resulting manifest probability is the log linear-by-linear association model (LLLA), which is a special form of log-linear model with second order interactions. By fitting the LLLA model, we fit the IRT model. Since MLE for LLLA is limited to small numbers of items, a pseudolikelihood estimation method was proposed for parameter estimation. PLE has the properties of consistency and asymptotic normality, which are also shared to MLE; and as demonstrated by the simulation studies and application to the example data, the lost of efficiency of PLE relative to MLE is negligible. A great advantage of PLE over MLE is its low computational demands that make it practical to fit the models on tests with large numbers of items.

In this final chapter, I would like to summarize the advantages of the log-linear-as-IRT method, followed by discussion of restrictions of the method, and point out directions for future research.

Flexibility of the Models

One of the features of the log-linear-as-IRT methods is its flexibility. Under this framework, many IRT models have equivalent LLLA models and can be fit by a single log-linear model procedure. I have implemented the method in an R package ‘plRasch’. The families of IRT models that can now be fit by ‘plRasch’ are listed in Table 24. The method is general and can also be implemented in other statistical programs that are favored by

a researcher. These IRT models incorporate features including polytomous items, multiple latent traits, and item and person covariates. The list of the models will grow as more research is conducted in this area.

Table 24

Models That can be Fit by R Package ‘plRasch’

IRT model	LLLA model
<i>Unidimensional, Dichotomous</i>	
Rasch model	LLLA
LLTM	LLLAi
Latent regression Rasch model	LLLAp
<i>Unidimensional, Polytomous</i>	
Partial credit model (PCM)	LLLA
PCM with item covariates (PCMi)	LLLAi
PCM with person covariates (PCMp)	LLLAp
<i>Multidimensional, Dichotomous</i>	
Multidimensional Rasch	LLLA
MultiD Rasch with item covariates	LLLAi
MultiD Rasch with person covariates	LLLAp
<i>Multidimensional, Polytomous</i>	
Multidimensional polytomous Rasch	LLLA
MultiD polytomous Rasch with item covariates	LLLAi
MultiD polytomous Rasch with person covariates	LLLAp

Flexibility on Latent Trait Distribution

Another flexibility of LLLA models comes from its assumptions on the distribution of the latent traits. Under LLLA models, the latent trait distribution is a mixture of normal distributions. The mixture distribution can be used to approximate many distributions that may appear in real applications. It is especially useful when the composition of the population is highly heterogeneous resulting in an irregularly shaped latent trait distribution. Under such a situation, it is better to use a mixture-of-normals distribution as assumed by

LLLA models rather than to use a normal distribution for the latent traits as assumed by MML based methods. While the mixture-of-normals distribution in LLLA models is not as flexible as in CML based methods, where no assumptions on latent trait distributions are imposed at all, CML methods do require the existence of sufficient statistics for the person parameters and that the sufficient statistics are unrelated to the item parameters. Thus CML methods are only applicable to a limited family of models including the Rasch model, LLTM, PCM, and PCMi. CML cannot be applied to the IRT 2PL model, Bock's model, or IRT models with person covariates, where such sufficient statistics do not exist.

Pseudolikelihood is Fast

As demonstrated by simulation studies and applications to the example data, pseudolikelihood estimation runs very fast. It has a clear advantage to the MLE, where the computational cost grows exponentially as the number of items increases. Indeed it is only possible to use MLE to fit the LLLA models when the number of items is small. It is faster than MML based methods, such as the generalized (non)linear mixed effect models, where the person parameters have to be integrated out numerically.

The pseudolikelihood functions for LLLA models have the same form as the likelihood functions of a logistic regression model (for dichotomous items) or a multinomial logit model (for polytomous items). With this fact, existing algorithms to maximize the likelihood functions for logistic regression models or multinomial logit models can be used to maximize the pseudolikelihood functions for LLLA models. This saves the cost of implementing the maximum pseudolikelihood method from scratch, such as with algorithms employing the Newton-Raphson method.

Non-collapsibility of LLLA Models

One of the issues that have been raised against LLLA models is that LLLA models are not collapsible. Suppose a test with I items follows an LLLA model with I items, if we remove an item from the test, the remaining $(I - 1)$ -item test will not exactly follow an LLLA model with $(I - 1)$ items. On the other hand, the Rasch model does not have this problem, after removing an item, the test that before followed the Rasch model still follows the Rasch model. The non-collapsibility of the LLLA model actually lies in its use of a mixture of normal distributions to describe the latent trait distribution. As pointed out in the simulation examples at the end of Chapter 2, the number of components in the mixture distribution is equal to the number of items plus 1 (which is the number of total scores). Suppose the population distribution of the latent traits is exactly an $(I + 1)$ -component mixture of normal distribution (as assumed by the I -item LLLA model), then it cannot be exactly an I -component mixture of normal distribution (as assumed by the $(I - 1)$ -item LLLA model). However, we are using the mixture of normal distribution to *approximate* the latent trait distribution. As demonstrated in the simulation studies in Chapter 2 and Chapter 7, none of them had the latent trait generated from an $(I + 1)$ -component mixture of normal distribution, thus the simulated data did not exactly follow the LLLA model. However, by fitting the LLLA model on the simulated data we were able to recover the true item parameters. Non-collapsibility of the LLLA model does not prevent us obtaining valid inference on item parameters and estimating the person parameters. Therefore the non-collapsibility is not an actual problem.

Interpretation of the LLLA Parameters

The equivalence between many IRT models and LLLA models as presented in this thesis has given the interpretations of the parameters in the LLLA models under the context of IRT models. As we interpret parameters in LLLA models, we must keep in mind that some

parameters, such as the variance parameter and the person covariate effects, are conditional on response patterns. To get the corresponding marginal parameters that are appropriate for the whole population, one needs to utilize the relationships developed in this thesis.

Differential Item Functioning

One interesting topic that is not addressed in this thesis is differential item functioning (DIF) (Holland & Wainer, 1993). DIF, also called item bias, describes the phenomenon that some items have difficulty parameters that are biased against certain groups of examinees. Therefore detecting biased items and removing them from the test is important for maintaining the fairness and validity of the test. Many methods have been proposed to detect DIF (Holland & Wainer, 1993). One way is to treat the DIF effects as the interaction between the item to the person group covariate (De Boeck & Wilson, 2004). Therefore we can use an IRT model that contains the item and person covariate interaction, and use the Dutch Identity to derive its equivalent LLLA model. If we use PLE to fit the model, then the hypothesis testing problem of the existence of DIF effect will require the development of the test procedures for pseudolikelihood estimation. Geys, Molenberghs, and Ryan (2002b) give examples of pseudolikelihood based Wald, score and likelihood ratio tests, and the idea can be applied for testing DIF effects in our model.

Estimation for IRT 2PL Model and Bock's Model

The pseudolikelihood methods presented in this thesis are applicable to LLLA models, where the score parameters are fixed to specific values. The score parameters in more general LMA models correspond to the discrimination parameters in the IRT 2PL model for dichotomous items and Bock's model for polytomous items. When scores are unknown parameters, it is possible to fit the LMA model by MLE as implemented in the statistical package ℓ_{EM} (Anderson and Vermunt, 2000). However, since MLE only works for data with

small numbers of items, it is worthwhile to develop estimation methods for the LMA models that are equivalent to the IRT 2PL model and Bock's model so that they can be used for tests with large number of items.

To summarize, the log-linear-as-IRT methods in this thesis have provided useful and efficient tools to fit many IRT models for dichotomous or polytomous items, for uni- or multidimensional latent traits, and with item or person covariates. The research will continue to grow and surpass over the limitations of the current methodology and contribute to the field.

References

- Ackerman, T. (1994). Using multidimensional item response theory to understand what items and tests are measuring. *Applied Measurement in Education*, 7(4), 255–278.
- Agresti, A. (2002). *Categorical data analysis*. Hoboken, NJ: Wiley-Interscience.
- Anderson, C., & Böckenholt, U. (2000). Graphical regression models for polytomous variables. *Psychometrika*, 65(4), 497–509.
- Anderson, C., Li, Z., & Vermunt, J. (2007). Estimation of models in a Rasch family for polytomous items and multiple latent variables. *Journal of Statistical Software*, 20(6).
- Anderson, C., Verkuilen, J., & Peyton, B. (2010). Modeling polytomous item responses using simultaneously estimated multinomial logistic regression models. *Journal of Educational and Behavioral Statistics*.
- Anderson, C., & Vermunt, J. (2000). Log-multiplicative association models as latent variable models for nominal and/or ordinal data. *Sociological Methodology*, 30, 81–122.
- Anderson, C., & Yu, H. (2007). Log-multiplicative association models as item response models. *Psychometrika*, 72(1), 5–23.
- Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika*, 43(4), 561–573.
- Arnold, B., & Strauss, D. (1991). Pseudolikelihood estimation: Some examples. *Sankhya B*, 53, 233–243.
- Baker, F., & Kim, S. (2004). *Item response theory: Parameter estimation techniques*. New York, NY: Marcel Dekker.
- Besag, J. (1975). Statistical analysis of non-lattice data. *The Statistician*, 179–195.
- Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F. Lord & M. Novick (Eds.), *Statistical theories of mental test scores* (pp. 395–479). Reading, MA: Addison-Wesley.
- Bock, R. (1972). Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika*, 37(1), 29–51.
- Böckenholt, U., & Böcknholt, I. (1990). Canonical analysis of contingency tables with linear constraints. *Psychometrika*, 55(4), 633–639.
- Boomsma, A., Duijn, M. van, & Snijders, T. (2001). *Essays on item response theory*. New York, NY: Springer Verlag.
- Chang, H. (1996). The asymptotic posterior normality of the latent trait for polytomous IRT models. *Psychometrika*, 61(3), 445–463.

- Chang, H., & Mazzeo, J. (1994). The unique correspondence of the item response function and item category response functions in polytomously scored item response models. *Psychometrika*, 59(3), 391–404.
- Chang, H., & Stout, W. (1993). The asymptotic posterior normality of the latent trait in an IRT model. *Psychometrika*, 58(1), 37–52.
- Clogg, C. (1981). New developments in latent structure analysis. In D. Jackson & E. Borgotta (Eds.), *Factor analysis and measurement in sociological research* (p. 215–246). Beverly Hills: Sage Publications.
- Cressie, N., & Holland, P. (1983). Characterizing the manifest probabilities of latent trait models. *Psychometrika*, 48(1), 129–141.
- Dawid, A. P. (1970). On the limiting normality of posterior distributions. In *Proceedings of the cambridge philosophical society* (Vol. 67, p. 625–633). Cambridge, UK: Cambridge University Press.
- de Leeuw, J., & Verhelst, N. (1986). Maximum likelihood estimation in generalized Rasch models. *Journal of Educational and Behavioral Statistics*, 11(3), 183.
- De Boeck, P., & Wilson, M. (Eds.). (2004). *Explanatory item response models: A generalized linear and nonlinear approach*. New York, NY: Springer.
- Efron, B., & Tibshirani, R. (1997). *An introduction to the bootstrap*. Boca Raton, FL: Chapman & Hall.
- Embretson, S., & Reise, S. (2000). *Item response theory for psychologists*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Fischer, G. (1973). The linear logistic test model as an instrument in educational research. *Acta Psychologica*, 37(6), 359–374.
- Fischer, G., & Molenaar, I. (Eds.). (1995). *Rasch models: Foundations, recent developments, and applications*. New York, NY: Springer.
- Geys, H., Molenberghs, G., & Ryan, L. M. (2002a). Pseudo-likelihood estimation. In M. Aerts, G. Molenberghs, & L. M. Ryan (Eds.), *Topics in modelling of clustered data* (p. 89–114). Boca Raton, FL: Chapman & Hall/CRC.
- Geys, H., Molenberghs, G., & Ryan, L. M. (2002b). Pseudo-likelihood inference. In M. Aerts, G. Molenberghs, & L. M. Ryan (Eds.), *Topics in modelling of clustered data* (p. 115–126). Boca Raton, FL: Chapman & Hall/CRC.
- Gilula, Z., & Haberman, S. (1994). Conditional log-linear models for analyzing categorical panel data. *Journal of the American Statistical Association*, 89(426).
- Gilula, Z., & Haberman, S. (1995). Dispersion of categorical variables and penalty functions: Derivation, estimation, and comparability. *Journal of the American Statistical Association*, 90(432).

- Goodman, L. (1974). Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika*, 61(2), 215–231.
- Hambleton, R., Rogers, H., & Swaminathan, H. (1995). *Fundamentals of item response theory*. Newbury Park, CA: Sage Publications.
- Holland, P. (1990). The Dutch Identity: A new tool for the study of item response models. *Psychometrika*, 55(1), 5–18.
- Holland, P., & Wainer, H. (Eds.). (1993). *Differential item functioning*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Joe, H., & Liu, Y. (1996). A model for a multivariate binary response with covariates based on compatible conditionally specified logistic regressions. *Statistics & probability letters*, 31(2), 113–120.
- Kelderman, H., & Rijkes, C. (1994). Loglinear multidimensional IRT models for polytomously scored items. *Psychometrika*, 59(2), 149–176.
- Lauritzen, S., & Wermuth, N. (1989). Graphical models for associations between variables, some of which are qualitative and some quantitative. *Annals of Statistics*, 17(1), 31–57.
- Lehmann, E., & Casella, G. (1998). *Theory of point estimation* (2nd ed.). New York, NY: Springer-Verlag.
- Li, Z., & Hong, F. (2007). *plRasch*: an R package for fitting Log Linear by Linear Association Models using pseudolikelihood estimation. <http://cran.r-project.org/web/packages/plRasch/>.
- Lord, F. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Lord, F., & Novick, M. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.
- Masters, G. (1982). A Rasch model for partial credit scoring. *Psychometrika*, 47(2), 149–174.
- Mellenbergh, G. (1995). Conceptual notes on models for discrete polytomous item responses. *Applied Psychological Measurement*, 19(1), 91.
- Muraki, E. (1990). Fitting a polytomous item response model to Likert-type data. *Applied Psychological Measurement*, 14(1), 59.
- Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *Applied psychological measurement*, 16(2), 159.
- Nering, M., & Ostini, R. (Eds.). (2010). *Handbook of polytomous item response theory models*. New York, NY: Taylor & Francis Group.
- Ostini, R., & Nering, M. (2006). *Polytomous item response theory models*. Thousand Oaks, CA: Sage Publications.

- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Danish Institute for Educational Research.
- Rasch, G. (1961). On general laws and the meaning of measurement in psychology. In *Proceedings of the fourth Berkeley symposium on mathematical statistics and probability* (Vol. 4, pp. 321–333). Berkeley and Los Angeles, CA: University of California Press.
- Reckase, M. (1997a). A linear logistic multidimensional model for dichotomous item response data. In W. Van der Linden & R. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 271–286). New York, NY: Springer Verlag.
- Reckase, M. (1997b). The past and future of multidimensional item response theory. *Applied Psychological Measurement*, 21(1), 25.
- Reckase, M. (2009). *Multidimensional item response theory*. New York, NY: Springer Verlag.
- Samejima, F. (1969). *Estimation of latent ability using a response pattern of graded scores*. Iowa City, IA: Psychometric Society.
- Samejima, F. (1970). Estimation of latent ability using a response pattern of graded scores. *Psychometrika*, 35(1), 139–139.
- Shao, J., & Tu, D. (1995). *The jackknife and bootstrap*. New York, NY: Springer.
- Smit, A., & Kelderman, H. (2000). Pseudolikelihood estimation of the Rasch model. *Journal of Outcome Measurement*, 4(1), 513–23.
- Strauss, D. (1992). The many faces of logistic regression. *American Statistician*, 321–327.
- Tettegah, S., & Anderson, C. (2007). Pre-service teachers empathy and cognitions: Statistical analysis of text data by graphical models. *Contemporary Educational Psychology*, 32(1), 48–82.
- Thissen, D., & Steinberg, L. (1986). A taxonomy of item response models. *Psychometrika*, 51(4), 567–577.
- Tutz, G. (1990). Sequential item response models with an ordered response. *British Journal of Mathematical and Statistical Psychology*, 43(1), 39–55.
- van der Ark, L. (2001). Relationships and properties of polytomous item response theory models. *Applied Psychological Measurement*, 25(3), 273–282.
- Van der Linden, W., & Hambleton, R. (Eds.). (1997). *Handbook of modern item response theory*. New York, NY: Springer Verlag.
- Vermunt, J. (1997). Lem: A general program for the analysis of categorical data [Computer software manual].
- von Davier, M., & Carstensen, C. (Eds.). (2007). *Multivariate and mixture distribution Rasch models: Extensions and applications*. New York, NY: Springer Verlag.

- White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica: Journal of the Econometric Society*, 50(1), 1–25.
- Whittaker, J. (1990). *Graphical models in applied multivariate statistics*. New York, NY: Wiley.
- Wong, R. (1995). Extensions in the use of log-multiplicative scaled association models in multiway contingency tables. *Sociological Methods & Research*, 23(4), 507.
- Zhang, J., & Stout, W. (1997). On Holland’s Dutch identity conjecture. *Psychometrika*, 62(3), 375–392.
- Zwinderman, A. (1991). A generalized Rasch model for manifest predictors. *Psychometrika*, 56(4), 589–600.
- Zwinderman, A. (1995). Pairwise parameter estimation in Rasch models. *Applied Psychological Measurement*, 19(4), 369.