# Discrete latent variable models

Heinen, A.G.J.J.

*Publication date:*
1993

*Document Version*
Publisher's PDF, also known as Version of record

Link to publication in Tilburg University Research Portal

*Citation for published version (APA):*
Heinen, A. G. J. J. (1993). *Discrete latent variable models*. Tilburg University Press.

Work and Organization Research Centre

# Discrete Latent Variable Models

Ton Heinen

SERIES ON WORK AND ORGANIZATION

# Discrete Latent Variable Models

Promotor: Prof. dr. J.A.P. Hagenaars
Co-promotor: Dr. M.A. Croon


Printed in The Netherlands.

# Discrete Latent Variable Models

Proefschrift

ter verkrijging van de graad van doctor
aan de Katholieke Universiteit Brabant, op gezag van
de rector magnificus, prof. dr. L.F.W. de Klerk,
in het openbaar te verdedigen ten overstaan van
een door het college van dekanen aangewezen commissie
in de aula van de Universiteit
op vrijdag 10 september 1993 te 16.15 uur

door

Antonius Gerardus Joseph Johannes Heinen

geboren te Waalwijk

Tilburg
University
Press
1993

**Acknowledgements**

Several people have made very valuable contributions to this study. I am indebted to Jacques Hagenaars and Marcel Croon for their critical comments and their stimulating support during the course of this study. Thanks are due to Paul Vermaseren for writing the DILTRAN program. Mark Vitullo did an excellent job in editing the text. I am grateful for the financial support received from WORC (Work and Organization Research Centre). Finally, I would like to thank all my collegues at the Methodology Department of the Faculty of Social Sciences at Tilburg University for providing me the opportunity to write this thesis.

# Contents

# Chapter 1

# Introduction

In the last few decades, several types of latent variable models have been proposed. It is assumed in some of these models (for example, factor analysis) that the observed variables are essentially continuous. Other latent variable models, on the other hand, deal explicitly with discrete observed variables. Both latent class models and latent trait models belong to the latter category. In this book, attention is focused on these two models. The relationship between latent class analysis and latent trait models is explored in depth. By doing so, the results achieved in the one field of research can be utilized in the other.

Latent variable models have achieved great popularity in the social sciences. Bartholomew (1987) mentioned two reasons for the acceptance of models with latent variables. Firstly, measurement in the social sciences is hampered by the nature of the concepts which are of interest to researchers in this field. Many concepts that play a crucial role in social and behavioral theories cannot be observed directly. There is no way to infer directly how intelligent a certain person is or what the socio-economic status of a particular family is. The only way empirical knowledge can be obtained about these concepts is to look for variables that *can* be observed directly and which contain some information about the theoretical concepts. Hence, in many cases the theoretical concepts themselves are not measured directly. Rather, some observable variables that are thought to represent certain aspects of the theoretical concepts, or are believed to be influenced by those theoretical variables are observed empirically. Intelligence, for instance, can be measured by posing specific problems that can be solved through abstract reasoning because it is believed that individuals who are more intelligent will show greater ability at solving these problems correctly. In other words, the probability of responding correctly to these problems is believed to be influenced by a person's intelligence. In the case of socio-economic status, one observable variable that could be used is the income of the head of the family since this variable is thought to be an important aspect of the theoretical variable "socio-economic status". Such a measurement process is, for obvious reasons, sometimes called *indirect* observation. The theoretical variables that are not observed directly are then denoted by the term *latent variables*, while the variables which are observed directly and contain information on the latent variables are called *manifest variables* or *indicators*.

The second rationale Bartholomew (1987) gave for using latent variable models is more pragmatic. In many examples of social research, so many different variables are measured that it becomes necessary to compress this data into a smaller set of variables which are assumed to reflect the common substance of a number of original variables. The use (and misuse) of factor analysis is an excellent example of this practical use of latent variable models. It should, however, be stressed that when latent variable models are used for mere data reduction, the resulting clusters of observed variables are interpreted substantively. This is also true in the case at hand in which directly observed variables are used to indicate theoretical concepts, albeit in a more exploratory manner. As Clogg (1988) noted, many applications in latent class analysis also aim at this more exploratory data reduction. The use of latent class models for confirmatory analyses, i.e., the testing of explicit measurement models, has only recently received more attention.

The distinction between latent and manifest variables is, of course, essential to latent variable models. When, however, this distinction is used as the only defining element of latent variable models, the resulting class of latent variable models is so broad that further structuring is required. Furthermore, by organizing the great variety of latent variable models through the introduction of a number of classification principles, the scope of this book can be indicated more precisely.

This book concentrates on two types of latent variable models, namely, latent class and latent trait models. An attempt is made to indicate what the common elements of both types of models are; the situations in which these two types of models will yield the same results or comparable results are also investigated. Both models are often considered as members of a broader set of latent variable models, named *latent structure analysis*. In the following sections classification criteria are introduced which can be used to determine whether latent variable models belong to the class of latent structure models. The class of latent structure models are also internally structured by making a distinction according to the measurement level of both the latent and the manifest variables. Furthermore, examples of the different latent structure models that can be distinguished according to the proposed classification criteria are presented. In this way, it will become clear which types of models are discussed in this book and which models will not be dealt with. Finally, some remarks are made regarding the differences between latent class models and latent trait models. This highlights some of the issues that will be explored more deeply in the chapters to come.

## 1.1 Latent structure models

Most principles for classifying latent variable models that have thus far been proposed in the literature are based on a distinction between types of manifest and latent variables. Bartholomew (1987), for example, classifies both the latent and the manifest variables as either *metrical* or *categorical*. The term metrical is reserved for variables that can take on values in the set of real numbers. These variables can be continuous or discrete. Categorical variables are always discrete as they allocate subjects to one of a set of exclusive categories. In other words, categorical variables are variables measured on a nominal scale. If both the latent variable and the manifest variable can be either metrical or categorical, four types of latent variable models can be distinguished. Furthermore, Bartholomew reserves the phrase *latent structure analysis* for all models which use categorical latent variables, regardless of the nature of the manifest variables.

This latter definition differs from the position originally taken by Lazarsfeld and Henry (1968), who were the first to study latent structure analysis in depth. In their view, latent structure models are models which introduce latent variables to account for the observed pattern of association between the manifest variables. This is done by formulating a mathematical model that relates the latent variable to scores or categories of the manifest variables. Furthermore, these relations between the latent and the manifest variables are stochastic. The fact that the indicators contain only partial information regarding the theoretical concept, which makes the relationship between these two types of variables not unequivocal, is accounted for by stating a stochastical or probabilistical relation between these variables. The most important assumption of latent structure models is that these stochastic relations are governed by the axiom of *local independence* (Lazarsfeld and Henry, 1968, p.22).[1] Lord and Novick (1968) and Langeheine and Rost (1988) also consider local independence as the defining characteristic of latent structure analysis. The significance of this axiom will be made clear in the next section, after the type of models that belong to the class of latent structure models have been more precisely defined.

As was stated before, the definition given by Bartholomew will not be used in this study. Bartholomew's method of defining latent structure models excludes latent variable models with continuous latent variables. This type of model was, however, explicitly dealt with by Lazarsfeld and Henry (1968); examples are the latent content model, polynomial trace line models and test theory models. These latter models (i.e., test theory models) were only briefly touched upon in the book by Lazarsfeld and Henry, but they have become very popular since then under the heading *latent trait models*, and nowadays it is very common to regard the class of latent trait models as members of the family of latent structure models. Therefore, the definition of latent structure models given by Lazarsfeld and Henry has been adopted in this study. This includes models for both continuous and discrete variables at the latent and the manifest level.

This book deals with only certain specific latent structure models, namely, latent class models and latent trait models. A common element of these models is that they treat the manifest indicators as discrete variables, though they can be measured at every measurement level (i.e., nominal, ordinal and metric level). This *excludes* covariance-structure models (also called structural equation models) such as factor analysis and the more "traditional" LISREL models. Of course,

[1]Local independence can be stated both as an axiom or as an assumption. See, for example, Andersen (1988) for further details.

more recent developments in covariance structure analysis also allow for the analysis of discrete manifest data by analyzing the matrix of polychoric correlations with weighted least squares methods (Jöreskog and Sörbom, 1988). As covariance structure models will not be dealt with, these more recent LISREL models will likewise not be discussed in this book.

The restriction of dealing only with discrete manifest data also excludes the latent profile model proposed by Lazarsfeld and Henry (1968). The latent structure models that are of interest can now be classified by taking the measurement level into account for both the latent and the manifest variables. This is done in Table 1.1, where a number of cells are filled with names of already known latent structure models. Other cells are empty as no models have yet been suggested that belong to the combination of measurement levels for the latent and the manifest variables. It should be noted that distinguishing between different levels of measurement for the manifest variables is especially relevant when these manifest indicators are polytomous, i.e., take more than two different values, because, when the manifest variables are dichotomous, it does not make much sense to distinguish between nominal and metrical measurement levels.

Despite the fact that latent variables can only be observed indirectly, it makes sense to discuss the measurement level of latent variables since the measurement level assumed for these latent variables has direct consequences for the structure of the measurement model. It is necessary to know what the characteristics are of the numerical values that the latent variable can take in a specified situation for the latent structure models that will be discussed.

The various entries in Table 1.1 will be discussed very briefly in Section 1.3. A more thorough treatment of the majority of the models mentioned in this table will be presented in the following chapters. Before this is done, however, the principle of local independence will be outlined.

## 1.2 Local independence

All latent structure models assume the existence of a latent variable. Sometimes the focus is restricted to one single latent variable, as is

Table 1.1: A typology of latent structure models with discrete manifest variables

| | | Manifest variable: Nominal | Ordinal | Metrical |
|---|---|---|---|---|
| Latent Variable | Nominal | LCA | | LCA for rating data (Rost) |
| | Ordinal | | LCA with ordered classes (Croon) | |
| | Metrical | LCA with linear restrictions (Haberman) Nominal Response Model (Bock) | Graded Response Model (Samejima) | LCA with linear by linear restrictions (Haberman) Partial Credit Model (Masters) Rating Scale Model (Andrich) |

the case in most latent trait models. In other latent structure models, the existence of several latent variables is postulated; this is the case in many latent class models. In every latent structure model, however, the association between the manifest variables is assumed to depend on the relationship between the manifest and the latent variables. Thus, it is assumed that all the associations among the manifest variables can be explained by the dependence of these manifest variables upon the latent variable(s). In other words, when the latent variable is held constant, the manifest variables should be statistically independent. This basic assumption of the latent structure model as developed by Lazarsfeld and Henry is known as the assumption of local independence.

This assumption can easily be formalized. Assume that $n$ different manifest indicators are measured and that, for the sake of convenience, these indicators all are dichotomous.[2] Each item $j$ can take only the values 0 or 1 and $j = 1, \ldots, n$. Furthermore, it is assumed that there is a latent variable $\theta$ which can be continuous or discrete. When this variable is metrical and continuous, the latent score for subject $i$ on this latent variable will be denoted by $\theta_i$. When the latent variable is discrete, the number of categories for the latent variable will be equal to $T$ and an arbitrary category will be indicated as $t$. When the latent variables are discrete, the latent structure models will be parameterized using a fixed number of latent classes (equal to $T$), instead of the individual latent scores $\theta_i$. The notation $\theta_t$ is used to indicate an arbitrary latent class. In some cases, these latent classes are only nominal, then the numbers $t$ (with $t = 1, \ldots T$) only serve to assign an arbitrary integer to these nominal latent classes. In other circumstances, however, $\theta$ is thought to be discrete but metrical. In those situations, the $T$ different numbers $\theta_t$ should be interpreted as metrical values. When a subject responds to the $n$ different items, there are of course $2^n$ different response patterns that could result. An arbitrary response pattern is denoted by the symbol $\nu$.

The probability that individual $i$ with latent score $\theta_i$ will obtain score 1 on item $j$ is denoted by $p_{j1|\theta_i}$. When $\theta$ is discrete, this response probability for all individuals belonging to the latent class $\theta_t$ is written as $p_{j1|\theta_t}$. For the two situations that are distinguished here ($\theta$ continuous or discrete), the conditional probability that response pattern $\nu$ is observed is denoted by $p_{\nu|\theta_i}$ and $p_{\nu|\theta_t}$ respectively. Finally, a set of indicator variables $x_{\nu j}$ is introduced. If in response pattern $\nu$ item $j$ is responded to in category 1 , then $x_{\nu j} = 1$; otherwise $x_{\nu j} = 0$.

Because the assumption of local independence states that the manifest items are statistically independent for individuals with the same position on the latent variable, the conditional probability for observing response pattern $\nu$ can now be expressed as the product of the conditional response probabilities for the separate items:

$$p_{\nu|\theta_i} = \prod_{j=1}^{n} (p_{j1|\theta_i})^{x_{\nu j}} \cdot (1 - p_{j1|\theta_i})^{1-x_{\nu j}}, \tag{1.1}$$

[2]The assumption of dichotomous indicators is only made because it leads to convenient notation. The same principles hold for polytomous items. The formal expressions that result from the more general polytomous situation can be found in Chapter 2.

in case $\theta$ is continuous, and as:

$$p_{\nu|\theta_t} = \prod_{j=1}^{n} (p_{j1|\theta_t})^{x_{\nu j}} \cdot (1 - p_{j1|\theta_t})^{1-x_{\nu j}} \tag{1.2}$$

for the situation where $\theta$ is discrete.

Thus, when local independence holds, knowledge of the conditional response probabilities $p_{j1|\theta_i}$ or $p_{j1|\theta_t}$ is sufficient for calculating the conditional probability of observing an arbitrary response pattern $\nu$.

As Clogg (1988) has shown, when local independence holds for the set of $n$ indicators, it will also hold for any subset of these manifest variables. Collapsing response categories of the manifest variables will also not distort the pattern of local independence. However, the reverse need not be true. When local independence holds for variables with certain categories collapsed, the assumption need not necessarily be satisfied for the original variables.

The observation that collapsing categories of manifest variables has no impact on the local independence cannot, however, be generalized to collapsing values of the latent variable. In general, when local independence holds for a continuous latent variable $\theta$, it is not possible to group or condense $\theta$ without distorting local independence. But as Clogg (1988) stated, in many cases it should be possible to divide the latent continuum into disjoint sets, such that the original relation between $\theta$ and the manifest variables will be approximately retained. This practice of discretizing the latent space into a set of discrete latent classes will be used intensively for the remainder of this study.

The principle of local independence can be used in both unidimensional and multidimensional latent structure models. This is not always recognized, as a number of authors, in particular in the literature on latent trait models are inclined to define local independence as a special case of unidimensionality of the latent space (see, for example, Hambleton & Swaminathan, 1985; Lord & Novick, 1968; and Kelderman, 1984). The case of multidimensional latent trait models is one of the topics covered in Chapter 5. How local independence works in a multidimensional latent structure model will be indicated here briefly.

In Figure 1.1, a multidimensional latent structure model with two latent variables is shown. The line connecting the two latent variables indicates that these two variables are correlated. The question of whether one latent variable is causally dependent on the other is left open. The arrows between the latent variables and the manifest indicators signify that the manifest variables are dependent on the latent variables. These arrows can be interpreted as direct effects of the latent variables on the indicators.

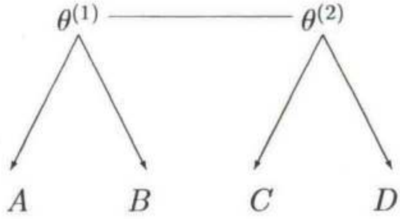


Figure 1.1: A multidimensional latent structure model

According to the relations sketched in Figure 1.1., the manifest indicators $A$ and $B$ serve only as indicators for the latent variable $\theta^{(1)}$, while the latent variable $\theta^{(2)}$ is measured through the manifest variables $C$ and $D$. It is clear that for the subset of variables $\theta^{(1)}$, $A$ and $B$, the principle of local independence will hold because $A$ and $B$ are only dependent on $\theta^{(1)}$, so if this latent variable is held constant, the manifest variables $A$ and $B$ will be independent. Of course, the same is true for the other part of this multidimensional model, i.e., for the variables $\theta^{(2)}$, $C$ and $D$. Thus, the association between $A$ and $B$ is explained by $\theta^{(1)}$, while the association between $C$ and $D$ is accounted for by $\theta^{(2)}$.

It is also possible to look at the association between manifest variables that indicate different latent variables, for example, $A$ and $C$. Because $A$ is only "caused" by $\theta^{(1)}$ and the only variable having a causal influence on $C$ is $\theta^{(2)}$, holding either $\theta^{(1)}$ or $\theta^{(2)}$ constant will make all association between $A$ and $C$ disappear. Finally, if manifest variables depend on two or more latent variables, all these latent variables must be held constant in order to make the association between the indicators disappear.

In general, the association between manifest indicators will disappear if the relevant latent variables (i.e., latent variables that have a causal influence on the manifest variables under consideration) are held constant. It follows that the association between indicators will also vanish if all latent variables (relevant or not) are held constant. Although these observations are common knowledge among those who are familiar with structural equation models, these facts are hardly ever stressed by item response theorists as item response theory is strongly dominated by models for the analysis of unidimensional latent traits.

## 1.3 Latent structure models for discrete data

In Table 1.1, a number of latent structure models were presented that can be used for the analysis of discrete manifest data. In this section, these models will be looked at more closely. As stated earlier, a more thorough treatment is postponed until Chapters 2 and 3.

The starting point for this discussion will be the top-left cell in Table 1.1. Situated in this cell is the standard *latent class analysis* (LCA) model, in which it is assumed that latent and manifest variables are nominal. This model was originally proposed by Lazarsfeld and Henry (1968), but its great popularity is due mainly to the work of Goodman, Haberman and Clogg (see, for example, Goodman, 1974a; Haberman, 1979; Clogg, 1981).

### 1.3.1 The unrestricted latent class model

The latent class model assumes that the population of subjects is divided into a number of $T$ exclusive and exhaustive latent classes. Each individual belongs to one and only one latent class. Within each latent class, the manifest variables are statistically independent. This assumption is the assumption of local independence. In other words, the association between the manifest indicators is assumed to be caused by the fact that people belong to different latent classes and that different classes have different conditional response probabilities. The membership in a particular latent class defines the probabilities for responding to items in specific categories. All individuals who belong to the same latent class have equal probabilities for responding to the items in certain categories. In this sense, the latent class model is a finite-mixture model (McLachlan & Basford, 1988), because the total population is a mix of a finite number of latent classes, which differ not only with respect to the conditional response probabilities, but also with respect to their relative sizes. This latter aspect will be taken into account in the notation used here by symbolizing the probability for observing a subject that belongs to latent class $t$ by $p_t^{\theta}$. The latent class model can now be parameterized by using the conditional response probabilities and the latent proportions. The probability that a subject will respond in category $g$ of the manifest item $A$, given that this subject belongs to latent class $t$, will be denoted by $p_{gt}^{\overline{A}\theta}$, with the bar above the superscript indicating that this probability is a *conditional* probability and not a probability in the joint distribution of $\theta$ and the manifest item.[3] The superscripts pertain to the variable labels, while the subscripts denote the specific categories of the variables. It will now be assumed that there are three manifest variables $A$, $B$ and $C$. Arbitrary response categories for these three variables $A$, $B$ and $C$ will be denoted by $g$, $k$ and $\ell$ respectively. For this simple situation, the basic equation for the probability of responding to these three items $A$, $B$ and $C$ in the categories $g$, $k$ and $\ell$, respectively, is (Goodman, 1974a):

$$p_{gk\ell}^{ABC} = \sum_{t=1}^{T} p_{gk\ell t}^{ABC\theta}, \tag{1.3}$$

where:

$$\begin{aligned} p_{gk\ell t}^{ABC\theta} &= p_t^{\theta} p_{gk\ell t}^{\overline{ABC}\theta} \\ &= p_t^{\theta} p_{gt}^{\overline{A}\theta} p_{kt}^{\overline{B}\theta} p_{\ell t}^{\overline{C}\theta}. \end{aligned} \tag{1.4}$$

The first of these two equations reflects the assumption that the population consists of a number of different and internally homogenous latent classes which are mutually exclusive. Therefore, the probability of observing a specific response pattern $\nu$ can always be found by summation of the probabilities for observing this response pattern in the various latent classes. Owing to the assumption of local independence, these latter probabilities (in this example, $p_{gk\ell t}^{ABC\theta}$) can be expressed as a very simple function of the latent proportions (so $p_t^{\theta}$) and the conditional response probabilities (for example, $p_{gt}^{\overline{A}\theta}$).

[3]Note that a slightly different notation for the conditional response probabilities has been used here, because the latent class model has been denoted for a hypothetical situation with three manifest variables $A$, $B$ and $C$. In due time, the expression $p_{gt}^{\overline{A}\theta}$ will be replaced by the more general notation $p_{jg|\theta_t}$.

This method of expressing the latent class model has the advantage that only parameters which have an intuitive meaning, and can therefore be interpreted very directly, are used. These parameters are the latent distribution in terms of proportions and the probabilities for observing specific responses on each of the different manifest items conditional upon the latent classes.

Another (equivalent) way of parameterizing the latent class model in terms of the log-linear model was suggested by Haberman (1979). The *joint* probability that a subject belongs to latent class $t$ and responds to items $A$, $B$ and $C$ in the categories $g$, $k$ and $\ell$, respectively, is written as:

$$\begin{aligned}\ln p_{gk\ell t}^{ABC\theta} &= u + u_g^A + u_k^B + u_\ell^C \\ &\quad + u_t^\theta + u_{gt}^{A\theta} + u_{kt}^{B\theta} + u_{\ell t}^{C\theta}. \end{aligned} \tag{1.5}$$

The interpretation of the parameters in log-linear models will be postponed until Chapter 2. It suffices here to remark that the most interesting parameters are the two-variable interactions (for example, $u_{gt}^{A\theta}$), because these parameters describe the relationship between the latent variable $\theta$ and the manifest items. Furthermore, Equation 1.5 describes a nonsaturated (i.e., restricted) log-linear model, in which the restrictions that have been imposed all result from the assumption of local independence. The appeal of the log-linear formulation of the latent class model lies in the possibility of putting additional restrictions on the two-variable interactions. Some examples of such restricted latent class models will be presented later in this section.

The standard latent class model is a very flexible tool for analyzing structural relationships between categorical variables with both unidimensional and multidimensional latent variables (Hagenaars, 1990). It can be seen as a natural extension of the log-linear model in order to take measurement error into account. It has become very popular among social scientists as an instrument for data reduction (see, for example, Formann, 1985 and Aitkin, Anderson and Hinde, 1981). In the literature, the value of latent class analysis for measurement purposes, i.e., the possibility of assigning quantitative scores to subjects, has been questioned. As Clogg (1988) noted:

> ... *a careful examination of the latent class models now available shows that none deal in a direct way with measurement, particularly if exacting standards are used to define how measurement should take place.*

This statement rests on the premise that it is desirable to develop measurement models which can be used to assign subjects scores on the latent variable. However, assigning scores to individuals is not without problems. One objection that can be raised against such a procedure is that the process of assigning subjects scores on a latent variable is hampered both by the presence of measurement error and by a number of identification problems. This topic will be explored further in Chapter 5. Secondly, if one would like to assign scores on a latent variable just to be able to explore the relationship between latent variables and some external variables, it should be noted that these relations can be explored quite satisfactorily *without* this assignment. This topic also will be dealt with in Chapter 5.

Examining the relations between latent variables and external variables without assigning scores to individuals becomes possible when measurement models are formulated in which both the relations between the latent variables and the manifest indicators, and the relations between the latent variables and external variables are included. Putting restrictions on the conditional response probabilities or the log-linear parameters will result in a class of restricted latent class models which are valuable in this respect. These models will be studied extensively in Chapter 2. However, one set of possible restrictions will be introduced in this chapter as it leads to a number of latent class models which are mentioned in Table 1.1. By using the restrictions pointed out here one tries to take the order information in the observed variables into consideration by linearizing the relationship between the latent and the manifest variables. Some of these models will be presented below.

### 1.3.2 Linear relations between the latent and the manifest variables

The log-linear formulation of the latent class model as expressed in Equation 1.5 uses two-variable interactions (such as $u_{gt}^{A\theta}$) to describe the relation between the latent and the manifest variables. The definition of these interactions, as well as the method by which they can be interpreted substantially, will be postponed until Chapter 2 but it is necessary to discuss here some characteristics of these log-linear parameters in order to make clear how the relation between the latent and the manifest variables can be linearized. Restricted latent class models with linear relations between the latent and the manifest variables can be found in the bottom row and the far right-hand column of Table 1.1.

Suppose that the categories for item $A$ have been labeled with successive integers $0, \ldots, g, \ldots, m_A$, so that item $A$ has $(m_A + 1)$ distinct categories. As was done before, the latent classes are numbered 1 through $T$, so there are $T$ different latent classes. When both the latent and the manifest variables are measured at a nominal level, the relation between $A$ and $\theta$ is described by a set of $(m_A + 1) \times T$ different two-variable interactions $u_{gt}^{A\theta}$. Because the number of non-redundant parameters equals $m_A \times (T - 1)$, certain restrictions are necessary in order to make this set of parameters identifiable. The analogy between the log-linear model in Equation 1.5 and the usual ANOVA model should be clear. As in ordinary ANOVA, the set of two-variable interactions can be submitted to the following restrictions:

$$\sum_g u_{gt}^{A\theta} = \sum_t u_{gt}^{A\theta} = 0.$$

In the literature these restrictions are known under the heading *effect coding*. However, other restrictions on the set of $u_{gt}^{A\theta}$-parameters are also possible. In regression analysis with dummy-variables, for example, identifiability restrictions are usually imposed by setting the parameters pertaining to one particular category equal to zero. This is called *dummy coding*. In dummy coding, the effect of falling in a specific category is expressed as a deviate from the reference category (i.e., the category for which the parameters are set equal to zero). In effect coding, effects are represented by deviations from the overall mean. In the following, it is assumed that the identifiability restrictions are imposed by dummy coding, so that for the set of two-variable interactions $u_{gt}^{A\theta}$ the following holds:

$$\begin{aligned} u_{g1}^{A\theta} &= 0 \qquad \text{for all } g = 0, \ldots, m_A \\ u_{0t}^{A\theta} &= 0 \qquad \text{for all } t = 1, \ldots, T. \end{aligned}$$

Hence, for item $A$ and latent variable $\theta$, categories 0 and 1, respectively, serve as the reference categories.

Although a more detailed study of the log-linear model is postponed until Chapter 2, some concepts have to be defined here as the linearizing restrictions are based upon these concepts. When the relation between the latent variable $\theta$ and a manifest indicator $A$ is studied and dummy coding is used, a natural question is in what manner the probability of scoring in category $g$ instead of in category 0 (the reference category) of item $A$ depends on the latent class to which an individual belongs. To answer such questions, the log-linear model uses as building blocks the *odds* of answering in category $g$ instead of category 0. The odds are defined as:

$$\text{odds} \; = \frac{p_g^A}{p_0^A}.$$

Because these odds are assumed to depend on the latent class to which an individual belongs, it makes sense to study the *conditional odds*, which are defined as the ratio of two conditional response probabilities:

$$\text{conditional odds} \; = \frac{p_{g\,t}^{\overline{A}\theta}}{p_{0\,t}^{\overline{A}\theta}}.$$

The logarithm of a conditional odds is called the *logit*. If there is association between the variables $\theta$ and $A$, these conditional odds differ for two different latent classes $t$ and $t'$. This can be expressed in the *odds ratio*:

$$\begin{aligned} \text{odds ratio} \quad &= \quad \frac{p_{g\,t}^{\overline{A}\theta}}{p_{0\,t}^{\overline{A}\theta}} \div \frac{p_{g\,t'}^{\overline{A}\theta}}{p_{0\,t'}^{\overline{A}\theta}} \\ &= \quad \frac{p_{g\,t}^{\overline{A}\theta} \times p_{0\,t'}^{\overline{A}\theta}}{p_{0\,t}^{\overline{A}\theta} \times p_{g\,t'}^{\overline{A}\theta}}. \end{aligned}$$

Because it is assumed that the manifest variable is causally influenced by the latent variable, the log-linear model in Equation 1.5 is often replaced by a logit model. Such a model expresses how the logit for a particular category of the manifest item depends on the latent variable. As stated above, the logits are defined by taking the ratio of the probability for responding in category $g$ on the one hand, and the probability for responding in the reference category 0 on the other hand, because a dummy coding scheme is proposed in the present example.[4] As will be shown in Chapter 2, this logit can be expressed as:

$$\ln\left[\frac{p_{gt}^{\overline{A}\theta}}{p_{0t}^{\overline{A}\theta}}\right] = u_g^A + u_{gt}^{A\theta}. \tag{1.6}$$

The logits represent the tendency to answer item $A$ in category $g$ rather than in category 0, and they are assumed to depend on the latent class $t$. Thus, there are $T$ different logits that can be defined for each category $g$ of item $A$, and the variation between these $T$ different logits can be explained by $(T-1)$ $u_{gt}^{A\theta}$-parameters.

More parsimonious models can be obtained by assuming specific contrasts with regard to these $T$ different logits. For example, if it is believed that the latent classes are ordered, then the logits could linearly increase or decrease with the class number $t$. Other sets of polynomial weights (quadratic, cubic, etc.) could also be used. It is clear that the maximum order for these polynomial weights is equal to $(T-1)$, and that when $(T-1)$ different contrast are used, the model gives the same solution as the "full" logit model given in Equation 1.6. The logit model becomes more parsimonious as fewer contrasts are used.

A very important subset of latent class models arises when only the linear contrast is used. The restrictions imposed on the two-variable interactions in this model can be expressed by:

$$u_{gt}^{A\theta} = u^{*A}_{g} \times \theta_t, \tag{1.7}$$

where the values $\theta_t$ are chosen such that they satisfy the requirements of a set of weights for a linear contrast. This restriction is referred to when it is said that the relationship between the latent variable and the manifest indicator is linearized.

When the restrictions as given in Equation 1.7 are imposed on the two-variable interactions, the latent variable $\theta$ is actually treated as a variable on the interval level.[5] The scores $\theta_t$ are equally spaced and

[4]With dichotomous dependent variables, only one relevant logit can be defined. However, for polytomous dependent variables several ways of defining the relevant logits are possible. See, for example, Fienberg (1980) for more details.

[5]From this point on the phrases "metrical variables" and "variables on an interval level" will be used interchangeably. One could object that this is not entirely correct because variables on a ratio level are also metrical. However, a ratio level is not relevant in the context of latent variables because the metric of these variables is always arbitrarily fixed. As regards manifest indicators, a ratio level will in practice only be relevant in a small number of situations.

should be interpreted as metrical values. All latent structure models mentioned in the bottom left-hand cell of Table 1.1 assume that the latent variable is measured on an interval scale. Some of these models (the latent class models with linear restrictions) consider the latent variable as discrete. In the context of log-linear analysis with directly observed variables, these linear restrictions are applied rather frequently (see, for example, Haberman, 1979 and Goodman, 1984). Within the framework of latent class analysis, these linear restrictions have only recently been applied (for instance, by Clogg, 1988). Using the linearizing restrictions as defined in Equation 1.7, the expression for the response probability $p_{gt}^{\overline{A}\theta}$ can now be rewritten as:

$$p_{gt}^{\overline{A}\theta} = \frac{\exp(u^{*A}_{g} \cdot \theta_t + u^{A}_{g})}{\sum_h \exp(u^{*A}_{h} \cdot \theta_t + u^{A}_{h})}. \tag{1.8}$$

At this point, it is not necessary to show how this expression for the conditional response probabilities can be derived. Equation 1.8 is only needed here to illustrate the remarkable resemblance to Bock's Nominal Response model.

While latent class models using linear restrictions on the relation between the latent and the manifest variables treat the latent variable as a *discrete* variable on a metrical scale, it is also possible to link a metrical *continuous* latent variable to manifest indicators on a nominal scale. A well known latent structure model that does this is the *Nominal Response model* proposed by Bock (1972). One can easily get a first intuitive grasp of this model by looking at a picture of the so-called trace lines as is given in Figure 1.2. These trace lines, also called category-characteristic curves, picture the conditional response probabilities as a function of the latent variable $\theta$. It can clearly be seen from Figure 1.2 that in the Nominal Response model for one category, this response probability is monotonically declining, while for one other category the conditional response probability is a monotonically increasing function of $\theta$. The other two categories in Figure 1.2 seem to take an intermediate position as up to a given $\theta$-value, the conditional response probability increases while past this particular $\theta$-value the probabilities consistently decrease. The mathematical expression for the conditional response

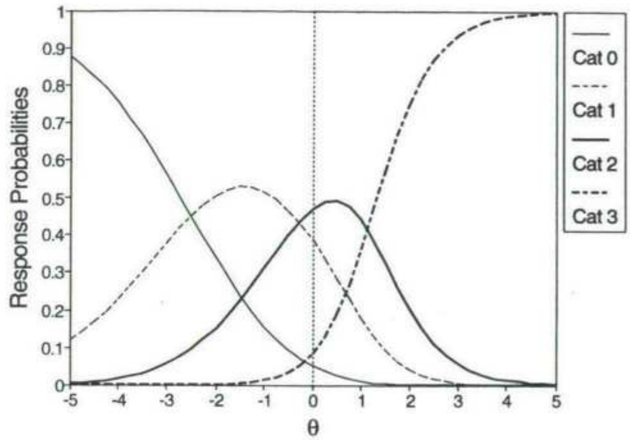


Figure 1.2: Trace lines in the Nominal Response model

probabilities in the Nominal Response model is:

$$p_{jg|\theta_i} = \frac{\exp\left(a_{jg} \cdot \theta_i + c_{jg}\right)}{\sum_{h=0}^{m_j} \exp\left(a_{jh} \cdot \theta_i + c_{jh}\right)}. \tag{1.9}$$

Note that in this equation the notation $\theta_i$ is used instead of $\theta_t$ to stress the continuous character of the latent variable. Furthermore, the number of categories of the manifest variables may vary over items. Therefore, the last numbered category is denoted by $m_j$ instead of just $m$. Finally, the more general notation $p_{jg|\theta_i}$ is used for the conditional response probabilities because a unique identification of certain manifest variables (such as $A$, $B$ and $C$) is no longer needed here.

When Equation 1.9 is compared with the expression for conditional response probabilities in the linear restricted latent class model (see Equation 1.8), the resemblance is quite remarkable, despite the differences in notation. Both models use the same functional form. The only difference seems to be that the Nominal Response model assumes the latent variable to be continuous, while $\theta$ is discrete in the case of the linear restricted latent class model. These differences will be further explored in the following chapters.

In the restricted LCA-models in the bottom left-hand cell of Table 1.1, log-linear parameters are linearized so that only the latent variable is considered metrical. The manifest variable is still thought of as measured on a nominal level. However, it is also possible to do just the reverse. The latent variable remains a set of nominal latent classes, while the log-linear parameters are linearized in a way that assumes that the manifest variables are metrical. This type of latent structure model was proposed by Rost (1985, 1988a and 1988b). Using Rost's original notation, the expression for the response probabilities in Rost's latent class model is:

$$p_{jg|\theta_t} = \frac{\exp\left[g \cdot \lambda_{j|\theta_t} - \sum_{x=0}^{g} \tau_{jx}\right]}{\sum_{h=0}^{m} \exp\left[h \cdot \lambda_{j|\theta_t} - \sum_{x=0}^{h} \tau_{jx}\right]}. \tag{1.10}$$

The LCA model proposed by Rost uses the same linearizing restrictions as the LCA models in the bottom left-hand cell, with the role of latent and manifest variables reversed. This means that if a logit is defined using conditional probabilities for belonging to a given latent class $t$ rather than to another class $t'$, given that the response to item $j$ was in category $g$, this logit will be linear related to the categories $g$ of the manifest item. This model will be studied in some detail in Chapter 2. For now it suffices to take an overall view of the expression of the conditional response probabilities. Firstly, it should be noted that the notation used for the conditional response probability is $p_{jg|\theta_t}$, indicating that this probability is the same for all individuals belonging to the same latent class $t$. Secondly, when the two-variable interaction pertaining to the latent variable $\theta$ and a given manifest indicator $j$ is denoted by $u_{jg\theta_t}$[6], these interactions are restricted by:

$$u_{jg\theta_t} = g \times \lambda_{j|\theta_t}.$$

Each two-variables interaction is assumed to equal the product of a class-specific item weight (i.e., $\lambda_{j|\theta_t}$) and the category number involved (i.e., $g$). The fact that the weights are multiplied by the *category numbers* implies that the manifest variables are seen as variables measured on an interval level, since otherwise there would be no sensible foundation for this restriction. On the other hand, no restrictions are put on the class-specific item parameters $\lambda_{j|\theta_t}$, thus indicating that the different latent classes can be seen as categories of a nominal level variable.

[6]The notation $u_{jg\theta_t}$ is more general than the previously given notation (for example, $u_{gt}^{A\theta}$) and will be used from this point on, except when it is necessary to label the manifest indicators in a more specific manner.

With this latent class model only one class-specific parameter $\lambda_{j|\theta_t}$ will be estimated for each item, while in the unrestricted LCA model, $m$ different independent parameters are estimated per item.

Finally, the one-variable parameters (for which the general notation $u_{jg}$ can be used) are rewritten in the Rost model as a function of a set of $\tau$-parameters. This reparameterization has the following form:

$$u_{jg} = -\sum_{x=0}^{g} \tau_{jx}.$$

The reparameterization is a direct consequence of the fact that Rost developed his LCA model using the so-called *threshold formulation*, originally proposed by Andrich (1978a) in the context of latent trait models for polytomous data. The $\tau$-parameters actually represent points on a latent continuum between two successive categories. In the formulation of Rost's model for ordered data as given in Equation 1.10, the set of threshold parameters is assumed to be the same for all latent classes. This reparameterization of the log-linear one-variable parameters is not very interesting for the present discussion, but this topic will be explored further in Chapter 3. Rost's latent class model itself will be explored in greater detail in Chapter 2. This model does not, however, play a crucial role in this study because it cannot be related directly to latent trait models. And, as said before, the relations between latent class analysis and latent trait models are the major subject of this book.

Until now, three corner cells in Table 1.1 have been discussed. The top left-hand cell contains the ordinary latent class models, linking a latent variable on a nominal level to manifest indicators that are also measured on a nominal scale. This model was discussed briefly in the former section. The bottom left-hand cell and the top right-hand cell treat either the latent variable or the manifest variable as a variable on an interval scale. The last corner cell, the bottom right cell, contains a number of models with latent and manifest variables measured on an interval scale. The latent structure models for this situation are again restricted LCA models if the latent variable is discrete, and certain latent trait models if the latent variable is considered continuous.

In order to obtain a latent class model in which both the latent and the manifest variables are treated as variables on an interval scale, the following restrictions must be imposed on the two-variables interactions in the unrestricted latent class model (see Equation 1.5):

$$
\begin{aligned}
u^{A\theta}_{gt} &= u^{*A} g\theta_t \\
u^{B\theta}_{kt} &= u^{*B} k\theta_t \\
u^{C\theta}_{\ell t} &= u^{*C} \ell\theta_t.
\end{aligned} \tag{1.11}
$$

The log-linear parameters reflecting the association between the latent and the manifest items are written as a weighted product of the latent and manifest scores. The weights may vary depending on the item. The model defined by this set of restrictions is called a linear-by-linear interaction model. It was first applied in the context of log-linear analysis for observed variables by Haberman (1978) and Goodman (1979). The parameters $u^{*A}$, $u^{*B}$ and $u^{*C}$ are often interpreted as coefficients reflecting the degree of association between $\theta$ and the manifest indicators. This interpretation prevails in the tradition of latent class analysis. However, there is yet another interpretation. As will be shown in Chapter 3, the parameters $u^{*A}$, $u^{*B}$ and $u^{*C}$ can also be regarded as scaling parameters, which take into account the fact that the metric of the latent and manifest variables, as reflected by the quantities $g$ and $t$, is arbitrarily chosen. This interpretation is frequently encountered in latent trait analysis.

With the restrictions given by Equations 1.11, the expression of the response probability $p^{\overline{A}\theta}_{gt}$ now becomes:

$$
p^{\overline{A}\theta}_{gt} = \frac{\exp(u^{*A} \cdot g \cdot \theta_t + u^{A}_{g})}{\sum_h \exp(u^{*A} \cdot h \cdot \theta_t + u^{A}_{h})}. \tag{1.12}
$$

This LCA model is more restricted than the model in Equation 1.8 as in Equation 1.12 the two-variable interactions are restricted by:

$$
u^{*A}_{g} = u^{*A} \cdot g.
$$

There are also fewer parameters to estimate in 1.12 than in 1.8. In the unrestricted latent class model (see Equation 1.5), the total number of non-redundant parameters that is used to describe the relation between $\theta$ and $A$ is equal to $m_A \times (T-1)$. When the relation between the latent and the manifest variable is linearized, only one parameter is needed to describe the linear relation between the latent classes and the logit for category $g$ of item $A$ (i.e. the slope parameter). Since $m_A$ different relevant logits can be defined for item $A$ when dummy coding is used, the number of non-redundant parameters in the LCA model given by Equation 1.8 drops to $m_A$. Finally, the model expressed by Equation 1.12 needs only one parameter (i.e., $u^{*A}$) to characterize the relation between $\theta$ and $A$.

There are also a number of latent trait models for polytomous manifest items which assume that both latent and manifest variables are measured on an interval level. One example of such a latent trait model is the *Partial Credit model* proposed by Masters (1982). In this model, the response probabilities are modelled as follows:

$$p_{jg|\theta_i} = \frac{\exp(g \cdot \theta_i + c_{jg})}{\sum_{h=0}^{m} \exp(h \cdot \theta_i + c_{jh})}. \tag{1.13}$$

When the expression for the conditional response probabilities, as given in Equation 1.13, is compared to the one in Equation 1.9 (the Nominal Response model), it becomes clear that the Partial Credit model sets the following restriction on the $a_{jg}$-parameters:

$$a_{jg} = g.$$

One noticeable difference between the LCA model in Equation 1.13 and the Partial Credit model is that the LCA model allows for one separate association parameter for each manifest item (i.e., $u^{*A}$), while the Partial Credit model assumes that the strength of the association between $\theta$ and the manifest variables is the same for all indicators. The Partial Credit model, as well as a number of other interesting latent trait models, will be studied in Chapters 3 and 4.

### 1.3.3 Latent structure models for ordinal data

If only nominal and metrical measurement levels were distinguished, then Table 1.1 would be a complete picture of the known latent structure models. This is, as a matter of fact, what is done in many articles on the analysis of *ordered* data (see, for example, Andrich, 1978a and 1978b; Clogg, 1982; and Goodman, 1984). In the view of many authors, ordered data are simply metrical data, and the way they are dealt with in latent structure models nearly always amounts to the linearization of certain parameters. This, in turn, leads to the assumption that the distances between the values of the latent and/or manifest variables are equal or at least known. When, however, data is interpreted in this manner, quite a bit more is assumed than is allowed for variables that are truly *ordinal*, i.e., variables for which only the order relations among the numerical values assigned to the various categories can be interpreted in a meaningful fashion. There are, however, a number of latent structure models that have been developed for the analysis of ordinal data. These models can all be located in the middle row or middle column of Table 1.1.

The five cells in the middle row and middle column of Table 1.1 will be dealt with consecutively, starting with the middle cell in the first column. In this cell, an ordinal latent variable is linked to manifest indicators measured on a nominal scale. Though at present no models have been discussed that belong to this category, it is possible to indicate how such models could be developed. Keeping in mind that the top left-hand cell assumes that both the latent and the manifest variables are measured on a nominal scale, it is possible to transform this LCA model into a LCA model with latent variables on an interval scale by imposing certain linear restrictions on the log-linear parameters. These restrictions would imply that the relationship between the set of relevant logits for the manifest variable on the one hand, and the $\theta_t$-values for successive latent classes on the other hand, are linear. A model which assumes that the latent variable is measured on an ordinal scale while the manifest indicators are measured on a nominal scale would depart from the assumption that, for each category $g$ of a manifest indicator, the relationship between the set of logits and the successive class numbers is monotonically increasing (or decreasing).

If it is possible to develop models which link an ordinal latent variable to nominal indicators along the lines outlined here, then it must also be possible to formulate models for the middle cell in the top row by reversing the role of the manifest and the latent variable. The relevant logits are now defined in terms of the conditional probabilities for belonging to latent class $t$ rather than latent class $t'$, given that item $j$ is responded to in category $g$. A latent class model that links ordinal manifest variables to a latent variable assumed to be measured on a nominal scale assumes that the relation between these logits and the categories $g$ for the manifest item $j$ is monotonically increasing (or decreasing). The middle cell in the top row is also empty as no such models have yet been developed.

The cell in the center of Table 1.1 combines ordinal latent variables with ordinal manifest variables. A latent class model for these combinations of variables was proposed by Croon (1990). This model is a LCA model using certain inequality restrictions imposed upon the response probabilities. The idea of latent class models for ordered latent classes can be illustrated very easily for the situation with dichotomous manifest indicators. Suppose there are $T$ different latent classes, numbered $1, \ldots, T$, which are ordered according to a given criterion. Assuming a proper ordering of the latent classes, the probability of a "positive" response (indicated by $p_{j1|\theta_t}$) should increase with the class number, thus:

$$p_{j1|\theta_{t+1}} \geq p_{j1|\theta_t}.$$

The same idea can be used to model ordered relations between latent and manifest variables for polytomous indicators. There are, however, a number of ways of defining a "positive response" for polytomous items. Croon (1990) tackled this problem by dichotomizing the polytomous item in a number of different ways. This ordinal latent class model will be studied in more detail in the following chapter.

The last model mentioned in Table 1.1 is the *Graded Response model* suggested by Samejima (1969). This model is a genuine latent trait model, because it assumes a metrical continuous latent variable. As it models the *cumulative* response probabilities and not the original response probabilities (thus, the probabilities for responding in certain response categories, given a specific position on the latent continuum), the Graded Response model deals more explicitly with the ordinal information in the manifest indicators, without assuming an interval metric for these indicators. This model be reviewed only briefly in this chapter. The reason for not discussing this model more extensively is that the Graded Response model cannot be brought within the same general log-linear framework as the other latent trait and latent class models. Because this general log-linear framework makes it possible to link latent trait models to latent class models, the Graded Response model is not very interesting from the perspective central in this study. This model falls in the category of what Thissen and Steinberg (1986) call *difference models.*

The model can be expressed as follows. If the probability of an individual $i$ responding to item $j$ in category $g$ or larger is denoted by $p^*_{jg|\theta_i}$, then this probability can be written as:

$$p^*_{jg|\theta_i} = \Phi(a_{jg} \cdot \theta_i + c_{jg}), \tag{1.14}$$

where the function $\Phi$ must have the properties of a cumulative distribution function. Often, $\Phi$ is the normal ogive or the cumulative logistic distribution. For the latter choice the resulting equation is:

$$p^*_{jg|\theta_i} = \frac{\exp(a_{jg} \cdot \theta_i + c_{jg})}{1 + \exp(a_{jg} \cdot \theta_i + c_{jg})}. \tag{1.15}$$

The term *difference models* is derived from the fact that the usual response probabilities for the intermediate categories can be derived by taking the difference between the complements of two cumulative probabilities. So:

$$p_{jg|\theta_i} = p^*_{jg|\theta_i} - p^*_{jg+1|\theta_i}, \tag{1.16}$$

for $1 \leq g \leq m - 1$, where $m$ denotes the number of the "highest" category of item $j$. For the two extreme categories, the conditional response probabilities can, of course, be found by:

$$\begin{aligned} p_{j0|\theta_i} &= 1 - p^*_{j1|\theta_i} \\ p_{jm|\theta_i} &= p^*_{jm|\theta_i}. \end{aligned}$$

There are some important restrictions that must be imposed upon the parameters $a_{jg}$ and $c_{jg}$. First, the slopes $a_{jg}$ should be equal for each item $j$, i.e. $a_{j0} = \ldots = a_{jm_j}$. The reason for this is quite obvious. If the slopes were not equal, the curves representing the relationship between the probabilities $p^*_{jg|\theta_i}$ on the one hand and the latent variable $\theta$ on the other hand would intersect at a given point. This means that for some values of $\theta$ the difference $p^*_{jg|\theta_i} - p^*_{jg+1|\theta_i}$ will become negative, which is of course not allowed. If the slopes are all equal, the resulting model is denoted as the homogeneous case of the Graded Response model. Second, the intercept-parameters $c_{jg}$ should satisfy the following inequality constraints:

$$c_{j1} \geq c_{j2} \geq \ldots \geq c_{jg} \geq c_{jg+1} \geq \ldots \geq c_{jm_j}.$$

These constraints are necessary in order to guarantee that $p_{jg|\theta_i} \geq 0$.

When the two sets of restrictions are satisfied, the curves describing the relationship between $p^*_{jg|\theta_i}$ and the latent variable $\theta$ run parallel and they are ordered from $p^*_{j1|\theta_i}$ (located at the extreme left) to $p^*_{jm|\theta_i}$ (located at the extreme right). An example for an item with four categories (i.e., $g = 0, \ldots, 3$) is depicted in Figure 1.3.

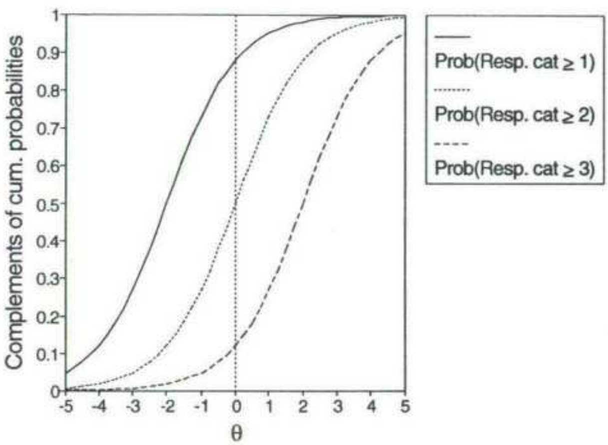


Figure 1.3: Graded Response model: curves describing the complements of the cumulative probabilities

When the proper restrictions are imposed on the slopes $a_{jg}$ and the intercept-parameters $c_{jg}$, the Graded Response model thus accounts for the ordering of the categories of the polytomous indicators. Moreover, as McCullagh and Nelder (1989) noted, models for ordinal data which are based on cumulative probabilities are also advantageous because conclusions based on such a model are not affected by the number of response categories. This means that the combination of certain categories of the manifest indicators will not lead to substantially different conclusions on the relation between the latent and the manifest variables, i.e., the item parameters in this model.[7] In general, this cannot be said of the latent trait models presented in Chapter 3.

The Graded Response model has a number of intuitively appealing characteristics (that is, for the homogeneous case and given that the

[7] This is, however, not true for estimates of $\theta_i$ as was shown by Jansen and Roskam (1986).

inequality restrictions mentioned above have been fulfilled). Firstly, the probability of a response in the highest category (i.e. $p_{jm_j|\theta_i}$) is a monotonically increasing function of the latent variable. Secondly, the probability of a response in the lowest category is a monotonically decreasing function of the latent variable. Finally, the curves describing $p_{jg|\theta_i}$ for all intermediate values of $g$ in function of $\theta$ are unimodal with the maxima ordered according to the ordering of the categories of $j$. An example for an item with four categories is pictured in Figure 1.4.

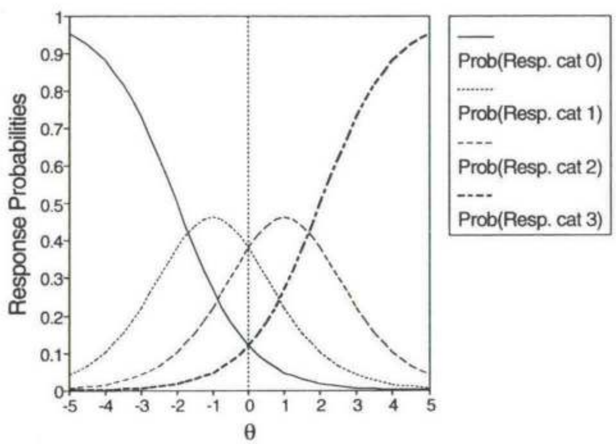


Figure 1.4: Graded Response model: curves describing the response probabilities

Despite all these interesting properties, difference models have not become very popular among applied researchers. This is probably partly due to the fact that some influential statisticians and methodologists have strongly propagated the use of Rasch-like models.[8] The existence of sufficient statistics for both item and person parameters is seen as an important advantage of these types of models. The advantage of the existence of known sufficient statistics for the parameters does not pertain to latent trait models for polytomous data which are based on the Birnbaum model, such as Bock's Nominal Response model. However, these models share with the Rasch-like models the benefit of bearing a very clear resemblance to restricted latent class models, as was noted

[8]This type of models will be dealt with extensively in Chapter 3.

in the previous section. Because it is this relationship between latent trait models and latent class analysis that is of interest in our study, the Graded Response model will not be discussed any further in this study.

A final question that can be raised is whether it is possible to link an ordinal latent variable to metrical indicators (the right-hand cell in the middle row of Table 1.1). In the context of *graphical models*, a number of models have been developed in which (observed) ordinal and metrical variables are related to each other (see, for example, Whittaker, 1990). However, such models do not (presently) exist for ordinal *latent* variables. Therefore, the right-hand cell in the middle row has been left empty.

Having reviewed all the cells in Table 1.1, this introductory chapter will be concluded with some comments on the differences between latent trait models and latent class models and a preview of the chapters to come.

## 1.4 Latent class analysis versus latent trait models

In this chapter a number of different latent structure models were surveyed. The latent structure models for discrete manifest data surveyed in this chapter were either latent class models or latent trait models. The criterion that was used to impose some order on this set of latent structure models was based upon the measurement scale for both the latent and the manifest variables. An overview of this arrangement of latent structure models was given in Table 1.1. It was noted that some cells in this table (i.e., the categories with latent variables on an interval scale) cover both latent class models and latent trait models. This raises the question of what distinguishes these two types of models.

Andersen (1990) mentioned two main points of difference. Firstly, the latent trait model assumes a continuous latent variable, while the latent class model postulates the existence of a number of mutually exclusive latent classes. The latent variable in the latent class model is thus discrete and not even necessarily unidimensional.[9] Secondly, the manner in which the conditional response probabilities are linked

[9]Latent trait models can also be multidimensional, but the latent trait models that are most frequently applied in practice are all unidimensional.

to values or categories of the latent variable is different. In latent trait models explicit functional relations are used to model the dependence of the response probabilities on the latent values; in latent class analysis such explicit functional relations are not used. The most frequently used functional forms are the logistic distribution and sometimes the normal ogive. This difference is seen by Masters (1985) as the most significant difference between the two types of models.

The latter point is of some importance. The unrestricted models estimate the conditional response probabilities $p_{jg|\theta_t}$ in such a way that they give the best fit to the data, whereas the latent trait models propose a functional relationship between these probabilities and the values on the latent continuum. Hence, the latent class model uses many more parameters to fit a certain data set than the latent trait model does. When latent trait models fit the data, they should be preferred since they use much fewer degrees of freedom. For the case in which standard non-restricted latent class models are compared with latent trait models, the difference mentioned by Andersen and Masters is very relevant.

However, in this chapter a number of restricted latent class models were described which also specify an exponential form for the relationship between the response probabilities and the latent classes. Examples are the latent class models with linear restrictions on the relationship between the latent and the manifest variables. Hence, in a number of restricted latent class models, explicit functional relationships are also used to model the response probabilities. This means that the point mentioned by Andersen and Masters is not always relevant. Whenever linearizing restrictions are imposed upon the two-variables interactions reflecting the relations between the latent variable and the manifest indicators, the response probabilities are necessarily functionally related to the latent scores.

So, the only important difference that remains between latent class and latent trait models is the difference with respect to the discrete or continuous character of the latent variable. From a conceptual point of view, this difference may be relevant. Thissen and Mooney (1989) also stressed this point.

It is, however, doubtful whether this difference between continuous and discrete latent traits is important from a more pragmatic point of view. Firstly, there are a number of latent trait models that can be expressed as log-linear models. This means that the estimation of the parameters in the log-linear model will lead to the same results as estimation of the parameters in the (continuous) latent trait model. Secondly, when one tries to estimate the latent scores $\theta_i$ on the basis of the estimated parameters in some latent trait model and the observed response patterns, the maximum number of different $\theta_i$ values that can be estimated is equal to the number of different response patterns that can be observed. For an important number of latent trait models, the maximum number of $\theta_i$ values that can be estimated is much smaller, namely, $(n-1)$ with $n$ equal to the number of manifest indicators. So, the "measurement" of the latent variable is in practice always discrete. These observations raise doubts about the assertion that the difference between continuous and discrete latent traits is always relevant.

Besides these practical considerations, an important question is whether latent class models and latent trait models will yield different results when applied to the same set of data. A number of authors have noticed the similarity between the results of both kinds of analyses when the same set of data is evaluated with the two different models (see, for example, Bock & Aitkin, 1981 and Haertel, 1990). The equivalence of the estimated parameters has been recognized for some time with respect to those latent trait models that can be expressed as certain kinds of log-linear models (Tjur, 1982 and Cressie & Holland, 1983). However, important results can also be obtained for latent trait models that do not fall in this category. So far, only vague comments have been made with respect to the agreement of fit statistics and predicted response probabilities for the two types of models. Haertel (1990) obtained some results regarding this agreement using restricted latent class models for manifest dichotomous data and latent trait models based upon the normal ogive. The reason Haertel did not obtain very rigorous results concerning the identity of results obtained with both types of latent structure models is twofold. Firstly, Haertel used a normal ogive to model the trace lines instead of the cumulative logistic distribution. Secondly, Haertel uses a parameterization for the latent class model in terms of latent proportions and conditional response probabilities instead of log-linear parameters. When the log-linear parameterization is used for the latent class model and the latent trait models are based upon the cumulative logistic distribution, some firm results can be achieved regarding the equivalence of latent class models and latent trait models. The relation between both types of models will be explored in more depth in the following chapters.

## 1.5 An outline of this book

This book is a study of the differences and similarities between latent trait models and latent class models. In this chapter, the latent trait models and latent class models that can be of interest within this perspective were ordered according to classification criteria based upon the measurement level of both the latent and the manifest variables. Chapter 2 starts with a discussion of log-linear models. Next, latent class analysis will be studied in detail. Various ways of parameterizing this model will be explored. Also, the procedures that can be used for the estimation of the parameters in these models will be reviewed, as well as methods for model selection. Chapter 3 concentrates on latent trait models. In this chapter a number of known latent trait models will be studied. Attention is given to the correspondence between latent trait models and certain types of restricted latent class models. This approach leads to new insights for both kinds of latent structure models. Estimation of the parameters in latent trait models is one of the topics that is of great interest with respect to the discussion regarding the correspondence between latent class and latent trait models. Therefore, Chapter 4 will be dedicated entirely to a discussion of the estimation procedures for parameters in latent trait models from this perspective. Finally, Chapter 5 will explore two important extensions of the latent trait model from the very same perspective. One concerns the relation between the latent variable and some other exogenous variables. The other extension concentrates on multidimensional latent trait models.

# Chapter 2

# Latent Class Analysis

## 2.1 Introduction

In this chapter attention is focussed on latent class analysis. The latent class model, which was introduced briefly in Chapter 1, can be parameterized in several ways. One parameterization uses conditional response probabilities, while another parameterization is possible through application of the log-linear model. This second parameterization is instrumental to the present study, because the log-linear formulation makes it possible to link specific latent class models to certain latent trait models. For this reason, log-linear models are dealt with in Section 2.2. Topics that are reviewed are the estimation of parameters in this model (Section 2.2.1) and the problem of testing and selecting models (Section 2.2.2).

Afterwards, latent class analysis (LCA) will also be reviewed in Section 2.3. In this section the different methods of parameterizing the latent class model are discussed and compared. Several procedures that can be employed for obtaining maximum likelihood estimates for the parameters in the latent class model are also reviewed (Section 2.3.1). The problem of testing and selecting models in LCA is comparable to the same kind of problem in the context of log-linear analysis. One complicating aspect of latent class models, however, is the problem of identifiability. This topic is considered in Section 2.3.2.

Whether a specific latent class model leads to a better understanding of the relations between the latent variable and the manifest indicators not only depends on the identifiability of that model and the outcome of formal statistical tests, but also on the strength of the association between the latent variable and the manifest indicators. Measures for the strength of this association are, as shown in Section 2.3.3, to a large extent related to the question of whether individuals can be allocated to certain latent classes on the basis of their observed response patterns.

In Section 2.4, restricted latent class models are studied. Restrictions can be applied to parameters in the different parameterizations of the latent class model. While unrestricted latent class models are not the most suitable class of models when the analysis is aimed at measurement rather than mere data reduction, a number of LCA models that are interesting from this measurement point of view can be obtained by restricting the parameters. Firstly (Section 2.4.1), a number of models are reviewed which have their roots in Guttman's scalogram analysis (Guttman, 1950). These models all make use of restrictions on the conditional response probabilities. Secondly (Section 2.4.2.), a number of latent class models suitable for Likert-type data that were suggested by Clogg (1977, 1981) are reviewed briefly. These models also impose restrictions on the conditional response probabilities. Thirdly, ordinal latent class analysis is examined (Section 2.4.3). This technique uses inequality restrictions on the conditional response probabilities. Finally, a category of restricted latent class models can be formulated by restricting the *log-linear parameters*. In this study, the most important restrictions on the log-linear parameters are restrictions that linearize the relationship between the latent and the manifest variables. Models that use these restrictions are explored in depth in Section 2.4.4. This class of LCA models is of particular interest in this book, because LCA models with linear restrictions can be compared to some known latent trait models. This topic will be explored in more depth in Chapter 3.

The last two sections of this chapter discuss extensions of the latent class model. Firstly (Section 2.5), latent class model with two or more latent variables are examined briefly. Afterwards (Section 2.6), attention is focussed on the relations between latent variables and exogenous manifest variables. Both topics are explored in detail in Chapter 5.

## 2.2 Log-linear models

Although the first discussion of the log-linear model can be traced back to Birch (1963), the popularity of this model is largely due to the work done by Goodman. In a series of articles, which were reprinted in Goodman (1978), the formulation of the model as well as methods for estimating parameters and assessing the fit of the model to the data were discussed. These articles were a major breakthrough in the analysis of discrete data. Until then this type of analysis was based on the classic elaboration procedures which were not well suited to the analysis of more than four variables and did not provide for statistical tests for specific hypotheses about the association between the variables. It is without doubt Goodman's greatest contribution that he formulated a firm statistical basis for the analysis of discrete data. This section on log-linear models is not meant to reconsider all aspects of the log-linear model, but only those topics necessary for understanding the latent class model.

Suppose there are three discrete variables $A$, $B$ and $C$. Arbitrary categories on these three variables will be denoted by $g$, $k$ and $\ell$, respectively. Furthermore, let $p_{gk\ell}^{ABC}$ denote the joint probability for observing an individual who responded to the items $A$, $B$ and $C$ with categories $g$, $k$ and $\ell$, respectively. The general log-linear model states that the natural logarithms of these probabilities can be decomposed as follows:

$$\ln p_{gk\ell}^{ABC} = u + u_g^A + u_k^B + u_\ell^C + u_{gk}^{AB} + u_{g\ell}^{AC} + u_{k\ell}^{BC} + u_{gk\ell}^{ABC}. \tag{2.1}$$

The various terms in the equation at the right can be interpreted as follows. The $u$-term is merely a normalizing constant to ensure that the sum of the probabilities over all possible combinations of $g$, $k$ and $\ell$ is equal to 1. The one-variable terms (i.e., $u_g^A$, $u_k^B$ and $u_\ell^C$) indicate the partial skewness of a variable.[1]

At first sight, there is a strong similarity between the model in Equation 2.1 and the well-known ANOVA model. However, in ANOVA the dependent variable is an individual score, while in the log-linear model the dependent variable is a probability of observing a particular response pattern. As a result, the one-variable parameters (i.e., the "main effects") are not very interesting in the log-linear model, as opposed to analysis of variance, where these effects are considered highly relevant. However, as will be shown in Chapter 3, these one-variable parame-

[1] As Davis (1974) has noted, these one-variable parameters do not indicate the skewness of the variables per se, but they could be interpreted as indices of mean skewness within the combined categories of the other variables.

ters have a clear substantive interpretation in the context of specific restricted latent class models.

The two-variable interactions (i.e., $u^{AB}_{gk}$, $u^{AC}_{g\ell}$ and $u^{BC}_{k\ell}$) can be taken as measures of the strength of the (partial) statistical relationship between two variables. Because the log-linear model assumes that the variables are measured on a nominal level, the partial association between two variables $A$ and $B$ (the "mean" association between $A$ and $B$ over categories of variable $C$) is indicated by a different $u^{AB}_{gk}$ parameter for each different combination of the categories $g$ and $k$. These two-variable interactions are defined in terms of log-odds ratios.

Insofar as the logarithm of the joint probabilities $\ln p^{ABC}_{gk\ell}$ cannot be reproduced by the grand mean $u$, the partial skewness parameters (the one-variable interactions $u^{A}_{g}$, $u^{B}_{k}$ and $u^{C}_{\ell}$) and the partial association parameters (i.e., the two-variable interactions $u^{AB}_{gk}$, $u^{AC}_{g\ell}$ and $u^{BC}_{k\ell}$), a set of three-variable interactions $u^{ABC}_{gk\ell}$ is introduced to guarantee a perfect reproduction. These three-variable interactions denote how the log-odds ratios for two specified variables vary at different levels of the third variable.

It should be noted that the log-linear model in Equation 2.1 is not identified unless certain restrictions are imposed on the parameters. This point was already noted in Chapter 1. The restrictions can be imposed in several ways. First, it can be required that the sum of the $u$-terms over any subscript is equal to zero, i.e. :

$$\sum_g u^{A}_{g} = \sum_k u^{B}_{k} = \sum_\ell u^{C}_{\ell} = 0,$$
$$\sum_g u^{AB}_{gk} = \sum_k u^{AB}_{gk} = \sum_g u^{AC}_{g\ell} = \sum_\ell u^{AC}_{g\ell} = \sum_k u^{BC}_{k\ell} = \sum_\ell u^{BC}_{k\ell} = 0,$$
$$\sum_g u^{ABC}_{gk\ell} = \sum_k u^{ABC}_{gk\ell} = \sum_\ell u^{ABC}_{gk\ell} = 0.$$

This is called effect coding. In the literature on log-linear models, effect coding is the most frequently used method of putting identifiability restrictions on the parameters. Effect coding implies that the interactions measure partial skewness or partial association in terms of deviations from the mean log-probabilities.

A second method of making the model in Equation 2.1 identifiable is to use dummy coding. With this coding scheme, the parameters for one particular category (in the case of one-variable parameters) or one particular combination of categories (in the case of all other parameters) are set equal to zero. In this case, the $u$-terms for the other (combinations of) categories are defined as departures from the fixed $u$-terms. This parameterization is used in Section 2.4.3 of this chapter as well as in Chapter 3.

Both effect coding and dummy coding are arbitrary methods of imposing the necessary identifiability restrictions. Both coding schemes lead to the same estimated expected frequencies and thus to the same model fit indices which are presented later on in this chapter. An excellent introduction to these two coding schemes can be found in Kerlinger and Pedhazur (1973). Long (1984), however, focuses on the differences which arise when interpreting parameter estimates obtained under the two different coding schemes.

It is possible to obtain maximum likelihood estimates for the parameters in model 2.1 using data obtained under a number of possible sampling schemes (see, for instance, Fienberg, 1980). One of the most frequently encountered sampling schemes is derived from the multinomial distribution. The total sample size $N$ is fixed and each unit is assigned to a cell of the cross-classification according to its values on the observed variables. Another possible sampling scheme is derived from the product multinomial distribution. In this case there is a multinomial distribution for each joint category of the variables for which the distributions are fixed in advance.

Both of the coding schemes mentioned earlier (dummy coding and effect coding) make it possible to estimate the values of the non-redundant parameters in the log-linear model using the cross-classified counts $f_{gk\ell}^{ABC}$ observed in a sample. Details about the various methods that can be employed to obtain maximum likelihood estimates for the parameters in a log-linear model will be presented later. The estimated parameters can be used to obtain estimated expected frequencies $\hat{f}_{gk\ell}^{ABC}$

using the same expression as in Equation 2.1:[2]

$$\ln \hat{f}^{ABC}_{gk\ell} = \hat{u} + \hat{u}^{A}_{g} + \hat{u}^{B}_{k} + \hat{u}^{C}_{\ell} + \hat{u}^{AB}_{gk} + \hat{u}^{AC}_{g\ell} + \hat{u}^{BC}_{k\ell} + \hat{u}^{ABC}_{gk\ell}. \qquad (2.2)$$

In the example presented in Equations 2.1 and 2.2, the number of non-redundant parameters in 2.2 is equal to the number of independent observed frequencies.[3] Therefore, this log-linear model gives a perfect fit to the data. So the estimated expected frequencies are equal to the observed counts and there are no degrees of freedom left to test the fit of the model. This log-linear model is called a *saturated model.* For most practical applications this model is not very interesting because this model in fact does nothing more than reparameterize the observed counts in terms of functions of the various $u$-parameters.

More interesting models can be obtained by specifying models with fewer parameters, in other words, by fixing some $u$-parameters equal to zero. In this respect the difference between hierarchical and non-hierarchical models is important. A hierarchical model is an unsaturated model in which, if any $u$-parameter is set equal to zero, all other $u$-parameters of higher order in which the same subscripts appear as in the restricted $u$-parameter are also set equal to zero. For example, the model obtained by setting only the three-variable parameter $u^{ABC}_{gk\ell}$ equal to zero is a hierarchical model. The model in which the two-variable $u$-parameter $u^{AB}_{gk}$ is set equal to zero is only hierarchical when the three-variable $u$-parameter is also set equal to zero. When this last parameter is not zero, the resulting model is non-hierarchical. The difference between hierarchical and non-hierarchical models is important because some of the estimation procedures can only be employed in the case of hierarchical log-linear models. This is also the reason why not every author in this field takes up the study of non-hierarchical log-linear models. The estimation procedures that can be used in the context of log-linear models are presented in the next section.

[2] Equation 2.1 expressed the probabilities $p^{ABC}_{gk\ell}$ as a function of a set of $u$-parameters, whereas Equation 2.2 is in terms of estimated expected frequencies. This difference will be reflected in the value of the overall $u$-parameter. In Equation 2.1 the parameter $u$ ensures that the sum of the probabilities is equal to 1, while in Equation 2.2 $\hat{u}$ merely serves to guarantee that the sum of the estimated expected frequencies is equal to the sample size $N$ under the multinomial sampling scheme.

[3] The number of non-redundant parameters will not always equal the number of independent observed frequencies. Especially in the case of incomplete tables, specific complications can arise. However, we will not deal with these more complicated situations in this chapter.

### 2.2.1 Estimation in the log-linear model

The parameters of log-linear models can be estimated in a number of ways. The most common estimation procedures are methods that yield maximum likelihood estimates. In this chapter only maximum likelihood methods are discussed. The popularity of the maximum likelihood (=ML) procedures can be explained by the attractive statistical properties of the resulting estimators: under relatively mild regularity conditions (see, for example, Bishop et al., 1975), ML estimators are consistent and asymptotically unbiased and efficient. Moreover, ML estimators are asymptotically normally distributed. It is also possible to obtain estimates for the asymptotic variance-covariance matrix of the ML estimators so that interval estimates for the model-parameters are relatively simply obtained and statistical tests on hypotheses with regard to these model parameters can be carried out readily.

There are two numerical procedures that are widely used for obtaining ML estimates in log-linear models, namely, *iterative proportional fitting* and the *Newton-Raphson procedure*. These procedures will be discussed and compared briefly.

Iterative proportional fitting (=IPF) is used mostly in the context of hierarchical log-linear models. IPF takes advantage of the fact that hierarchical log-linear models offer extremely simple expressions for the sufficient statistics which can be used to estimate the parameters in such a model. This can easily be illustrated for a simple model with three variables, $A$, $B$ and $C$. The log-linear model for which the parameters are to be estimated assumes that there are no three-variable interactions, but that all possible two-variable interactions are present. So, the parameters $u_{g\,h\,\ell}^{ABC}$ are all set equal to zero. Because the model is hierarchical, this model can also be denoted by mentioning only the *highest* possible interaction terms or the corresponding marginal tables, i.e., $(AB)$, $(AC)$ and $(BC)$. All the other (*lower*) parameters are implied by the hierarchical nature of this log-linear model. Now it can be shown that the likelihood equations for estimating the parameters simply state that the marginal expected frequencies corresponding to the three interaction terms mentioned should equal the matching marginal observed frequencies, i.e. :

$$\hat{f}_{g\,k\,+}^{ABC} = f_{g\,k\,+}^{ABC},$$

$$\begin{aligned} \hat{f}^{ABC}_{g+\ell} &= f^{ABC}_{g+\ell}, \\ \hat{f}^{ABC}_{+k\ell} &= f^{ABC}_{+k\ell}. \end{aligned} \tag{2.3}$$

The observed marginal two-way tables $(AB)$, $(AC)$ and $(BC)$ are exactly reproduced by the model and the corresponding marginal observed frequencies are the sufficient statistics that can be used to estimate the parameters in the model (Bishop et al., 1975: 64). This result is used in the IPF procedure. In an iterative process the estimated expected frequencies are adjusted to the observed marginals that are to be reproduced until they converge. Details about this procedure can be found in many books on this subject (for instance, Bishop et al, 1975; Fienberg, 1980; Hagenaars, 1990; Andersen, 1990).

The main advantage of IPF is that the procedure is very simple to implement and that in those cases in which direct estimates[4] exist, the iterative procedure takes only two cycles to converge. Although in some situations it takes many iterations before convergence is reached, the total amount of computing time does not have to be excessive since each iteration takes very little time. The main drawback of this method is rather obvious: one has to restrict oneself to the estimation of parameters in hierarchical log-linear models.

Most known programs which use IPF are based upon likelihood equations in terms of complete marginal tables. With these programs it is not easy to estimate parameters in non-hierarchical models, although some non-hierarchical models can be transformed into hierarchical models using particular tricks (see, for instance, Duncan, 1975). Darroch and Ratcliff (1972) have proposed a generalized variant of IPF that can be used to estimate the parameters in non-hierarchical log-linear models. And, as Vermunt (1992) has shown, it is also possible to use IPF on the level of individual cells in the joint $A \times B \times C$ cross-classification, instead of on the level of complete marginal tables. With such an estimation scheme, IPF can easily be employed to estimate the parameters in a number of non-hierarchical models. So, IPF's frequently mentioned drawback of not being able to handle non-hierarchical log-linear models applies more to existing computer programs which use IPF than to the IPF method per se.

[4]Direct estimates are estimates for which closed form expressions exist.

Another frequently reported disadvantage is that IPF does not automatically generate estimates for the variance-covariance matrix of the parameter estimates. Although this is true, it should be realized that this also implies that computations with IPF are relatively simple in contrast to the calculations that must be performed when using the Newton-Raphson procedure. Moreover, nothing in the procedure inhibits the calculation of an estimated variance-covariance matrix, but it takes some extra programming and calculations; indeed, the same kind of calculations that are required by Newton-Raphson. But even these calculations can be circumvented if one is willing to settle for estimates for the standard errors of the parameter estimates, and one does not require estimates for the covariances between the parameter estimates. In this situation, it is very easy to calculate the standard errors for the parameter estimates in a saturated model (see Goodman, 1972b). Because these standard errors can be regarded as the upper bounds for the standard errors of the corresponding parameter estimates in any unsaturated model, in the context of unsaturated log-linear models this procedure will lead to conservative testing procedures which favor the null hypotheses. As the number of parameters in log-linear models tends to grow fast, one could hardly object to these conservative tests. If these inflated standard errors are not acceptable, it is still necessary to compute the variance-covariance matrix of the parameter estimates separately.

In order to describe how ML estimates can be found using the Newton-Raphson procedure, the notation for the log-linear model has to be modified slightly. This facilitates the derivations and also clearly shows how that algorithm works.

In Equation 2.1, the logarithms of the probabilities for observing some response pattern were written as a sum of a number of $u$-terms. The distinction made between one-variable, two-variable and three-variable interaction-terms will now be dropped, and each of these $u$-parameters will be denoted by $u_s$, where $s = 1, \ldots, q$, so that the number of independent $u$-terms, for a particular model without counting the overall $u$-term is equal to $q$. Assuming that the variables $A$, $B$ and $C$ are all dichotomous, there are seven independent $u$-terms and so $q = 7$ for the example in Equation 2.1. Using this new notation: $u_g^A = u_1$, $u_k^B = u_2$ and, finally, $u_{gk\ell}^{ABC} = u_7$. Furthermore, an arbitrary response pattern will again be denoted by the symbol $\nu$. Now the logarithm of the probability of observing $\nu$ can be written as:

$$\ln p_\nu = u + \sum_{s=1}^{q} u_s \cdot x_{\nu s}. \tag{2.4}$$

The $x$-variables are the elements in the design matrix specifying the log-linear model under study. These $x$-variables define the weight for the various $u$-terms for each response pattern. The $x$-variables could be defined such that dummy-coding is applied, or alternatively they can provide an effect-coding scheme. In these two cases, the $x$-variables merely indicate the absence or presence of specific effects. But it is also possible for the $x$-variables to represent polynomial weights. The polynomial weights that are used most frequently are linear weights. Linear weights are, for example, used in a number of latent class models that were introduced briefly in Chapter 1. These models are studied in more detail in Section 2.4.3. Representing log-linear models with the aid of design matrices is clearly very flexible and also allows for the inclusion of non-hierarchical models. For more information regarding the use of design matrices, the reader is referred to Evers & Namboodiri (1978).

Maximum likelihood estimation of the $u_s$-parameters requires finding estimates of those parameters that minimize the log-likelihood function. As shown in Appendix A, this log-likelihood can be expressed as:

$$\ln L = C + \sum_{\nu} f_\nu \ln p_\nu, \tag{2.5}$$

where C is some normalizing constant and $f_\nu$ denotes the observed frequency for response pattern $\nu$. Minimization of this function means that the first derivative of the log-likelihood to the unknown $u_s$-parameters must be set equal to zero. This leads to the following expression (see Appendix A for details):

$$\frac{\partial \ln L}{\partial u_s} = \sum_{\nu} (f_\nu - N \cdot p_\nu)\, x_{\nu s}. \tag{2.6}$$

As there is no closed form solution for this set of equations, the problem may be solved using the Newton-Raphson procedure. The Newton-Raphson procedure for calculating ML estimates solves likelihood equations which do not have explicit solutions by using a Taylor expansion for the first derivatives of these equations up to the first term (see, for example, Rao, 1973). This means that it is necessary to also derive second derivatives of the log-likelihood to the parameters of the model. The second derivatives can be expressed as follows:

$$\frac{\partial^2 \ln L}{\partial u_s \partial u_r} = -N \left[ \sum_{\nu} x_{\nu s} \cdot x_{\nu r} \cdot p_\nu - \left( \sum_{\nu} x_{\nu s} \cdot p_\nu \right) \cdot \left( \sum_{\nu} x_{\nu r} \cdot p_\nu \right) \right]. \tag{2.7}$$

Now the $u_s$-parameters are written in a vector $\boldsymbol{u}$ of order $q$:

$$\boldsymbol{u} = \begin{pmatrix} u_1 \\ \vdots \\ u_q \end{pmatrix}. \tag{2.8}$$

Using the first derivatives, the gradient vector $\mathbf{g}$ is defined, and using the second derivatives, a matrix $\mathbf{H}$ is defined:

$$\begin{aligned} \mathbf{g} &= \frac{\partial \ln L}{\partial \boldsymbol{u}}, \\ \mathbf{H} &= \frac{\partial^2 \ln L}{\partial \boldsymbol{u} \partial \boldsymbol{u}'}. \end{aligned} \tag{2.9}$$

New estimates for the parameters can be obtained by solving iteratively the following equation:

$$\boldsymbol{u}_{(new)} = \boldsymbol{u}_{(old)} - \mathbf{H}^{-1}_{(old)} \cdot \mathbf{g}_{(old)}. \tag{2.10}$$

As has been shown by many authors (see, for example, Haberman, 1978, or Andersen, 1990) this is equivalent to an iterated weighted least squares solution for Equation 2.4. This method was first proposed by Nelder and Wedderburn (1972).

Although the Newton-Raphson procedure is computationally more complex than IPF, Newton-Raphson is a very flexible tool for obtaining ML estimates in both hierarchical and non-hierarchical models. The choice of the $x_{\nu s}$-variables in 2.4 makes it possible to estimate parameters in a great number of different log-linear models.[5] The Newton-Raphson algorithm, when employed in the context of log-linear models,

[5]The choice of the $x$-variables is, however, not completely free because the design matrix should be of full rank.

is also relatively insensitive to the specific choice of a set of initial estimates for the parameters, just as IPF is. When Newton-Raphson is used for estimating parameters in latent class models, however, the algorithm is much more sensitive to the choice of the initial estimates.

Computing time may be saved by calculating the inverse of the matrix **H** in only the first two or three steps and working in the remainder of the iterations with this inverse matrix. This has the disadvantage that at the end of the computations one does not have a correct matrix of second derivatives. This is an important point because the inverse of this matrix can be shown to be the estimate of the asymptotic variance-covariance matrix of the parameter estimates.[6] In the case of unsaturated log-linear models, the standard errors of the parameter estimates are not inflated, as was the case with IPF. Another time saving possibility is to calculate only the diagonal elements of the matrix **H**. Because the inverse of a diagonal matrix is calculated very easily, this saves lots of time. However, with this modification of the algorithm a multivariate Newton-Raphson procedure is broken down to a series of univariate Newton-Raphson procedures, which in some situations can have a negative influence on the stability of the algorithm.[7] Moreover, the advantage of an automatic calculation of the estimated variance-covariance matrix of the parameter estimates is then also given up.

The final practical problem concerning the use of Newton-Raphson is that with most computer programs that use Newton-Raphson, it is necessary to supply a design matrix with the $x_{\nu s}$-variables. In some instances this can be a tedious job. Some programs offer more or less automated facilities to generate design matrices.

### 2.2.2 Testing and assessing model fit

After estimates of the model parameters are obtained with one of the methods described in the previous section, the question is whether the

[6]In the case of log-linear models with direct observations (i.e., no latent variables) the variance-covariance matrix of the parameter estimates can be estimated by either the inverse of the matrix of second derivatives or by the inverse of the information matrix. Both procedures will give the same numerical results (see Haberman, 1977a and 1977b).

[7]Replacing the multivariate Newton-Raphson procedure by a series of univariate Newton-Raphson procedures is for a number of models in fact identical to the modified IPF strategy proposed by Vermunt. See Vermunt (1992) for details.

model fits the data in an acceptable way. There are several possible ways to answer this question. First, one could try to solve the problem by carrying out an unconditional statistical test. One of the most important tests that could be applied is an overall goodness of fit test. A specific log-linear model is specified for the population. The null hypothesis states that the sample is drawn from a population for which the specified model holds. This null hypothesis is tested against the alternative that the observed response patterns stem from a general multinomial distribution. In other words, the alternative hypothesis specifies a saturated model so that under this hypothesis, the observed frequencies are identical to the frequencies that would have been expected when this alternative hypothesis is true. If the likelihood of the data under the optimal solution for the specified model is denoted by $L_o$, and the likelihood under the general multinomial distribution by $L_g$, then the test statistic (the log-likelihood ratio $G^2$) is:

$$\begin{aligned} G^2 &= -2 \times \ln\left(\frac{L_o}{L_g}\right) \\ &= -2 \times \sum_{\nu} f_\nu \cdot \ln\left(\frac{N \cdot p_\nu}{f_\nu}\right). \end{aligned} \tag{2.11}$$

Provided that some regularity conditions are satisfied (see Bishop et al. 1975, p. 510 and pp. 513-518), it can be proven that the log-likelihood ratio test statistic has an asymptotic $\chi^2$ distribution with degrees of freedom equal to the number of independent observed counts minus the number of independent parameters that have to be estimated according to the model under study. Another test statistic that could be used to test the same null hypothesis is the well-known Pearson test statistic, which is defined as:

$$\chi_P^2 = \sum_{\nu} \frac{(f_\nu - N \cdot p_\nu)^2}{N \cdot p_\nu}. \tag{2.12}$$

This test statistic also approximately follows a $\chi^2$ distribution with the same number of degrees of freedom as the log-likelihood ratio. Because both test statistics have the same asymptotic distribution, the values of these both statistics should be nearly the same. However, in practice one may observe that the values of both statistics are quite different.

This is specially true when the observed counts are relatively small. The problem also occurs in the analysis of sparse tables, i.e., tables in which some cells have observed frequencies equal to zero. These observations have led to a discussion concerning the question of how well the two test statistics approximate the theoretical $\chi^2$ distribution in the problematic situations of small frequencies and sparse tables if the null hypothesis is true.

In the case of small observed counts, some Monte Carlo studies have indicated that in certain situations the Pearson statistic seems to behave well in this respect, while the log-likelihood ratio systematically yields values that are too high and thus underestimates the $p$-values, i.e., the probability of a type I error (Fienberg, 1980; Read and Cressie, 1988). Very little is known about how both test statistics behave in the analysis of very large sparse tables. One of the specific problems encountered in the analysis of these kinds of tables is that some of the marginal cells which are necessary for obtaining estimates of the parameters are empty. In this case, one frequently adjusts the number of degrees of freedom (Bishop et al. 1975, pp. 114-116). However, the way these adjustments are made seems to be based on intuitive arguments as a formal foundation is lacking (Kreiner, 1987).

All these objections against the use of the two test statistics have raised the question whether it is possible to replace the asymptotic tests by exact tests. However, the numerical difficulties which arise with exact tests have been a serious drawback. Recently, there have been some attempts to tackle these problems (Kreiner, 1987). Unfortunately, these efforts have only been made in the context of so-called graphical models (Whittaker, 1990), which are a subset of specific log-linear models. Hence, a general solution for this problem is still lacking. A recent survey of exact tests for contingency tables was given by Agresti (1992).

Besides performing an unconditional overall test on the goodness of fit, it is also possible to test more specific hypotheses by using conditional tests. Conditional tests make use of the fact that the log-likelihood ratio can be partitioned into components which can be attributed directly to specific effects, i.e., $u$-parameters. Therefore, precise hypotheses can be tested concerning the need to retain specific $u$-parameters in the model. For that purpose two models are specified. The first model contains certain parameters (for instance, all main effects) and, in addition, the particular parameters in which one is interested (for example, certain association parameters). The second model is nested in the first model, because it uses only a subset of the parameters of the first model (for instance, only the main effects). Then a conditional test can be carried out with the difference in $G^2$ values for the two models as the test statistic. This statistic also approximately follows a $\chi^2$ distribution with degrees of freedom equal to the difference in degrees of freedom for the two models. The null hypothesis tested states that all $u$-parameters that are present in the extended model, but are missing in the restricted model, are zero in the population. Of course, these conditional tests can also be used to test hypotheses concerning one specific parameter instead of a subset of several parameters. Conditional tests, however, can only be carried out if the extended model is true. Readers familiar with regression analysis will note the similarity to testing hypotheses pertaining to differences in the proportion variance explained by different nested models.

When the hypothesis formulated concerns just one specific parameter, an alternative for the conditional testing procedure is performing a test using the standard error for the parameter. Because these standard errors can be estimated and the ML estimates of the parameters are asymptotically normally distributed, a simple test for the significance of each $u$-parameter can be carried out which is comparable to the $t$ test for regression parameters in regression analysis.

When the hypothesis of interest pertains to a subset of $u$-parameters rather than just one single parameter, the conditional testing procedures should be preferred because the inflation of the type I error-rates can be reduced by using conditional tests. In the case of unconditional tests, this can only be achieved by making use of special simultaneous testing procedures (Goodman, 1964) or by adapting the significance level according to a Bonferroni procedure (see, for example, Andersen, 1990). Bonferroni procedures are, however, very conservative and therefore not very powerful. Holm (1979) has developed a sequential Bonferroni procedure which has more power than the original procedure. Simultaneous testing of several restrictions can, however, best be done using conditional tests.

Both testing procedures which use the estimated standard errors for the parameters, and the conditional tests that are based upon differences in the log-likelihood ratios for nested models, can be employed in step-wise procedures to search for the "best" model in an exploratory way. Both "backward" procedures (which start from the saturated model) and "forward" procedures (that use some baseline model as a starting point) have been proposed in the literature. Detailed guidelines for these exploratory stepwise procedures can be found in Goodman (1971, 1973a) and Bishop et al. (1975, pp. 165-175).

There are two reasons to prefer the conditional testing procedures to the unconditional tests. First, conditional tests are relatively robust. Even in the case of very sparse tables, conditional test statistics seem to behave very well in the sense that the approximation of the actual sampling distribution of the test statistic by the theoretical $\chi^2$ distribution is still satisfactory. Second, conditional tests have more power than the corresponding unconditional tests (Goodman, 1981) because the alternative hypothesis in a conditional test specifies a more restricted model than the alterative hypothesis for an unconditional test (where the specified model for the alternative hypothesis is always the unrestricted saturated model).

Another guideline in the exploratory log-linear analysis is given by inspecting the residuals, i.e., the differences between the expected and the observed frequencies. In most cases these residuals are standardized by dividing them by the square root of the expected frequencies:

$$\text{Standardized residual } = \frac{f_\nu - N \cdot p_\nu}{\sqrt{N \cdot p_\nu}}. \tag{2.13}$$

These standardized residuals are the building blocks for the Pearson statistic. Another frequently developed method is to calculate adjusted residuals, i.e., residuals divided by there respective standard errors. Formulas for these adjusted residuals can be found in Haberman (1978; 78-79). These adjusted residuals are asymptotically standard normally distributed.

Assessment of model fit has until now been discussed in the context of statistical tests which are, with a few exceptions, asymptotic tests. This leads to some serious problems because, as stated before, with relatively small sample sizes, the expected frequencies will also be small and the approximation of the sampling distribution by the theoretical $\chi^2$ distribution will often not be very satisfactory. On the other hand, when the sample size is large, almost any difference between the observed and the expected counts will lead to rejection of the null hypothesis. In addition, when stepwise procedures are used, the number of tests that will be carried out can increase very rapidly. This, of course will inflate the actual error rates. These problems have led some researchers to use the test statistics in a purely descriptive sense. However, the log-likelihood ratio and the Pearson statistic are not very good descriptive indices since the values that these two quantities take depend heavily on the sample size. This effect can easily be eliminated by dividing these measures by that sample size, which results in the so-called effect sizes:

$$e_1 = \frac{G^2}{N} \quad \text{and} \quad e_2 = \frac{\chi_P^2}{N}. \tag{2.14}$$

It should be noted that these effect sizes not only have a descriptive meaning but are also used in determining the power of both conditional and unconditional $\chi^2$ tests (Bonnet and Bentler, 1983; Cohen, 1977).

Another objection that can be made against the use of the log-likelihood ratio and the Pearson statistic is that the values of these statistics depend on the number of parameters in a model. Both statistics will take smaller values as the number of parameters in a model grows. So parsimonious models will nearly always be in an unfavorable position when compared to less parsimonious models. Moreover, there is no clear criterion that can be used to decide when a specific model has to be rejected or not. These problems can be resolved by dividing the statistics through the corresponding number of degrees of freedom:

$$F_1 = \frac{G^2}{df} \quad \text{and} \quad F_2 = \frac{\chi_P^2}{df}. \tag{2.15}$$

Both $F_1$ and $F_2$ are asymptotically distributed as $F$ variables with degrees of freedom $df_1 = df$ and $df_2 = \infty$ (Goodman, 1971 and 1975; Haberman, 1978; Wheaton, 1987). When the null hypothesis is true, the expected value of the $F$ statistics is equal to one, so that models with large $F$ values are not likely to be accepted.

Another descriptive use that can be made of the $\chi^2$ statistics is in terms of the improvement of the fit of a model by adding one or more parameters to the log-linear model. Goodman (1972a, 1972b) has developed measures that were originally interpreted as having some analogy to partial and multiple correlations. The general expression in terms of the log-likelihood ratio for these measures is as follows:

$$\frac{G_r^2 - G_u^2}{G_r^2}. \tag{2.16}$$

where $G_r^2$ and $G_u^2$ are defined as before, i.e., $G_r^2$ is the log-likelihood ratio for a restricted model and $G_u^2$ is the same statistic for an unrestricted model. These measures always take values in the interval between 0 and 1. Because both $G_r^2$ and $G_u^2$ are based on the same sample size $N$, the value of this index is not affected by changes in the sample size. This may suggest that the value of these coefficients is not influenced by any sample size effects, but this is not true, as was proven by Bollen (1990). He showed that there is a relation between the mean of the sampling distribution of this statistic and sample size. In other words, for the valid population model different sample sizes will lead to other expected values for the statistic presented above.

When the restricted and the unrestricted models differ from each other in just one parameter, this measure is called the coefficient of partial determination. High values of this measure indicate that the parameter that is missing in the restricted model is responsible for an important improvement in the fit of the unrestricted model, whereas low values demonstrate that it can be left out in the restricted model. When the restricted and the unrestricted models differ with respect to several parameters, the measure is called the coefficient of multiple determination; the coefficient then indicates whether including this set of parameters in the restricted model leads to a substantial improvement of the fit of this unrestricted model.

In the context of log-linear modelling the use of these measures has become obsolete in the past few years because the values they take on depend on the chosen baseline model. The question whether baseline models should contain prior information has been extensively studied by Sobel and Bohrnstedt (1975). In some applications the coefficient of multiple determination has been calculated with an equi-probability model as the restricted baseline model. In other examples, the baseline model is defined as the model in which all two-variable parameters and all higher-order interactions are set equal to zero. With such restricted baseline models, the coefficient of multiple determination will be approximately one for almost any model that is less restrictive. As such, these measures are not very informative. Further, it should be noted that these measures reflect differences in the fit of two nested models, but sometimes they have been incorrectly interpreted as measures of explained variation. As a number of authors have pointed out (see, for example, Hagenaars, 1990), the analogy with multiple and partial correlations is rather weak. High values of these fit indices do not necessarily indicate a strong association between the variables.

The coefficient of multiple determination has also been widely used in the context of structural equation modelling. Bentler and Bonnet (1980) introduced the name *normed fit index* for the very same coefficient. They also presented some modifications of this index, for example, the *non-normed fit index*:

$$\frac{G_r^2/df_r - G_u^2/df_u}{G_r^2/df_r - 1}. \tag{2.17}$$

The adjustment for the degrees of freedom is made to insure that the index can obtain a value equal to 1.0. This improvement is relevant because it can be proven that the original normed fit index does not reach 1.0 even when the unrestricted model is the correct model. This is especially true for small samples. The non-normed fit index does however have the disadvantage that it can fall outside the range of permitted values, i.e., $0 - 1$. Moreover, the non-normed fit index has a larger standard error than the normed fit index. In an attempt to solve these problems, Bentler (1990) developed another modification of these indices by assuming that the test statistics for both the restricted and the unrestricted model approximate a noncentral $\chi^2$ distribution in large samples. This modified index is then defined in terms of the estimated noncentrality parameters of the two models being compared. This index, which has been named the *comparative fit index*, has the advantage of a small standard error and its value can range from 0 to 1. McDonald and Marsh (1990) also prefer an index which makes use of the noncentrality parameters. Still other fit indices can be found in Bollen (1989).

The next approach for model selection that will be discussed here was proposed by Akaike (1974, 1987). Akaike discussed the problems of both estimation and model selection from an information theoretical viewpoint. From a set of models, that model will be chosen that contains the greatest amount of information. In such a selection procedure, the goodness-of-fit of a particular model is weighted against the parsimony of that model. That can be done by calculating the Akaike's information criterion (AIC) for each of the models.

$$\text{AIC} = G^2 - 2 \times df. \tag{2.18}$$

The model that will be preferred is the one with the lowest AIC value. The correction for degrees of freedom can be seen as a penalty for choosing models with too many parameters. Bozdogan (1987) proposed a modification of AIC in order to obtain a more stringent penalty for overfitting and to get an index which is asymptotically consistent (as contrasted with AIC, which is not asymptotically consistent). This *consistent Akaike information criterion* (CAIC) is defined as:

$$\text{CAIC} = G^2 - (\ln N + 1) \times df. \tag{2.19}$$

Another variation of AIC was proposed by Raftery (1986a, 1986b) and is also discussed by Schwarz (1978) and Sclove (1987). This index, the *Baysian information criterion* (BIC), can be computed by:

$$\text{BIC} = G^2 - \ln N \times df. \tag{2.20}$$

The BIC statistic may be interpreted within a Baysian context. It is directly related to the posterior odds for the model specified under the null hypothesis against the model under the alternative hypothesis. Like CAIC, BIC favors models with fewer parameters than does AIC, except for very small samples. One of the benefits of the information criteria is that they permit the comparison of models which are not nested. As such, they can be used in a much broader context than the fit indices presented earlier. A typical characteristic of this approach is that one questions the idea of a "true" model. There is a set of models which are in principle acceptable and from this set that model is chosen that contains the most information. This strategy of not searching for the "true" model, but instead selecting one or more models from a subset of relevant models that provide an acceptable approximation to reality is advocated by more authors (see, for example, McCullagh & Nelder, 1989, and Cudeck & Browne, 1983).

The main objection to the use of these information criteria is the fact that model selection will depend on the sample size (McDonald and Marsh, 1990). The model that yields the smallest AIC value will be more complex for larger sample sizes. At some given sample size the saturated model will necessarily be selected. The adjustments made by CAIC and BIC just seem to slow down the rate at which the saturated model will be selected.

The conclusion then should be that there is still no reasonable method available for model selection in exploratory analyses. Neither the information criteria nor the fit indices presented earlier seems to be satisfactory. Almost all indices suffer from sample size effects, while the fit indices also raise the problem of selecting a useful baseline model. As such, they are not necessarily better alternatives than the classical asymptotic testing procedures. The approach based upon exact tests proposed by Kreiner (1987) may be very promising, but needs to be worked out for other than graphical models.

All methods reviewed in this section can help in selecting models to understand the structure of the observed data. It goes without saying that this understanding should always be based upon some theoretical notions about the relations between the variables. An inspection of the magnitude and sign of the estimated parameters will give a first important impression of the correctness of these global notions. Statistical tests, descriptive statistics and information criteria can only help to gain a more firmly based idea about the theoretical hypotheses.

In this and subsequent chapters a number of illustrative analyses will be presented. Because Chapters 3 and 4 focus upon a comparison of different latent trait models and of various ML estimation procedures, the values of the likelihood ratios that are found in these examples are always interesting for this reason. Moreover, some of the latent trait models studied in Chapters 3 and 4 are compared with certain latent class models reviewed in this chapter. Hence, in this chapter the value of $G^2$ will also be reported for each analysis. The first impression concerning the goodness-of-fit of all these models will therefore also be based upon conditional and unconditional log-likelihood ratio tests. To better understand these models in the context of the examples, the estimated parameters will also be reported in many cases.

## 2.3 Latent class models

The latent class model was developed by Lazarsfeld (1950a, 1950b) as a measurement model for categorical data.[8] Latent class analysis may be seen as a qualitative version of the factor analytic models for the analysis of quantitative data. However, the use of latent class models was severely hampered by the lack of good estimation procedures for the parameters of these models. The work of Lazarsfeld and Henry (1968) and of Wiggins (1955, 1973) did not solve these problems. Goodman (1974a; 1974b) and Haberman (1974a, 1976, 1979) were the first to formulate maximum likelihood estimation procedures which could be used in the field of latent class analysis. In the following section some of these ML methods are considered. First, however, the basic assumptions of latent class analysis as well as the basic equations for the model will be presented. The core of these matters was presented in Chapter 1, but a brief introduction will probably elucidate the kind of restrictions that can be imposed on the parameters in order to develop some interesting latent class measurement models.

A central assumption in latent class analysis is that the population consists of a set of mutually exclusive and exhaustive homogeneous subpopulations. Together these subpopulations make up a latent classification which is by definition discrete. The notation used here is the same as that used in Chapter 1. The latent variable will be denoted by the Greek character $\theta$. The number of latent classes is equal to $T$ and the probability that a subject drawn at random from the population belongs to subpopulation (=latent class) $\theta_t$ is written as $p_{\theta_t}$. Because the latent classification is exclusive and exhaustive, it is obvious that $\sum_t p_{\theta_t} = 1$. The subpopulations are homogeneous in the sense that the probability for giving some particular response on a particular item given that one belongs to latent class $\theta_t$ is the same for all individuals belonging to that specific latent class.

[8]Although latent class analysis was developed as a measurement model and the accent in this chapter is on measurement models, the use of latent class analysis has reached a much more broader scope. Clogg (1981) and Hagenaars (1990) have given many examples of the use of latent class models in the analysis of causal relations.

The second central assumption is the assumption of local independence. This assumption states that the association between the manifest variables can be explained by their dependence upon the latent classification. Consequently, within a given latent subpopulation, all manifest indicators will be statistically independent.

Assume that subjects are measured on three variables $A$, $B$ and $C$, and let arbitrary categories for these three variables be denoted by $g$, $k$ and $\ell$, respectively. The conditional probabilities are written by $p_{gt}^{\overline{A}\theta}$, $p_{kt}^{\overline{B}\theta}$ and $p_{\ell t}^{\overline{C}\theta}$. Using the assumption of local independence, the probability for observing response pattern $\nu$, given that a subject belongs to latent class $\theta_t$, can be written as:

$$p_{\nu|\theta_t} = p_{gt}^{\overline{A}\theta} \cdot p_{kt}^{\overline{B}\theta} \cdot p_{\ell t}^{\overline{C}\theta}. \tag{2.21}$$

The probability that a randomly chosen subject will come up with response pattern $\nu$ is given by:

$$p_\nu = \sum_{t=1}^{T} p_{\nu|\theta_t} \cdot p_{\theta_t}. \tag{2.22}$$

Only two sorts of intuitive parameters appear in the preceding expression; the latent proportions $p_{\theta_t}$ and the conditional response probabilities $p_{gt}^{\overline{A}\theta}$, etc.

This simple unrestricted latent class model will be illustrated using data from a Dutch study of socio-cultural developments in the Netherlands. Information about this project can be found in Felling et al. (1987). The data were gathered in 1985 using a sample of the Dutch adult population. For this analysis, the data of a subsample consisting of respondents with at least some years of secondary education were used.[9] The analysis in this section uses the answers to five items which formulate different opinions about male and female role patterns. The

[9]To be more specific, all respondents who had only primary education or who had only attended a lower vocational school were excluded. The reason for this is that analyses for the total sample yielded rather unsatisfactory results for the latent trait models presented in Chapter 3. Because it is very frustrating to present a set of analyses, in which all null hypotheses must be rejected, that part of the sample that seemed to be responsible for the bad fit was excluded.

items were stated as follows:

1. Women's liberation sets women against men.
2. It's better for a wife not to have a job because that always poses problems in the household, especially if there are children.
3. The most natural situation occurs when the man is the breadwinner and the woman runs the household and takes care of the children.
4. It isn't really as important for a girl to get a good education as it is for a boy.
5. A woman is better suited to raise small children than a man.

The responses on all of these items were rated on a five-point scale. For the illustrative analyses in this section, a dichotomy is used in which the two categories have the following meaning:

- category 0 : agree entirely, agree or neutral (= don't agree, don't disagree)
- category 1 : don't agree or don't agree at all

The percentage of respondents who did not agree (at all) with the items is presented in Table 2.1. These data are given for both the male and the female respondents, who satisfied the educational condition mentioned above. Also given are the data for the male and female respondents together.[10] The reason for presenting the data for the male and the female respondents separately is that in some later subsection, a comparative analysis for the male and the female samples will be presented. Looking at the data for the male respondents, it is obvious that item 5 is the most difficult, in the sense that the smallest percentage of men takes a pro-women's lib position on this item. The next most difficult item is item 3. Item 4 however is an easy item; nearly 9 out of every 10 men do not agree with this item. The order of difficulty

[10] All respondents who had a missing value on at least one of the items were excluded in order not to complicate the analyses. However, if certain assumptions are met, it is possible to perform latent class analysis with missing data. See Hagenaars (1990, pp. 249-263) for more details.

Table 2.1: Percentage of category 1 responses

| | Male | Female | Total |
|---|---|---|---|
| Item 1 | 50.4 | 59.8 | 55.3 |
| Item 2 | 71.6 | 77.7 | 74.8 |
| Item 3 | 46.7 | 51.7 | 49.3 |
| Item 4 | 89.1 | 93.6 | 91.4 |
| Item 5 | 33.9 | 57.1 | 46.0 |
| $N$ | 542 | 592 | 1134 |

is somewhat different for the female respondents. Here item 3 is the most difficult one, followed by item 5. The remainder of the order in difficulty is the same for male and female respondents. Further, it is very clear that women are in general more pro-women's lib than men are.

In order to perform a latent class analysis it is necessary to know with what frequency each of the $2^5 = 32$ response patterns occurs. This information is given in Table 2.2, both for male and for female respondents who satisfy the educational conditions mentioned earlier.

Table 2.2: Response patterns with frequencies

| Item 12345 | Male | Female | Item 12345 | Male | Female |
|---|---|---|---|---|---|
| 00000 | 30 | 23 | 10000 | 5 | 4 |
| 00001 | 1 | 0 | 10001 | 2 | 1 |
| 00010 | 48 | 49 | 10010 | 28 | 13 |
| 00011 | 7 | 14 | 10011 | 3 | 10 |
| 00100 | 0 | 0 | 10100 | 1 | 0 |
| 00101 | 0 | 0 | 10101 | 0 | 1 |
| 00110 | 10 | 4 | 10110 | 7 | 7 |
| 00111 | 1 | 3 | 10111 | 11 | 3 |
| 01000 | 8 | 3 | 11000 | 3 | 2 |
| 01001 | 1 | 0 | 11001 | 0 | 1 |
| 01010 | 68 | 45 | 11010 | 51 | 45 |
| 01011 | 19 | 29 | 11011 | 15 | 47 |
| 01100 | 2 | 2 | 11100 | 2 | 0 |
| 01101 | 2 | 0 | 11101 | 2 | 1 |
| 01110 | 36 | 15 | 11110 | 59 | 42 |
| 01111 | 36 | 51 | 11111 | 84 | 177 |

First, the results of the latent class analyses for the entire subsample (i.e., for male and female respondents together) is presented. A very intuitive starting point is given by the assumption that the association between the five items can be explained by a twofold latent variable; one latent class consisting of people who take a pro-women's lib point of view and one latent class for those who have more traditional positions concerning male and female role patterns.

The subject of testing and model selection in latent class models is discussed in Section 2.3.2. With these models, the same hypothesis can be tested that can be tested with log-linear models. This null hypothesis states that the sample is drawn from a population for which the specified model is valid. Again, two test statistics can be used whose distribution approximates a theoretical $\chi^2$ distribution, namely, the Pearson's statistic and the log-likelihood ratio $G^2$. The values found for these two test statistics in the present example, the number of degrees of freedom and the corresponding probability levels for the present example are given in Table 2.3.[11]

Table 2.3: Testing results for unrestricted latent class analyses with 2 and 3 latent classes

| | | $T = 2$ | | $T = 3$ | |
|---|---|---|---|---|---|
| | | | $p$ | | $p$ |
| Male | Pearson $\chi^2$ | 40.66 | .004 | 14.68 | .474 |
| | $G^2$ | 42.20 | .003 | 15.14 | .442 |
| | $df$ | 20 | | 15 | |
| Female | Pearson $\chi^2$ | 54.30 | .000 | 19.31 | .311 |
| | $G^2$ | 53.78 | .000 | 17.53 | .419 |
| | $df$ | 20 | | 17 | |
| Total | Pearson $\chi^2$ | 66.82 | .000 | 14.26 | .506 |
| | $G^2$ | 67.90 | .000 | 13.37 | .574 |
| | $df$ | 20 | | 15 | |

The latent class model with two latent classes has to be rejected. The same is true for the analyses carried out separately on the male and female subsamples. Obviously the assumption that the population can be divided in two homogeneous subpopulations which explain the association between the five indicators is not correct. In other words, the

[11] The computations were carried out with the program LCAG. See Hagenaars & Luijkx (1987) for more details.

hypothesis that a dichotomous latent variable can explain the observed association between the manifest indicators has to be rejected.

Next, an unrestricted latent class model with three latent classes is considered. One justification for the assumption that there are three latent classes could be that the items differ with respect to difficulty. Then it is possible that there is a latent class of people who take a medium position in the sense that they agree with the easier items (i.e., the items 1, 2 and 4), but do not agree with the more difficult items (i.e., item 3 and item 5). This model does not have to be rejected (see Table 2.3). The number of degrees of freedom for this model is 14. However, one estimated conditional response probability had a boundary value of 0.0 (see Table 2.4). If this parameter had been restricted a priori to that value, the number of degrees of freedom would have been 15. Of course, it is not entirely correct to impose this restriction after the parameters have been estimated, but it is a practical way out of the problem encountered here, and it is also a widely adopted practice. Therefore, that particular conditional response probability has been set equal to zero, although it is acknowledged that this restriction is somewhat problematic.

Since there is some difference between men and women with respect to the order of difficulty for the five items, the unrestricted model with three latent classes was also tested for male and female respondents separately. The results for both the male and the female subsamples are given in Table 2.3.[12]

How the three latent classes should be characterized depends on the pattern of conditional response probabilities. For male and female respondents separately, and for the total subsample the values of the estimated probabilities of answering in category 1 (i.e., a pro-women's lib point of view) are given in Table 2.4 (these conditional probabilities are symbolized by $p_{jg|\theta_t}$; this notation will be introduced in Equation 2.26). Also reported are the estimated latent proportions $\hat{p}_{\theta_t}$ and the log-linear two-variable parameters for the relations between the latent and the manifest variables, indicated by $u_{jg\theta_t}$. These log-linear parameters will be interpreted substantively in Section 2.3.3.

The pattern of conditional probabilities shows that there is one class

[12] For the men there was one estimate with a boundary value of 1.0, but for the women there were three such estimates. Hence, the difference in the number of degrees of freedom.

Table 2.4: Latent proportions, conditional probabilities for scoring in category 1 and log-linear two-variable interactions for a model with three latent classes

| | | Male | | | Female | | | Total | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | $t$ | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 |
| | $\hat{p}_{\theta_t}$ | 0.12 | 0.48 | 0.40 | 0.11 | 0.43 | 0.46 | 0.11 | 0.44 | 0.45 |
| Item 1 | $\hat{p}_{jg\|\theta_t}$ | 0.16 | 0.41 | 0.72 | 0.09 | 0.50 | 0.81 | 0.11 | 0.44 | 0.77 |
| | $\hat{u}_{jg\theta_t}$ | | 1.34 | 2.61 | | 2.38 | 3.80 | | 1.83 | 3.29 |
| Item 2 | $\hat{p}_{jg\|\theta_t}$ | 0.13 | 0.69 | 0.93 | 0.07 | 0.72 | 1.00[a] | 0.08 | 0.69 | 0.97 |
| | $\hat{u}_{jg\theta_t}$ | | 1.48 | 4.46 | | 3.55 | — | | 3.31 | 5.83 |
| Item 3 | $\hat{p}_{jg\|\theta_t}$ | 0.00[a] | 0.20 | 0.94 | 0.00[a] | 0.24 | 0.90 | 0.00[a] | 0.21 | 0.88 |
| | $\hat{u}_{jg\theta_t}$ | | — | — | | — | — | | — | — |
| Item 4 | $\hat{p}_{jg\|\theta_t}$ | 0.41 | 0.95 | 0.97 | 0.59 | 0.96 | 0.99 | 0.49 | 0.94 | 0.99 |
| | $\hat{u}_{jg\theta_t}$ | | 3.29 | 3.87 | | 2.82 | 5.73 | | 2.82 | 4.50 |
| Item 5 | $\hat{p}_{jg\|\theta_t}$ | 0.05 | 0.17 | 0.63 | 0.00[a] | 0.41 | 0.86 | 0.01 | 0.26 | 0.76 |
| | $\hat{u}_{jg\theta_t}$ | | 3.87 | 3.46 | | — | — | | 3.38 | 5.59 |

[a]These values are restricted according to boundary conditions. See the text for further details.

consisting of people who take a pro-women's lib point of view and a class of people who are more traditional in this respect. The first group consists of about 45% of the population, whereas the second group contains some 11%. Between these two extremes there is another class in which people take a pro-women's lib position with respect to the easier items (item 2 and 4), but they express traditional views when responding to the more difficult items (especially item 3 and 5). The estimated proportion of people belonging to this latent class is about 44%. A comparison of the response probabilities patterns for men and women shows that the differences between the latent classes are more pronounced for the female subsample.

The pattern of response probabilities among the three latent classes seems to indicate a dimension along which these latent classes could be ordered. Whether this is the case can, however, not be decided on the basis of the unrestricted latent class analyses presented here. This topic will be returned to later on in this chapter.

In this example the traditional latent class model yields results that can be interpreted rather easily. This parameterization also leads to a number of possible restricted latent class models which have proven to be useful in the context of the analysis of measurement problems. As will be shown in one of the following sections, all probabilistic variants of the Guttman scaling model proposed so far can be expressed as latent class models with restrictions imposed on the conditional response probabilities.

However, it is possible to express the latent class model as a log-linear model with a latent variable. This is the notation used by Haberman (1979). Because of the assumption of local independence, this log-linear model is readily expressed as:

$$\begin{aligned} \ln p_{gk\ell t}^{ABC\theta} &= u + u_g^A + u_k^B + u_\ell^C \\ &\quad + u_t^\theta + u_{gt}^{A\theta} + u_{kt}^{B\theta} + u_{\ell t}^{C\theta}. \end{aligned} \tag{2.23}$$

Sometimes, the joint probability $p_{gk\ell t}^{ABC\theta}$ will also be denoted by $p_{\nu\theta_t}$ in order to remain consistent with the notation used earlier for the Lazarsfeld/Goodman parameterization. The relationship between these unknown joint probabilities and the known probabilities for observing some specific response pattern $\nu$ is then:

$$p_\nu = \sum_{t=1}^{T} p_{\nu\theta_t}. \tag{2.24}$$

This way of expressing the latent class model as an unsaturated log-linear model with a latent variable is a very general formulation of the latent class model (see, for example, Langeheine, 1988). By expressing the logarithm of the joint probabilities as the sum of a number of log-linear parameters, it is possible to place restrictions upon these parameters. This leads to a number of latent class models which lie beyond the scope of the latent class model as parameterized by Lazarsfeld and Goodman.[13] The models that deserve mentioning in this context are the models using linear restrictions on the relationship between the

[13]The reverse, however, is also true. When the latent class model is parameterized as a log-linear model, equality restrictions upon the conditional response probabilities are also not straightforward.

latent and the manifest variables.[14] Moreover, by supplying an estimation method and an algorithm suited for this parameterization, Haberman also made it possible to perform multi-group analyses as well as analyses in which the response patterns are weighted differentially. The topic of estimating parameters in the latent class model is taken up in the next section.

The log-linear formulation makes it possible to derive an expression for the conditional probabilities of scoring in some category of an item, given that a person belongs to a specific latent class (i.e., probabilities like $p_{gt}^{\bar{A}\theta}$). Thus, it is very easy to establish a relationship between these conditional probabilities, which play such a distinctive role in the Lazarsfeld/Goodman formulation, and the log-linear parameters for Haberman's approach. Moreover, these expressions will prove helpful in exploring the relation between latent class models and latent trait models, as these latter models are nearly always formulated in terms of conditional response probabilities. The probability that an individual belonging to latent class $\theta_t$ will respond to item $A$ in category $g$ can be written as:

$$\begin{aligned} p_{gt}^{\bar{A}\theta} &= \frac{p_{gt}^{A\theta}}{p_{\theta_t}} \\ &= \frac{\sum_k \sum_\ell p_{gk\ell t}^{ABC\theta}}{\sum_g \sum_k \sum_\ell p_{gk\ell t}^{ABC\theta}} \\ &= \frac{\exp(u_g^A + u_{gt}^{A\theta}) \sum_k \sum_\ell \exp(u + u_k^B + u_\ell^C + u_t^\theta + u_{kt}^{B\theta} + u_{\ell t}^{C\theta})}{\sum_g \exp(u_g^A + u_{gt}^{A\theta}) \sum_k \sum_\ell \exp(u + u_k^B + u_\ell^C + u_t^\theta + u_{kt}^{B\theta} + u_{\ell t}^{C\theta})} \\ &= \frac{\exp(u_g^A + u_{gt}^{A\theta})}{\sum_g \exp(u_g^A + u_{gt}^{A\theta})}. \end{aligned} \tag{2.25}$$

The latter expression shows how to translate restrictions in terms of conditional response probabilities into restrictions put upon the log-linear parameters, and vice versa. It is also very convenient in exploring the relations between latent trait models and restricted latent class models. Another advantage of this notation is that it is more general, in the sense that it does not depend on the number of variables as do the parameterizations used by Goodman and Haberman. The notation in

[14]These models were introduced briefly in Section 1.3.2 in Chapter 1. These models are reviewed in more detail in Section 2.4.3 of this chapter.

the following will therefore be changed slightly. This modified notation was also used in the previous chapter. The index $j$ will be used for an arbitrary item, so item $A$ will be indicated by $j = 1$, item $B$ by $j = 2$, etc. The expression for the conditional response probabilities then becomes:

$$p_{jg|\theta_t} = \frac{\exp(u_{jg} + u_{jg\theta_t})}{\sum_g \exp(u_{jg} + u_{jg\theta t})}. \tag{2.26}$$

From this expression it can be verified that the logits can be expressed as

$$\ln \left[ \frac{p_{jg|\theta_t}}{p_{j0|\theta_t}} \right] = u_{jg} + u_{jg\theta_t}.$$

This equation was presented earlier in Chapter 1 using Goodman's notation (Equation 1.6).

There is one drawback to using the modified notation proposed here because not all parameters that have to be estimated appear in this formulation. Besides the sets of log-linear parameters (or equivalently the conditional response probabilities), we also have to estimate either the one-variable log-linear parameters which pertain to the latent variable $\theta$, i.e., the parameters $u_{\theta_t}$, or the latent proportions $p_{\theta_t}$.

It should be noted that the parameterization given in Equation 2.26 was already given by Haberman himself (Haberman, 1979). The parameterization in terms of latent proportions and response probabilities that are expressed as functions of log-linear parameters is also a special case of Formann's *linear logistic latent class analysis* (Formann, 1992). The linear logistic LCA model is more general because it expresses not only the response probabilities, but also the latent proportions as a function of log-linear parameters. Moreover, with linear logistic LCA, these log-linear parameters do not necessarily pertain to just the manifest and the latent variables; they can also apply to other external variables. Because the relationship with other external variables is not explored in depth in this chapter, linear logistic LCA will also not be discussed here.

The two parameterizations presented in this section have also given rise to different estimation procedures. This is not to say that a specific formulation of a latent class model will necessarily lead to a particular estimation method. But both Goodman and Haberman proposed not only different parameterizations, but also different estimation procedures. The estimation of the parameters in latent class models is the topic of the next section.

### 2.3.1 Estimation in the latent class model

ML estimation of the parameters in the unrestricted latent class model[15] can be performed in a number of ways, each with its own benefits and drawbacks. The most frequently encountered numerical methods for obtaining maximum likelihood estimates are Newton-Raphson, the scoring algorithm and the EM algorithm.[16] Estimation in the latent class model is more difficult than in the ordinary log-linear models with direct observations since the data-matrix is incomplete because of the latent variable. This results in an expression for the log-likelihood with a more complex structure. First, the Newton-Raphson procedure will be reviewed briefly. Next, the scoring algorithm and a modified Newton-Raphson procedure proposed by Haberman (1988) are paid attention to. Finally, a great deal of this section is devoted to estimation of parameters in the latent class model with EM.

In the log-linear parameterization of the latent class model as formulated by Haberman, it can be shown that the same likelihood equations hold as in the case of the ordinary log-linear model. But in these equations the expected frequencies $(N \cdot p_{\nu\theta_t})$ appear, which are not observed (see Haberman, 1979 pp 543-544). It is possible to estimate these frequencies, but the resulting likelihood equations are not solved easily by means of Newton-Raphson. The procedure is numerically not very stable. As a result Newton-Raphson is not widely used for the estimation of parameters in the latent class model. Haberman (1979) proposed the scoring method as an alternative. In this method the matrix of second derivatives of the log-likelihood to the parameters is replaced by its expected value, the Fisher information matrix, also called the expected information matrix. It should be noted that, in contrast to the situation for log-linear models, the expected information matrix for latent class models is not the same as the matrix of second derivatives (see

[15]The estimation procedures have to be changed in order to include restrictions on the parameters. These changes will be described later on, after the different kinds of restrictions that can be applied have been discussed.

[16]Formann (1984) also uses the gradient method. This last method is not discussed in this book. The reader is referred to Formann (1984) for further details.

Appendix A). A minor drawback is that the information matrix is somewhat more complex to compute than the matrix of second derivatives. But one gains in terms of numerical stability by replacing Newton-Raphson by the scoring algorithm. The scoring method can also be rewritten as an iterated weighted least squares solution, as was the case with the Newton-Raphson procedure for log-linear models. This scoring algorithm is used in Haberman's program LAT (Haberman, 1979). As Haberman himself already noted, it is often very difficult to find initial values for the parameters which will lead to convergence of the algorithm. Because of this, the scoring algorithm has not become very popular among social researchers. Recently, Haberman (1988) suggested a modified Newton-Raphson procedure where the matrix of second derivatives is adjusted whenever this matrix is not negative definite. The resulting algorithm has been used in Haberman's program *NEWTON*. The algorithm converges very fast, it is less sensitive to the choice of initial estimates, but it also requires a lot of computing time. Hagenaars (1990) reports satisfactory results obtained with *NEWTON* even when the initial estimates were rather crude figures as compared to the final solution. He used the program for a number of restricted latent class models, with restrictions on the conditional response probabilities. However, when linear restrictions are imposed on the association coefficients between the latent and the manifest variables it is much more difficult to get the algorithm to converge. Although very flexible and definitely a lot more satisfactory than LAT, the *NEWTON* program is not well suited for estimating all sorts of restricted latent class models.

An alternative to these Newton-Raphson and scoring methods is the EM method proposed by Dempster et al. (1977). This procedure can be used in the wide context of models with missing data. As the latent variable is not observed directly, the scores on this latent variable can be seen as missing. The EM algorithm consists of two steps:

1. In the first step (this is called the E(xpectation)-step) the sufficient statistics for the parameters are estimated. In the context of latent class models this means that the probabilities for the complete data-matrix $p_{\nu\theta_t}$ are estimated using the observed proportions $(f_\nu/N)$.

2. In the second step (which is called the M(aximization)-step) the

log-likelihood for the complete data matrix is maximized in order to obtain new estimates of the model parameters. Using these new model estimates, another E-step can be performed in order to obtain new estimates for the complete data, etc.

The EM algorithm thus consists of an E-step and an M-step which are alternatively performed until the procedure converges. The application of this algorithm in the context of latent class models is very simple.[17] First, some points concerning the notation that was introduced at the end of the previous section, and that was also used in Chapter 1, should be recapitulated. The probability that an individual will respond in category $g$ of item $j$, given that this individual belongs to latent class $\theta_t$, is denoted by $p_{jg|\theta_t}$ where $j = 1, \ldots, n$ and $g = 0, \ldots, m_j$, so that the number of categories for item $j$ is equal to $(m_j + 1)$. As before, the number of individuals that has responded with response pattern $\nu$ is equal to $f_\nu$. Furthermore, some indicator variables $x_{\nu jg}$ are needed. These dichotomous indicator variables are used to simplify the notation for the log-likelihood. The following rules determine when $x_{\nu jg}$ is equal to 0 or 1:

| | |
|---|---|
| $x_{\nu jg} = 1$ | if in response pattern $\nu$ item $j$ is responded to with category $g$, |
| $x_{\nu jg} = 0$ | otherwise. |

Thus, for each item $j$, it will be true that

$$\sum_{g=0}^{m_j} x_{\nu jg} = 1. \tag{2.27}$$

Assuming local independence, the probability of observing response pattern $\nu$ in latent class $\theta_t$ can be expressed as:

$$p_{\nu|\theta_t} = \prod_{j=1}^{n} \prod_{g=0}^{m_j} p_{jg|\theta_t}{}^{x_{\nu jg}}. \tag{2.28}$$

Starting with initial estimates for the conditional response probabilities $p_{jg|\theta_t}$ (or with initial estimates for the log-linear parameters $u_{jg}$ and

[17]The notation used here, as well as the explanation of the EM algorithm, leans heavily on Croon (1990).

$u_{jg\theta_t}$), Equation 2.28 can be used to compute estimates for the probabilities $p_{\nu|\theta_t}$. Then Equation 2.22 can be used to compute initial estimates for the probabilities $p_\nu$:

$$p_\nu = \sum_{t=1}^{T} p_{\nu\theta_t}.$$

Furthermore, by Bayes' theorem:

$$p_{\theta_t|\nu} = \frac{p_{\nu|\theta_t} \cdot p_{\theta_t}}{p_\nu}. \tag{2.29}$$

These estimates for the probability of belonging to latent class $\theta_t$ given that one has responded with response pattern $\nu$ (i.e., the so-called allocation probabilities) can be used for estimating the complete data. When the estimated number of persons who responded with pattern $\nu$ and also belong to latent class $\theta_t$ is denoted by $e_{\nu\theta_t}$, initial estimates for these complete table frequencies can be computed by:

$$e_{\nu\theta_t} = f_\nu \cdot p_{\theta_t|\nu}. \tag{2.30}$$

It is important to note that in the estimation procedure for parameters in the latent class model two kinds of expected frequencies appear. Firstly, there are $e_{\nu\theta_t}$ frequencies, the estimated observed frequencies. These are estimates of the frequencies within the complete table under the restriction that the observed marginals $f_\nu$ should be reproduced exactly. In the calculation of the test statistics, however, the expected frequencies $N \cdot p_\nu$ are used. These expected frequencies will not reproduce the marginals $f_\nu$.

The calculation of the estimated observed frequencies $e_{\nu\theta_t}$ completes the E-step. Now an M-step has to be performed in order to derive new estimates for the log-linear parameters $u_{jg}$ and $u_{jg\theta_t}$ (or for the conditional response probabilities $p_{jg|\theta_t}$) and for the latent proportions $p_{\theta_t}$. These new estimates can then be used again for another E-step, etc.

The M-step can be carried out in a number of ways. Since at the start of the M-step a provisional complete table is known, the estimation procedures developed for the ordinary log-linear model with direct observations can be used. This means that both Newton-Raphson and iterative proportional fitting can be employed in order to obtain new parameter estimates by maximizing the log-likelihood. Because IPF is the easiest procedure to apply, it is the most natural choice[18] and it is applied in two common programs for latent class models, Hagenaars' LCAG and Clogg's MLLSA. Moreover, the conditional response probabilities (or the log-linear parameters $u_{jg}$ and $u_{jg\theta_t}$) can be estimated independently from the latent proportions $p_{\theta_t}$. It must be remembered that after an E-step has been carried out, the estimated complete data matrix can be used in the following M-step. This means that in the M-step a log-likelihood based upon the complete data can be used. This likelihood function has the following form:

$$L = \prod_{\nu} \prod_{t=1}^{T} [p_{\nu\theta_t}]^{e_{\nu\theta_t}} . \tag{2.31}$$

By using Equation 2.28, this likelihood function can be rewritten as:

$$L = \left( \prod_{j} \prod_{t} \prod_{g} p_{jg|\theta_t}{}^{q_{jg\theta_t}} \right) \times \prod_{t} p_{\theta_t}^{e_{+\theta_t}}, \tag{2.32}$$

where $q_{jg\theta_t}$ denotes the number of individuals who belong to latent class $\theta_t$ and respond to item $j$ within category $g$. This number can be expressed as:

$$q_{jg\theta_t} = \sum_{\nu} e_{\nu\theta_t} \cdot x_{\nu jg}.$$

The symbol $e_{+\theta_t}$ stands for the frequency with which individuals belong to latent class $\theta_t$ and this frequency is given by:

$$e_{+\theta_t} = \sum_{\nu} e_{\nu\theta_t}.$$

In the E-step provisional estimates for the observed frequencies in the complete table ($e_{\nu\theta_t}$) and provisional estimates for the bivariate

[18] The argument that Newton-Raphson has an advantage in that it automatically generates estimated standard errors no longer holds in this situation, because in the M-step only a subset of the parameters is estimated using Newton-Raphson. The latent proportions $p_{\theta_t}$ are also estimated in the M-step, but not with Newton-Raphson. For the estimation of $p_{\theta_t}$, Equation 2.33 is used. For this reason, the inverted matrix of second derivatives used in the M-step can no longer be used to compute such estimated standard errors. Another topic that is relevant in this context is the difference between the observed and the expected information matrix. This topic will be dealt with again later on.

marginals ($q_{jg\theta_t}$) were computed. These estimates can be used to estimate the conditional probabilities $p_{jg|\theta_t}$ (or the log-linear parameters $u_{jg}$ and $u_{jg\theta_t}$) in the M-step by maximizing the likelihood in Equation 2.32. The expression for this log-likelihood also clearly shows that the latent proportions $p_{\theta_t}$ can be estimated independently from the other parameters. It can be shown quite easily that the maximum likelihood estimates for the latent proportions are given by:

$$\hat{p}_{\theta_t} = \frac{e_{+\theta_t}}{N}. \tag{2.33}$$

The maximum likelihood estimates for the conditional response probabilities $p_{jg|\theta_t}$ in the unrestricted model are given by:

$$\hat{p}_{jg|\theta_t} = \frac{q_{jg\theta_t}}{q_{j+\theta_t}}, \tag{2.34}$$

with

$$q_{j+\theta_t} = \sum_g q_{jg\theta_t}.$$

When certain restrictions are imposed on the conditional response probabilities, the estimation is less straightforward. Estimation can, in fact, be rather complex when certain equality constraints are combined with restrictions that fix certain conditional response probabilities at particular values, as was shown by Mooijaart and van der Heijden (1992). Further details on this subject can be found in Section 2.4. Details for applying equality constraints in less complex situations are given by Hagenaars (1990). The procedure for estimating the log-linear parameters in the M-step is directly comparable to the one used for ordinary unsaturated[19] log-linear models. As mentioned earlier, this can be done by applying either IPF or Newton-Raphson.

The discussion mentioned earlier concerning the relative merits of IPF and Newton-Raphson in the context of log-linear models also applies to the case of latent class models. The question is whether parameters should be estimated by applying EM on the one hand or Newton-Raphson or scoring on the other. Standard Newton-Raphson suffers from numerical instability; the scoring algorithm only gives satisfactory results if proper initial values are specified, and these may be hard

[19]The fact that this log-linear model is not saturated is a consequence of the assumption of local independence.

to find. The improved Newton-Raphson algorithm proposed by Haberman seems, in many cases, to be a good alternative. The possibility of making use of any of these three estimation procedures, often mentioned as an advantage, is that an estimate of the variance-covariance matrix of the estimated parameters can be produced, while EM does not offer such information without additional manipulation. However, this advantage is a relative one. The Newton-Raphson procedures and the scoring algorithm are time consuming because of the calculation of these matrices, in contrast to the EM procedure. Also, nothing prevents the calculation of an estimated variance-covariance matrix in the context of EM, but the calculations can only be performed in addition to the estimation of the parameters.

It should be noted that the estimation of the variance-covariance matrix of the parameters in the context of latent class analyses can be performed in two ways. The more traditional method of tackling this problem is to take the negative of the expected information matrix. The inverse of this matrix can be shown to be an estimate of the variance-covariance matrix. However, another possibility is to use the matrix of second derivatives of the log-likelihood to the parameters. The inverse of the negative of this matrix can also be shown to be an estimator for the variance-covariance matrix. The observed information matrix is defined as minus the matrix of second derivatives evaluated for the ML estimates. Efron and Hinkley (1978) give a number of examples where the observed information matrix should be preferred to the expected information matrix, but this position has been challenged.[20] The most direct method for estimating the variance-covariance matrix is, however, to perform additional calculations using the usual expressions for either the observed or the expected information matrix. These expressions can be found in Appendix A. Both procedures lead asymptotically to the same results.

In the aforementioned, the estimation of parameters in the unrestricted latent class model was discussed. The procedures sometimes

[20]See also the comments on the Efron & Hinkley article in the same issue of Biometrika. Louis (1982) has pointed out how the observed information matrix can be calculated when using the EM algorithm, while Meilijson (1989) proposed a modification of the EM algorithm in order to estimate the standard errors of the parameters via the empirical information matrix, a less known variant of the information matrix.

require adaptation when restrictions are imposed on the parameters. The kinds of restrictions that are sensible in the context of latent class analysis will be the subject of one of the following sections. In the next section some special aspects of testing and model selection with latent class models will be discussed.

### 2.3.2 Testing and model selection in latent class analysis

In the discussion of model selection and testing for log-linear models, several strategies were mentioned that can be used in the context of latent class models. However, there are several differences that should be kept in mind when selecting and testing latent class models.

Firstly, it should be noted that in order to perform a statistical test on a model, it is necessary for the model to be identifiable. This is not necessarily the case for latent class models, as situations exist in which the parameters cannot be uniquely determined. A necessary condition for identifiability is that the number of degrees of freedom is not negative. The number of parameters that must be estimated independently cannot be greater than the number of independently observed frequencies. However, this condition is nevertheless insufficient. Thus, there are situations in which the model is not identified, despite the fact that the number of degrees of freedom is not negative. Goodman (1974b) has formulated a sufficient condition for local identifiability. In order to check the local identifiability, a matrix is constructed in which the rows consist of the $\left[\prod_{j=1}^{n}(m_j+1)\right]-1$ nonredundant response probabilities $p_\nu$ as defined by the model, and the columns pertain to the non-redundant model parameters (i.e., $p_{jg|\theta_t}$, $p_{\theta_t}$ or, when the log-linear representation is used, $u_{jg}$ and $u_{jg\theta_t}$). The cells of this matrix are filled with expressions for the first derivatives of $p_\nu$ to the model parameters.[21] These expressions are given at the end of Appendix A. Sample estimates for these first derivatives can be found by substituting the maximum likelihood estimates $\hat{p}_{jg|\theta_t}$, $\hat{p}_{\theta_t}$ and $\hat{p}_{\nu|\theta_t}$ in these expressions for the first derivatives. By calculating the rank of this matrix, a test for

[21] Some authors (i.e., Hagenaars (1990) and McCutcheon (1987) state that one should use the derivatives of the *observed* probabilities i.e. $f_\nu/N$ to the nonredundant model parameters. This is, of course, incorrect because the observed probabilities cannot be expressed as a function of the model parameters.

local identifiability can be performed. The matrix must be of full column rank; if the matrix is not of full rank, the model is not locally identifiable.

This test of local identifiability can also be performed using the estimated expected information matrix. Dayton and MacReady (1980) stated that the theoretical variance-covariance matrix for the parameters should be of full rank. Thus, by investigating the rank of estimated asymptotic variance-covariance matrix, a test of local identifiability can be carried out. Formann (1985) discussed the properties which the expected matrix of the second-order partial derivatives should have in order for the solution to correspond to a local maximum of the likelihood function. If all the eigenvalues of this matrix are negative (and, therefore, the matrix is negative definite), the model is locally identifiable.[22]

Models which are not identified can be made identifiable by putting restrictions on the parameters.[23] The various possibilities for applying restrictions will be surveyed in the next section.

With regard to the statistical testing of the usual null-hypothesis (i.e., the model is valid for the population) it is still true that the log-likelihood ratio asymptotically follows a $\chi^2$ distribution when certain regularity conditions are met. This means that for unrestricted latent class models, the usual testing procedures can be carried out as was explained in the section concerning log-linear models. However, when certain parameter estimates are on the boundary of the parameter space (e.g., estimated probabilities which are equal to 0 or 1), this may indicate a situation in which the parameters are also on the boundary. In that case, the standard regularity conditions will break down and the likelihood ratio test statistic will no longer approximate a $\chi^2$ distribution. When such a *terminal solution* is obtained, one can usually assume that the parameter with an estimated value on the boundary was restricted to that value on an a priori basis so that with one extra degree of freedom the model can still be tested. This convention

[22]Because the observed information matrix equals the negative of the matrix of second-order derivatives, this observed information matrix (and also its inverse, i.e., the estimated variance-covariance matrix) should be positive definite, i.e., all eigenvalues should be greater than zero for the model to be locally identifiable.

[23]Unfortunately, the opposite is also true. An identifiable model can be made unidentifiable by putting certain restrictions on the parameters. Examples are given in Hagenaars (1990) and Goodman (1974b).

was proposed by Goodman (1974a; 1974b) and has been widely applied since its proposal; as was noted in the previous section, this practice is, strictly speaking, not admissible.

The same type of problem also arises when one wants to test hypotheses about the number of latent classes, assuming that one would like to test whether a model with two or a model with three latent classes should be used to explain the data under study. It is not permitted to take the difference in the log-likelihood ratios and test this statistic with the difference in degrees of freedom for the two models, because the difference in the two $G^2$-values is not asymptotically distributed as a $\chi^2$-variable. The reason for this is that, again, standard regularity conditions break down as a consequence of the fact that the model with two latent classes can only be expressed as a restricted version of the model with three latent classes by setting one $p_{\theta_t}$ in the latter model equal to 0 (for a more thorough treatment of these ideas, see McLachlan and Basford, 1988).[24] Thus, one would test the hypothesis that one parameter will take a value on the boundary of the parameter space. In that case, the regularity conditions will no longer hold.

It could be presumed that this problem could be solved by comparing models with a different number of latent classes using the Akaike information criterion or one of the modified information criteria described in the previous sections. However, these criteria rely on the same regularity conditions as the log-likelihood ratio does in order for $G^2$ to be asymptotically distributed as a $\chi^2$-variable. Therefore, the information criteria also do not resolve this problem.

All that has been said previously should not lead to the conclusion that it is impossible to perform statistical tests for certain restrictions on the parameters. It is possible to test whether certain parameters can be restricted to take the same value as opposed to the alternative where they are allowed to vary freely. The point is that it is not possible to test hypotheses when parameters are set equal to values on the boundary of the parameter space. Some researchers have suggested modifications which should be applied to the likelihood ratio statistic and/or to the number of degrees of freedom in order to improve the approximation of the actual sampling distribution by a theoretical $\chi^2$

[24]It should be noted that the log-linear $u$-parameters can be set equal to 0 because 0 is not a boundary value for the log-linear parameters.

distribution when parameter-estimates are on the boundary of the parameter space. Monte Carlo research by Holt and MacReady (1989) has shown, however, that the deviance between the actual sampling distribution and the $\chi^2$ distribution cannot be readily predicted from a number of relevant variables such as sample sizes or the skewness of the latent distribution. This casts doubt upon the possibility of applying standard corrections when parameter estimates take values near 0 or 1.

Besides performing formal statistical tests, it is always possible to calculate the more descriptive statistics indicating the fit of a specified model. All indices presented earlier can be applied in the context of latent class analysis.

### 2.3.3 Strength of association between latent and manifest variables

Apart from the question of whether a model provides a good fit to the data, it is worthwhile to pay attention to the quality of the measurement model by investigating the relationship between the latent variable and the manifest indicators. The idea behind this notion is simple: the measurement model improves as the association between the latent and the manifest variables grows stronger. One way to indicate the association between the latent and the manifest variables is through the two-variable log-linear parameters (i.e., $u_{jg\theta_t}$ in Equation 2.26). The estimated values of these parameters for the unrestricted latent class model with $T = 3$ latent classes are reported in Table 2.4. Because the latent variable has three categories and dummy coding is used, these log-linear parameters are only estimated for the second and the third latent class. The parameter $u_{jg\theta_2}$ is equal to the log odds of the $2 \times 2$ table consisting of a dichotomous manifest indicator and the latent variable with only categories one and two. The log-linear parameter $u_{jg\theta_3}$ equals the log odds when only categories one and three of the latent variable are taken into account.

The estimated values of the log-linear parameters were all positive and for each item $u_{jg\theta_3}$ was greater than $u_{jg\theta_2}$. This means that the probability for responding in category 1 increased for the three successive latent classes. The larger the value of these coefficients was, the stronger the association between the latent variable and a manifest indicator were. Because some estimated conditional response probabilities were equal to zero, not all log-linear parameters could be estimated.

Another possibility is to operationalize the strength of association in terms of the ability to predict the membership of specific latent classes on the basis of observed response patterns. A central role in this strategy was played by the allocation probabilities $p_{\theta_t|\nu}$, which were already defined in Equation 2.29. On the basis of these allocation probabilities, the *maximum a posteriori* or the MAP estimator can be obtained for each response pattern $\nu$. The MAP estimator for a specific response pattern $\nu$ is defined as the value of the latent class for which the allocation probability takes the greatest value. This maximum allocation probability is denoted by $p^*_{\theta|\nu}$, so:

$$p^*_{\theta|\nu} = \max \left[ \, p_{\theta_1|\nu}, \;\; p_{\theta_2|\nu}, \;\; \ldots, \;\; p_{\theta_T|\nu} \, \right].$$

The strength of the degree of association between the latent and the manifest variables can then be measured by the proportion misclassified $E$ or by the coefficient $\lambda_{\theta_t.\nu}$.

The proportion misclassified can be obtained as follows. When the MAP-estimator is used to allocate individuals to latent classes, the probability of assigning an individual with response pattern $\nu$ to the wrong latent class is given by:

$$1 - p^*_{\theta|\nu}.$$

The overall probability of assigning individuals to a wrong latent class following the modal assignment rule is given by:

$$E = \sum_{\nu} p_\nu (1 - p^*_{\theta|\nu}). \tag{2.35}$$

The higher the value $E$ takes, the weaker the relationship between the latent variable and the manifest indicators. Of course, one could also take the proportion correctly classified, i.e., $1-E$. It is merely a matter of taste whether one uses the proportion misclassified or the proportion correctly classified.

Another method for expressing the relationship between latent and manifest variables was proposed by Clogg (1981). This modification resulted in the well-known coefficient $\lambda$ developed by Goodman and

Kruskal (1954). This measure of association indicates the relative improvement in predicting class-membership with the aid of the observed indicators as compared with the situation where class-membership is predicted without knowledge of the responses on the manifest variables. This coefficient $\lambda$ is defined as:

$$\lambda_{\theta.\nu} = \frac{(1 - p^*_\theta) - E}{(1 - p^*_\theta)}, \tag{2.36}$$

where:

$$p^*_\theta = \max\left[\ p_{\theta_1},\ p_{\theta_2},\ \ldots,\ p_{\theta_T}\ \right].$$

Thus, $p^*_\theta$ refers to the unconditional modal latent class.

By replacing the population parameters in these equations with their maximum likelihood estimates, sample estimates can be obtained for the proportion misclassified as well as the coefficient $\lambda_{\theta.\nu}$. For the latent class models that were presented in Section 2.2, the estimated values $\hat{\lambda}$ and the proportion *correctly* classified (so $1 - \hat{E}$) are reported in Table 2.5. These results clearly show that as the number of latent

Table 2.5: Proportion correctly classified and the coefficient $\hat{\lambda}$ for unrestricted latent class analyses with 2 and 3 latent classes

| | $1 - \hat{E}$ | | $\hat{\lambda}_{\theta.\nu}$ | |
|---|---|---|---|---|
| | $T = 2$ | T=3 | $T = 2$ | T=3 |
| Male | .905 | .831 | .791 | .675 |
| Female | .923 | .851 | .784 | .723 |
| Total | .911 | .823 | .777 | .678 |

classes increased it became more difficult to allocate individuals to latent classes correctly, which is not surprising. Furthermore, although the model with two latent classes did not fit according to the usual likelihood ratio test, the values of both the percentage correctly classified as the coefficient $\lambda$ were relatively high. This indicates that these measures of the strength of the relationship between the latent variable and the manifest indicators reflected other characteristics of the latent class model than the fit indices did.

In this section and the preceding one, the possibility of putting certain restrictions on the parameters in a latent class analysis was mentioned several times. The various types of restrictions that can be used to define more restricted latent class models are discussed in the next section.

## 2.4 Restricted latent class models

Unrestricted latent class models can be useful for an exploratory analysis of a multidimensional cross-classification, but restricted latent class models are more interesting. Restricted models make it possible to test and investigate a number of interesting hypotheses concerning the structure of the measurement model, i.e., the relations between the latent variable and manifest indicators. Furthermore, by imposing restrictions on the model parameters, one may perform detailed multi-group comparisons, include external variables in the measurement model, or construct multidimensional latent class models (see, for example, Hagenaars, 1990). The restrictions that can be used, including a number of interesting measurement models that arise given specific sets of constraints, will be the topic of this section.

As was seen in the previous sections, the latent class model can be parameterized in several ways. The restrictions that can be imposed on the parameters in the LCA model are closely linked to the selected parameterization. The different parameterizations are summarized below:

1. Firstly, the latent class model can be parameterized in terms of latent proportions $p_{\theta_t}$ and conditional response probabilities $p_{jg|\theta_t}$. This parameterization was originally proposed by Lazarsfeld and Henry (1968) and is also used by, for example, Goodman (1974a, 1974b), Clogg (1977, 1981) and Hagenaars (1988, 1990).

2. Haberman (1979) chose a log-linear formulation of the latent class model. In this parameterization (see Equation 2.23), only log-linear parameters appear.

3. In Section 2.3, log-linear parameterization was used to obtain expressions for the conditional response probabilities $p_{jg|\theta_t}$ in terms of the log-linear parameters (see Equation 2.26). Together with the set of latent proportions $p_{\theta_t}$, this provides a third method of parameterizing the latent class model.

Each method of parameterizing the LCA model provides its own opportunities for imposing restrictions on the parameters used. If, for example, the LCA model is parameterized in terms of latent proportions and conditional response probabilities, the latter parameters can be submitted to restrictions. The computer programs written by Clogg (i.e., MLLSA) and by Hagenaars (i.e., LCAG) offer the possibility of formulating models with these kinds of restrictions (Clogg, 1981 and Hagenaars & Luijkx, 1987). These restrictions fall into two different categories. Firstly, it is possible to restrict some parameters so that they will be equal to each other. Secondly, certain parameters can be fixed at a specified value. Equality restrictions and fixed-value restrictions on the latent proportions $p_{\theta_t}$ are, for example, used in multi-group analyses (see, for example, Clogg & Goodman, 1986). In most cases, however, restrictions are imposed on the conditional response probabilities.

By restricting the conditional response probabilities, one can test multidimensional models, perform multi-group comparisons or construct a number of scaling models for dichotomous items. Most of these models require equality constraints, i.e., certain conditional response probabilities are restricted to take the same value. It is, however, also possible to restrict certain conditional response probabilities to take a specified value. In most cases, this will be the value 0 or 1. This makes error-free linkage of a given indicator to the latent variable possible. Besides equality constraints and fixed-value constraints, it is also possible to formulate inequality constraints with respect to the conditional response probabilities. This is done by Croon (1990) in his model for ordinal latent class analysis.

By applying the second method of parameterizing the LCA model, i.e., the log-linear representation proposed by Haberman, it is possible to impose linear restrictions on the log-linear parameters. The merit of this approach is that it allows the formulation of models for variables measured on an interval scale. A short introduction to this type of model was presented in Chapter 1. The computer programs provided by Haberman, LAT (Haberman, 1979) and NEWTON (Haberman, 1988), offer possibilities for imposing restrictions on the log-linear parameters not found in MLLSA or LCAG. However, it is not intuitively clear how restrictions on conditional response probabilities should be handled with Haberman's programs. Although it is not entirely impossible, it is a major drawback of the log-linear formulation that equality and fixed-value restrictions on the latent proportions and the conditional response probabilities are not easy to incorporate.

By focussing on the third method of parameterization, i.e., by expressing the conditional response probabilities $p_{jg|\theta_t}$ in terms of log-linear parameters as is done in Equation 2.26, it can be made clear for a number of simple situations how equality restrictions on the conditional response probabilities can be translated into restrictions on the log-linear parameters. The simplest situations occur when the manifest indicators are dichotomous.

In the following, it is assumed that there are equality restrictions on the conditional response probabilities of one variable $j$ or two variables $j$ and $j'$. Furthermore, the restrictions can pertain to probabilities within one latent class $\theta_t$ or to probabilities for two different latent classes $\theta_t$ and $\theta_{t'}$. Next, the two categories for the manifest items are scored 0 and 1. Finally, it is assumed that a dummy coding scheme has been applied with category 0 for the manifest variables and category $\theta_1$ for the latent variable treated as the reference-category (i.e., for these categories the log-linear parameters are restricted to zero).

It is now possible to write the two most general expressions for the relationship between equality restrictions among conditional response probabilities on the one hand and the resulting equality restrictions among sets of log-linear parameters on the other hand as follows:

$$\begin{aligned} p_{j1|\theta_t} = p_{j'1|\theta_{t'}} &\Leftrightarrow u_{j1} + u_{j1\theta_t} = u_{j'1} + u_{j'1\theta_{t'}}, \\ p_{j1|\theta_t} = p_{j'0|\theta_{t'}} &\Leftrightarrow u_{j1} + u_{j1\theta_t} = -(u_{j'1} + u_{j'1\theta_{t'}}). \end{aligned} \tag{2.37}$$

These general expressions are at this point not yet very informative, but in a number of situations simplified and more meaningful expressions can be obtained.

Supposing that $\theta_{t'}$ is equal to $\theta_1$, i.e., $\theta_{t'}$ is the reference-category, the expressions can now be simplified as follows:

$$\begin{aligned} p_{j1|\theta_t} = p_{j'1|\theta_1} &\Leftrightarrow u_{j1} + u_{j1\theta_t} = u_{j'1} \\ &\qquad \text{or} : u_{j1\theta_t} = u_{j'1} - u_{j1}, \\ p_{j1|\theta_t} = p_{j'0|\theta_1} &\Leftrightarrow u_{j1} + u_{j1\theta_t} = -u_{j'1} \\ &\qquad \text{or} : u_{j1\theta_t} = -u_{j'1} - u_{j1}. \end{aligned}$$

A further simplification follows when $\theta_t = \theta_{t'} = \theta_1$. In this case, the equality of corresponding probabilities results in the equality of corresponding one-variable parameters, and the equality of opposite probabilities is identical to the one-variable parameters which form each other's complement:

$$\begin{aligned} p_{j1|\theta_1} &= p_{j'1|\theta_1} \;\Leftrightarrow\; u_{j1} = u_{j'1}, \\ p_{j1|\theta_1} &= p_{j'0|\theta_1} \;\Leftrightarrow\; u_{j1} = -u_{j'1}. \end{aligned}$$

Assuming that the probabilities pertain to two different latent classes but to the same manifest item, i.e., $\theta_t \neq \theta_{t'}$ but $j = j'$, the expressions given in 2.37 can also be simplified. Consider first the case where $\theta_t \neq \theta_1$ and also $\theta_{t'} \neq \theta_1$. Under these assumptions, the expressions become:

$$\begin{aligned} p_{j1|\theta_t} &= p_{j1|\theta_{t'}} \;\Leftrightarrow\; u_{j1\theta_t} = u_{j1\theta_{t'}}, \\ p_{j1|\theta_t} &= p_{j0|\theta_{t'}} \;\Leftrightarrow\; u_{j1\theta_t} = -2u_{j1} - u_{j1\theta_{t'}}. \end{aligned}$$

Finally, in the simple case where $j = j'$, $\theta_t \neq \theta_{t'} = \theta_1$, the following results are obtained:

$$\begin{aligned} p_{j1|\theta_t} &= p_{j1|\theta_1} \;\Leftrightarrow\; u_{j1\theta_t} = 0, \\ p_{j1|\theta_t} &= p_{j0|\theta_1} \;\Leftrightarrow\; u_{j1\theta_t} = -2u_{j1}. \end{aligned}$$

The pattern of restrictions on the conditional response probabilities can be translated into a set of restrictions on the log-linear parameters. However, the translation is only simple if the manifest items are dichotomous. For polytomous items the relation between the two sets of restrictions is not transparent. In the case of polytomous items and/or a rather complex pattern of restrictions, it is more convenient to use the parameterization given by Goodman (1974a, 1974b).

A topic which has not yet been discussed concerns the estimation of parameters in restricted latent class models. Equality constraints and fixed-value constraints on the latent proportions $p_{\theta_t}$ can easily be incorporated into the EM estimation framework. The parameter estimates that result from a particular M-step are submitted to the constraints before the next E-step is carried out. In the case of fixed-value constraints, this entails that the latent proportions which have been left free are rescaled so that the sum of all latent proportions (fixed and free) equals one. When equality constraints are imposed on the latent proportions, the estimates obtained in the M-step for the constrained proportions are replaced by their arithmetic mean.

While constraints on class sizes are not problematic, restrictions on the conditional response probabilities $p_{jg|\theta_t}$ can cause severe problems. For a long time it was believed that restrictions on the conditional response probabilities could be handled in much the same way as restrictions on the latent proportions. This strategy was proposed by Goodman (1974a, 1974b). Thus, fixed-value restrictions were incorporated by rescaling the estimates obtained in the E-step, while equality restrictions were handled by calculating the weighted mean of the estimates for the restricted parameters, with the class sizes used as weights. As Mooijaart and van der Heijden (1992) pointed out, this estimation procedure will not work in some specific situations. In general, it can be shown that when equality restrictions or fixed value-restrictions are imposed on the conditional response probabilities, estimates for the parameters can be found by solving a system of non-linear equations. In a number of situations, these non-linear equations have a closed-form solution that is identical to Goodman's procedure. The reader is referred to Mooijaart and van der Heijden (1992) for more details on the specific situations in which the Goodman procedure provides the correct results. In all other situations, the system of non-linear equations has to be solved iteratively, for example, by Newton-Raphson. After each M-step, therefore, new parameter estimates for the response probabilities must be found iteratively before the next E-step starts. The available computer programs do not offer such possibilities.

As was stated earlier, in some simple situations equality restrictions on the conditional response probabilities can be translated directly into constraints on the log-linear parameters in Equation 2.26. It can be shown that these simple situations belong to the cases classified by Mooijaart and van der Heijden (1992) as situations allowing explicit solutions for the restricted parameters. In other words, in these simple situations the system of non-linear equations can be simplified, with the direct solutions equivalent to the solutions obtained with Goodman's strategy of calculating weighted means.

Even if the parameterization consists of latent proportions and a formulation of the conditional response probabilities in terms of log-linear parameters as was done in Equation 2.26, it is still convenient to estimate the parameters using an EM procedure. If the M-step contains a Newton-Raphson procedure, the restrictions can be accounted for directly by incorporating them in the expression for the log-likelihood. This procedure is used in Chapter 3. A drawback of this procedure is that for different kinds of restricted models, different expressions for the log-likelihood must be maximized.

Finally, when the parameterization is in terms of the log-linear parameters as in Equation 2.23, scoring algorithms or a (modified) Newton-Raphson procedure will almost always be employed to estimate the parameters. In the programs developed by Haberman, LAT and NEWTON, linear restrictions on the log-linear parameters can be taken into account by supplying the program with a properly adapted design matrix. This approach has, however, a minor drawback. Although not really difficult, it is also not very challenging to construct a design matrix for problems with a relatively large number of response patterns and a rather complex pattern of restrictions.

The various types of restrictions that were presented in this section can be employed to construct a number of measurement models, most of which are probabilistic variants of the Guttman scaling method. In the next section, these models are used to analyze the data on women's liberation that were previously analyzed using unrestricted latent class models.

### 2.4.1 Latent class models for scale response patterns

In this section, a number of restricted latent class models will be presented, all of which are probabilistic versions of the Guttman scale (for a discussion of the Guttman scale, see, for example, Torgerson, 1962). The principles on which Guttman's scalogram technique is founded, can be easily explained using an example with three dichotomous items $A$, $B$ and $C$. The two response alternatives are scored with the numbers 0 and 1. If the three items all refer to some common latent variable and this latent variable is considered continuous then both individuals and items are located along this unidimensional latent continuum. Furthermore, when an individual is located to the right of a given item along this latent continuum, then this individual should agree with that item. The response category 1 is reserved for agreeing with an item. If an individual is located to the left of an item that individual will not agree with that item because the item is too "difficult" for that individual. This will result in an observed score of 0. With three manifest items of which $A$ is the easiest item and $C$ is the most difficult item, only the response patterns denoted in Table 2.6 are legitimate under the Guttman scale model.

Table 2.6: Guttman-type response patterns with three items

| Item $A$ | Item $B$ | Item $C$ |
|---|---|---|
| 0 | 0 | 0 |
| 1 | 0 | 0 |
| 1 | 1 | 0 |
| 1 | 1 | 1 |

In the abovementioned, a set of five items concerning women's liberation was analyzed by means of unrestricted latent class analysis. The data for the male subsample will be used here again to illustrate some restricted latent class models that are constructed on the basis of the ideas formulated by Guttman.[25]

In order to explain the Guttman scale pattern it is necessary to order the items according to the degree of difficulty. It is assumed that this ordering can be made along a unidimensional continuum. As there is no a priori information that can be used to order the items, the observed data was employed for this aim. The difficulty of the items was related to pro-women's lib answers on the 5 statements that were presented to the respondents. Thus, an item was considered more difficult, if less respondents did not agree with an item (and thereby take a pro-women's lib point of view). This meant that item 4 was the easiest item, followed by the items 2, 1 and 3, respectively. Item 5 was the most difficult item for the male respondents (see Table 2.1).

The Guttman scale assumes that a person who agrees with the most difficult item (i.e., item 4) will also agree with all the other (easier) items. Therefore, the ordering of items is cumulative. Once a certain threshold is passed, i.e., when an individual agrees with a certain item, it is assumed that all thresholds belonging to easier items have also

[25] Again, the data for the subsample consisting of men that fulfilled the educational requirements mentioned earlier was used.

been passed (i.e., the individual should have agreed with all items that are easier).

It is obvious that, in general, there are $n+1$ (in this case 6) legitimate response patterns. For the example presented here, these patterns are denoted in Table 2.7. The frequencies in Table 2.2 indicate that a total

Table 2.7: Guttman-type conditional response probabilities; difficulty order for male respondents

| Item 1 | Item 2 | Item 3 | Item 4 | Item 5 |
|---|---|---|---|---|
| 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 1 | 0 |
| 0 | 1 | 0 | 1 | 0 |
| 1 | 1 | 0 | 1 | 0 |
| 1 | 1 | 1 | 1 | 0 |
| 1 | 1 | 1 | 1 | 1 |

of 340 males responded with one of these response patterns. Thus, 63% of all respondents used in this analysis conformed to the rules of the Guttman scale. Strictly speaking, the other 37% were not scalable because they responded with one of the remaining $2^n-(n+1)$ response patterns.

This rather brief exposition of the properties of the Guttman scale also reveals the demands that are placed upon the items. Firstly, the manifest items must be dichotomous. Secondly, the items must be monotone, i.e., the item characteristic curves should be monotone. An example of a monotone item characteristic curve (also called a trace line) that conforms to the Guttman scale is shown in Figure 2.1. It is clear from this figure that the trace lines in the Guttman scale model are step functions. This trace line clearly shows that up to some point along the continuum, the probability that a subject will give a positive response (in this example: will not agree) is 0, but once the difficulty location of that item is passed, the probability of a positive response is 1. Items that are more difficult have the same trace line, except that the threshold values for these items are located more to the right. Easier items have threshold values more to the left.

A nonmonotone item has a different kind of item characteristic curve. For these items, the probability for a positive response will increase up

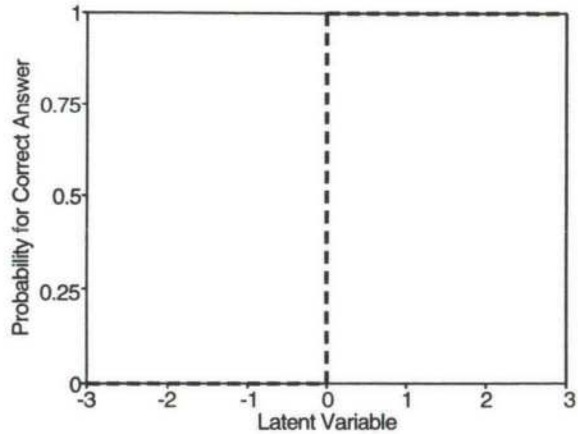


Figure 2.1: Item characteristic curve for a Guttman scale item

to some point and after reaching a maximum, that probability will again decrease. A trace line for a perfectly discriminating item (i.e., the probability of giving a positive response is either 0 or 1) that is non-monotonous is pictured in Figure 2.2. Especially with the study of attitudinal topics, one cannot always be sure in advance that items will be monotone. So there is no guarantee for *substantial monotonicity* (Jansen, 1983). Procedures for analyzing non-monotone items can be found in Formann (1988), Hoijtink (1990) and Croon (1993).

Although the Guttman scale rests upon a few simple and attractive ideas, it cannot be used in many practical situations, because the demands the Guttman scale puts upon the data are too rigid, i.e., in most cases it will be concluded that the data will not conform to the pattern predicted by the Guttman scale. It has been argued that when the data are not too deviant from the Guttman scale pattern, it is still possible to use the Guttman scale as a measurement device. In this context use is made of rules of thumb resting upon coefficients like the coefficient of reproducability and the coefficient of scalability (see, for example, Clogg and Sawyer, 1981, for an explanation of these coefficients). Apart from the problem that it is not clear what values these coefficients should take in order to justify the use of the Guttman scale, it is also true that in examples where the coefficients reach values that

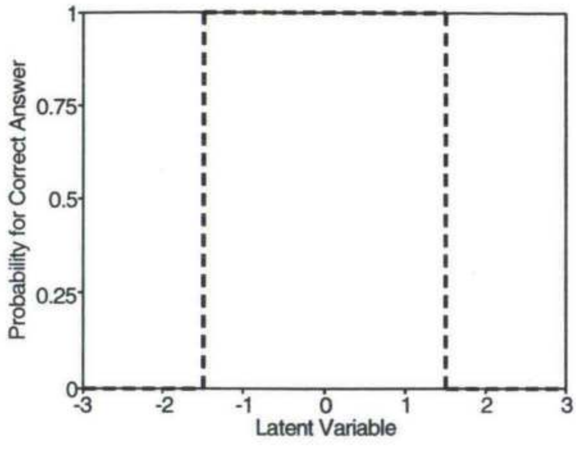


Figure 2.2: Item characteristic curve for a perfectly discriminating non-monotone item

are considered satisfactory, a closer analysis of the scale properties of the items reveals that the items are not in accordance with the Guttman scale type. Clogg and Sawyer (1981) gave an example in which both the coefficient of reproducability and the coefficient of scalability were satisfactory. However, latent class analyses in which models were tested that can be regarded as probabilistic (and therefore less demanding) variants of the Guttman scale, all models had to be rejected according to standard log-likelihood ratio tests. A number of attempts have been made to develop Guttman-like scaling models that would place less severe demands upon the data. Three of these variants will be discussed below.

#### 2.4.1.1 Models with error of measurement

One way to improve the flexibility of the Guttman scale is to allow for measurement error. The idea of a deterministic scaling model is discarded in favor of a probabilistic formalization of the measurement model. This idea can be made clear by looking at the trace lines for such *error-rate* models. In Figure 2.3 a trace line is shown for an item for which errors of measurement are allowed. There is a certain probability that individuals who have a position on the continuum left of

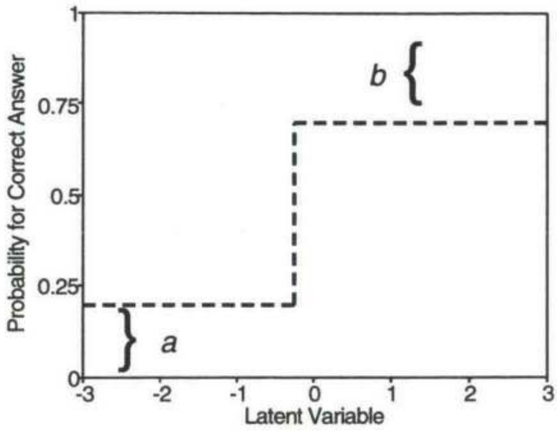


Figure 2.3: Item characteristic curve for an item with measurement error

the threshold value of the item will respond within the positive category, although they should ideally answer with the negative category. This error-rate for falsely giving a positive response is denoted by $a$. If an individual passes the threshold, then the probability of obtaining a positive response increases. But then again it still is possible that an incorrect (so negative) response is recorded. This error-rate for scoring incorrectly with the negative category is denoted by $b$. With this modification of the model the deterministic character of the original Guttman scale is replaced by a probabilistic formulation, which can account for the occurrence of response patterns that deviate from the Guttman pattern. The simplest model arises when it is assumed that both types of error probabilities have the same value (i.e., $a = b$) and that this error-rate is constant over items. This model was proposed by Proctor (1970). The Proctor model can be expressed as a restricted latent class model. A latent class is associated with each legitimate response pattern. The membership of a specific latent class then defines which responses are legitimate and which are false. When the error-rate is equal to $a$, the conditional probability of responding correctly is equal to $(1 - a)$. When the five items are ordered from easiest to most difficult, Table 2.8 shows the equality constraints that ought to be imposed on the conditional response probabilities, according to the Proctor model.

Table 2.8: Conditional response probabilities for the Proctor model

| Latent | Item 4 | | Item 2 | | Item 1 | | Item 3 | | Item 5 | |
|---|---|---|---|---|---|---|---|---|---|---|
| Class | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 |
| 1 | $a$ | $1-a$ | $a$ | $1-a$ | $a$ | $1-a$ | $a$ | $1-a$ | $a$ | $1-a$ |
| 2 | $1-a$ | $a$ | $a$ | $1-a$ | $a$ | $1-a$ | $a$ | $1-a$ | $a$ | $1-a$ |
| 3 | $1-a$ | $a$ | $1-a$ | $a$ | $a$ | $1-a$ | $a$ | $1-a$ | $a$ | $1-a$ |
| 4 | $1-a$ | $a$ | $1-a$ | $a$ | $1-a$ | $a$ | $a$ | $1-a$ | $a$ | $1-a$ |
| 5 | $1-a$ | $a$ | $1-a$ | $a$ | $1-a$ | $a$ | $1-a$ | $a$ | $a$ | $1-a$ |
| 6 | $1-a$ | $a$ | $1-a$ | $a$ | $1-a$ | $a$ | $1-a$ | $a$ | $1-a$ | $a$ |

By using the results stated earlier, this pattern of restrictions on the conditional response probabilities can be translated into restrictions on the log-linear parameters. It can easily be verified that in this case all one-variable terms $u_{j1}$ will be equal to some value, say $c$, for all $j = 1, \ldots, 5$ if dummy coding is used. The two-variable parameters $u_{j1\theta_t}$ will be equal to $-2c$ for every combination of $j$ and $\theta_t$ where a response in category 1 can be considered as correct. For every pair of $\theta_t$ and $j$ where a response in category 1 is not in accordance with the Guttman scale pattern (and so which will result in a conditional response probability equal to $1-a$), the parameters $u_{jg\theta_t}$ are equal to 0. This implies that the probability of giving a correct response is given by:

$$\frac{1}{1+\exp(c)},$$

and the probability of giving a response that is contradictory to the scale pattern is then of course:

$$\frac{\exp(c)}{1+\exp(c)}.$$

As can be seen from the results in Table 2.9, the Proctor model has to be rejected for the male subsample. The estimated error-rate under the Proctor model is equal to 0.13. So, when the Proctor model had fitted the data, there would have been a constant probability of some 13% of scoring in the wrong category of some item, given that one belongs to a certain latent class.

Table 2.9: Testing results for given scaling models with measurement error

| | Pearson $\chi^2$ | $p$ | $G^2$ | $p$ | $df$ |
|---|---|---|---|---|---|
| Proctor's model | 108.29 | .000 | 108.62 | .000 | 25 |
| True-type specific error-rates | 95.99 | .000 | 92.70 | .000 | 21[a] |
| Intrusion-Omission model | 98.78 | .000 | 94.86 | .000 | 24 |
| Item-specific error-rates | 22.16 | .390 | 27.37 | .159 | 21 |
| Latent distance model | 19.88 | .339 | 24.02 | .154 | 18 |

[a]Because one of the estimated probabilities turned out to be zero, that probability was restricted to that value making the number of degrees of freedom for this model 21 instead of 20

The legitimate response patterns corresponding to the various latent classes and their estimated latent proportions in the Proctor model are presented in Table 2.10.

Table 2.10: Estimated latent proportions; Proctor model

| Item 4 | Item 2 | Item 1 | Item 3 | Item 5 | $\hat{p}_{\theta_t}$ |
|---|---|---|---|---|---|
| 0 | 0 | 0 | 0 | 0 | .054 |
| 1 | 0 | 0 | 0 | 0 | .155 |
| 1 | 1 | 0 | 0 | 0 | .241 |
| 1 | 1 | 1 | 0 | 0 | .119 |
| 1 | 1 | 1 | 1 | 0 | .145 |
| 1 | 1 | 1 | 1 | 1 | .285 |

The conclusion that the Proctor model is not acceptable in the example presented raises the question whether it is possible to formulate Guttman-type models that put less severe restrictions on the parameters than Proctor's model does.[26] One possibility that has been proposed by Clogg and Sawyer (1981) is to assume that within latent classes there are constant error-rates, but these error-rates may vary over latent classes. In that case the trace lines are still symmetric, which means that the error-rate left of the threshold value of an item is equal to the error-rate to the right of the threshold. In other words,

[26]For the more complex models for scale response patterns that will be presented next, there is no simple relation between the marginal popularity as denoted in Table 2.1, and the item difficulties, i.e., the location of the items along the latent continuum. Still, because no more information is available, the same ordering of items will be used as in the Proctor model.

the probabilities for both types of error (scoring incorrectly in category 0 and scoring incorrectly in category 1) are equal.

This model has been denoted in Table 2.9 as the *true-type specific error-rates*. This model does not adequately explain the deviations from the deterministic Guttman pattern that are present in the data for the male subpopulation. The reason that both the Proctor model and the *true-type specific error-rates* model fail to fit the data may be due to the fact that both models use error-rates which are invariant across items. When error-rates do vary over items, other more flexible models could possibly fit the data. The most flexible model in this respect is the latent distance model proposed by Lazarsfeld & Henry (1968). This model allows for different error-rates per item. Moreover, the latent distance model allows for asymmetric trace lines, so that the probability of incorrectly responding with category 0 need not be equal to the probability of incorrectly responding with category 1. However, in this most general formulation, the model is not identified. One way to solve this problem is to require the two most extreme items to have symmetric trace lines. In the example presented here this means that only one error-rate will be estimated for item 4 and item 5, whereas for the other three items, two different error rates are estimated. This model fits the data well as can be seen from the results of the statistical tests displayed in Table 2.9. In Table 2.11, the estimated error-rates for the latent distance model are presented together with the estimated latent proportions.

Table 2.11: Estimated error-rates and latent proportions; Latent Distance model

| Scale type | Response Pattern | Item 4 | Item 2 | Item 1 | Item 3 | Item 5 | $\hat{p}_{\theta_t}$ |
|---|---|---|---|---|---|---|---|
| 1 | (0,0,0,0,0) | .028 | .200 | .275 | .066 | .105 | .086 |
| 2 | (1,0,0,0,0) | .028 | .200 | .275 | .066 | .105 | .183 |
| 3 | (1,1,0,0,0) | .028 | .094 | .275 | .066 | .105 | .156 |
| 4 | (1,1,1,0,0) | .028 | .094 | .327 | .066 | .105 | .074 |
| 5 | (1,1,1,1,0) | .028 | .094 | .327 | .135 | .105 | .204 |
| 6 | (1,1,1,1,1) | .028 | .094 | .327 | .135 | .105 | .297 |

One of the most striking results has to do with the high error-rates for item 1. A possible explanation for this result is that item 1 takes a rather average position on the scale when the items are ordered with respect to difficulty, using information from the marginal distributions. It seems justifiable to assume that error-rates will be lower for either very easy or very difficult items.

The restrictions that were proposed by Lazarsfeld and Henry are not the only way to circumvent the identifiability problems. Dayton and MacReady (1976) suggested adapting the latent distance model by preserving the idea of asymmetric trace lines for all the items, but requiring the two types of error probabilities to be equal across the items. The model they proposed thus contains two error-rates: one for the case in which, given the scale type, a respondent incorrectly comes up with a negative response, and the other error-rate pertains to the situation in which the subject erroneously answers with the positive category. This model (the *intrusion-omission model*) does not fit the data according to the values of the test statistics in Table 2.9. This is not very surprising given the rather substantial variation in the estimated error rates under the latent distance model.

Finally, the results will be presented for a model which can also be seen as a restricted latent distance model, but which allows the error-rates for the items to be different. This model, displayed in Table 2.9 under the heading of *item-specific error-rates*, assumes that the trace lines for all items are symmetric, but each item is characterized by its own error-rate. There is one error-rate per item and each item can have a different error-rate. This model fits the data very well. The estimated error-rates as well as the estimated latent proportions are displayed in the following table.

Table 2.12: Estimated error-rates and latent proportions; Item-specific error-rates

| Item | Error-rate | Scale type | $\hat{p}_{\theta_t}$ |
|---|---|---|---|
| 4 | .037 | 1 | .078 |
| 2 | .093 | 2 | .163 |
| 1 | .303 | 3 | .228 |
| 3 | .087 | 4 | .068 |
| 5 | .125 | 5 | .178 |
| | | 6 | .286 |

These results also clearly demonstrate that the first item has rather high error-rates. This simple scaling model, however, gives a very parsimonious description of the data of the male subsample.

#### 2.4.1.2 Models with an intrinsically unscalable class of respondents

The models described in the previous section account for the occurrence of response patterns that do not belong to the Guttman scale types by allowing some error probabilities. Another solution proposed by Goodman is relaxing the assumption of population homogeneity. The basic assumption is that a part of the population responds in line with the requirements of a Guttman scale (i.e., these subjects respond with one of the allowed response patterns, given a specified ordering of the items). The rest of the population is not scalable according to the Guttman model and is therefore denoted as the *intrinsically unscalable class*. This scale model as proposed by Goodman can be translated into a restricted latent class model. For the example presented here, there are six different scale types. The Goodman scale model will thus consist of seven different latent classes. The first six latent classes pertain to the six different Guttman scale types. This means, for example, that the first latent class will only consist of respondents who answered with the response pattern $(0,0,0,0,0)$ and the second latent class is reserved for people whose response pattern is $(1,0,0,0,0)$.[27] This linking of latent classes to Guttman scale types can be done by restricting the conditional response probabilities to be equal to either 0 or 1, depending on the item and the specific latent class. In this way the first six latent classes are linked to the Guttman scale types. The seventh latent class contains the intrinsically unscalable respondents. For this latent class no restrictions are put on the conditional response probabilities, but local independence should hold. Some of the results of this model are summarized in Table 2.13. Again, the data of the male subsample also used in the previous section were analyzed. One of the most striking results is the large estimated proportion of individuals that belong to the unscalable class. It was already observed that in the male sample some 63% of the respondents answered with one of the legitimate Guttman response patterns. So, the observed proportion

[27] The order of the items is according to the item difficulty, as measured by the marginal popularity.

Table 2.13: Conditional response probabilities for the intrinsically unscalable class

| Unscalable | Item 4 | | Item 2 | | Item 1 | | Item 3 | | Item 5 | |
|---|---|---|---|---|---|---|---|---|---|---|
| Class | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 |
| $\hat{p}_{\theta_\iota} = 0.690$ | .083 | .918 | .253 | .747 | .514 | .486 | .559 | .441 | .687 | .313 |

of unscalable respondents is only 37%, while the estimated proportion of unscalable individuals in this Goodman scale is 69%. This can be explained by noting that individuals who belong to the unscalable class can also produce a response pattern that is in accordance with the Guttman model. In the Goodman scale model the estimated proportion of unscalable individuals will always be greater than or equal to the observed proportion. If such a large percentage of the population belongs to the unscalable class, as is the case in our example, one may question the usefulness of this model.

Another point of criticism concerns the implicit assumption that responses to the items are independent for the unscalable class. Clogg & Sawyer have defended this assumption by stating that the unscalable class could be conceived of as consisting of a number of subgroups. When the ordering of the items is different for each subgroup, it could be argued that the responses within the grand total of these subgroups are independent. However, as the results in Table 2.13 clearly show, the ordering of the items that was used to construct the Guttman pattern is also apparent in the conditional probabilities for the unscalable class. Therefore it is questionable whether the assumption of independence within the unscalable class can be justified in this example.

Dayton and MacReady (1980) have criticized the Goodman scale model because even in situations where all respondents are scalable but some small response error is present in the data, the Goodman model yields a high estimated proportion of unscalable individuals. They also mentioned the problem that a test for the hypothesis that the proportion of unscalable respondents is zero makes no sense in the context of the Goodman model, because if this hypothesis is accepted, one is left with a deterministic model which cannot account for the occurrence of non-scale type response patterns.[28] Dayton and MacReady

[28]Dayton and MacReady do not resolve this problem, however. They suggested testing

therefore proposed a model which combines the assumptions underlying the Proctor model or the latent distance model (i.e., error-rates) with an intrinsically unscalable class. However, these kinds of models continue to suffer from the questionable assumption of the existence of an unscalable class whose responses on the items are independent.

### 2.4.1.3 Guttman scaling with Rasch models

The last probabilistic alternative for the deterministic Guttman scale that will be dealt with is the Rasch model. This model will be studied only briefly here because the Rasch model and other latent trait models are the subject of Chapter 3.

Andrich (1985) has made it clear in what sense the Rasch model can be seen as a probabilistic alternative for the Guttman scale. He criticized the attempts to formulate probabilistic Guttman-type models by introducing error-rates or the inclusion of an unscalable latent class. In this way too much weight is given to the manifest properties of the scale and too little attention to the fundamental measurement ideas which underlie the concept of the Guttman model. Andrich stresses that these fundamental ideas as formulated by Guttman are very similar to the methodological point of view taken by Rasch. Guttman demonstrated that the measurement scale should lead to an ordering of individuals which is invariant with respect to different subsets of items. In the same way the ordering of items should be invariant with respect to different subsamples of subjects. These requirements with respect to the measurement scale are similar to the idea of *specific objectivity* formulated by Rasch. Stated briefly, this idea means that scale parameters for items should be independent of the specific sample used, and scale-parameters for individuals have to be independent of the specific set of items used to derive these parameters. The formulation of Rasch is therefore stricter. Guttman only required invariance with regard to the ordering of items and/or individuals, while Rasch formulated this invariance in terms of scale-parameters which are used to character-

---

whether the proportion unscalable respondents is zero by comparing the likelihood ratio for a model with such an unscalable class with the likelihood ratio for a model without such a latent class. As was previously noted, a conditional likelihood ratio test cannot be performed in a situation where the restricted model is formulated by restricting a parameter at a boundary value. The procedure proposed by Dayton and MacReady is therefore not correct.

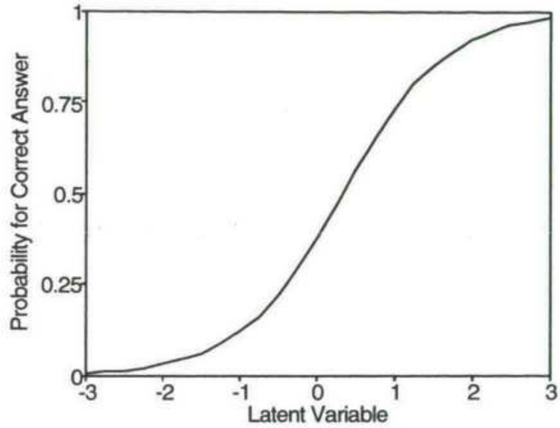


Figure 2.4: Item characteristic curve for the Rasch model

ize both individuals and items. These parameters are used to express a functional relationship between the conditional response probability for an item and the continuous latent scale. The logistic model used by Rasch to formalize this functional form leads to trace lines for the items as shown in Figure 2.4.

The form of trace lines shows that the latent scale is conceived of as a continuous variable. Further, it should be noted that the trace lines for different items are parallel, i.e., trace lines for different items will never intersect. The trace lines for the Rasch model will always be parallel. As will be made clear in Chapter 3, the idea of specific objectivity is translated into statistical terms by the requirement of the presence of sufficient statistics. The total sum of scores for an individual is a sufficient statistic for the latent scale parameter for that individual; at the same time, the total number of positive responses on an item has to be a sufficient statistic for the scale parameter pertaining to that item. The sum of scores plays a central role in the Rasch model. Andrich has pointed out the fact that this same sum of scores is also dominant in the Guttman model because for each scalable subject this sum of scores uniquely defines the response pattern for that subject. Ideas borrowed from the Guttman model are also used by Andrich to develop his *rating scale model*, a model that is in the tradition of Rasch and that is suited for the analysis of polytomous data.

Although Rasch models have been used in the context of the analysis of attitudinal data (see, for example, Duncan, 1984a and 1984b and also Reiser, 1981), it is not clear whether this was done as an attempt to replace the deterministic Guttman scale by a probabilistic alternative.

### 2.4.2 Latent class models for Likert-type data

All latent class models exemplified so far have pertained to dichotomous variables. Clogg (1977 and 1981) has suggested some models for the analysis of polytomous data. To illustrate these models, the data for the five items on women's liberation are analyzed again, but now the original five categories are reduced to three categories as follows:

- category 0: agree entirely or agree
- category 1: don't agree, don't disagree
- category 2: don't agree or don't agree at all

The marginal distributions for the five items and the frequencies with which the $3^5 = 243$ response patterns occur can be found in Appendix B. The data are given for the male and female subsample as well as for the total sample. Again, only those individuals are included that satisfy the educational conditions that were mentioned earlier.

The first model starts from the following assumptions. Since all manifest items have, say, three categories, a model with three latent classes is assumed. Each latent class is associated with one of the response categories. Hence, individuals belonging to the first latent class should respond only in category 0 of each of the manifest variables and individuals in the second latent class should only respond in category 1. Finally, subjects who belong to the third latent class would consequently score in category 2 if there was no measurement error. Measurement error is now introduced to relieve these rather strict assumptions. Individuals are allowed to score not only in their own "true" category, but also in a category that adjoins this "true" category. So, individuals belonging to the first latent class can respond in category 0 (the "true" category for these individuals), but also in category 1. They are, however, not allowed to respond in category 2. The opposite holds for the subjects belonging to the third latent class. They can only come up with a response in category 1 or 2. The probability of responding in category 0 is zero for these individuals. Finally, subjects in the middle latent class can respond in any of the three categories.

This model was tested for the male and female subsample, as well as for the total sample (as was stated before, only respondents that satisfied the educational conditions were included in the analyses).

Table 2.14: Testing results for latent class models for Likert-type data

| | Pearson $\chi^2$ | $p$ | $G^2$ | $p$ | $df$ | $1-\hat{E}$ | $\hat{\lambda}$ |
|---|---|---|---|---|---|---|---|
| Male | 303.81 | .000 | 242.30 | .000 | 221[a] | .936 | .819 |
| Female | 276.68 | .008 | 216.89 | .603 | 223[a] | .933 | .838 |
| Total | 357.25 | .000 | 328.01 | .000 | 221[a] | .938 | .872 |

[a]Degrees of freedom adapted owing to restrictions for parameters of which the estimated values were equal to boundary values

The values for the two test statistics in Table 2.14 are quite different, especially in the case of the male subsample. This is a consequence of the fact that the five-way table analyzed here is quite sparse. Because some of the estimated conditional response probabilities were equal to 1, these parameters were restricted to those boundary values and the number of degrees of freedom was adapted accordingly.

The association between the latent variable and the five manifest indicators is relatively large for both subsamples, as can be seen from the satisfactory values for the percentage of respondents that would be allocated correctly. The estimates for the parameters are given in Table 2.15.

The distribution of individuals over the three latent classes indicates that the women have a slightly more positive attitude concerning women's liberation on the average than the men.

This restricted latent class model thus clearly shows that the ordering of the categories corresponds to an ordering of the latent classes. This ordering is reflected by the relationship between the three latent classes and the conditional response probabilities: the probability of responding in the first category declines as one shifts from the first via the second to the third latent class. The opposite relation can be found for the probability of responding in the third category. Such a clear

Table 2.15: Estimated parameters for latent class models for Likert-type data

| | | Estimated conditional response probabilities | | | | | |
|---|---|---|---|---|---|---|---|
| | | Male sample | | | Female sample | | |
| | | Class 1 | Class 2 | Class 3 | Class 1 | Class 2 | Class 3 |
| Item 1 | Cat. 0 | .752 | .224 | .000[a] | .792 | .260 | .000[a] |
| | Cat. 1 | .248 | .339 | .437 | .208 | .289 | .186 |
| | Cat. 2 | .000[a] | .437 | .723 | .000[a] | .451 | .814 |
| Item 2 | Cat. 0 | .746 | .137 | .000[a] | .608 | .125 | .000[a] |
| | Cat. 1 | .254 | .209 | .044 | .392 | .238 | .000 |
| | Cat. 2 | .000[a] | .655 | .956 | .792 | .637 | 1.000[b] |
| Item 3 | Cat. 0 | 1.000[b] | .448 | .000[a] | .738 | .522 | .000[a] |
| | Cat. 1 | .000 | .241 | .129 | .262 | .232 | .131 |
| | Cat. 2 | .000[a] | .311 | .871 | .000[a] | .246 | .869 |
| Item 4 | Cat. 0 | .445 | .030 | .000[a] | .354 | .038 | .000[a] |
| | Cat. 1 | .555 | .063 | .005 | .646 | .018 | .000 |
| | Cat. 2 | .000[a] | .907 | .995 | .000[a] | .944 | 1.000[b] |
| Item 5 | Cat. 0 | .897 | .691 | .000[a] | 1.000[b] | .406 | .000[a] |
| | Cat. 1 | .103 | .147 | .228 | .000 | .242 | .132 |
| | Cat. 2 | .000[a] | .162 | .772 | .000[a] | .352 | .868 |
| Lat. Prop. | $\hat{p}_{\theta_t}$ | .047 | .649 | .304 | .042 | .586 | .373 |

[a]Restricted a priori

[b]Restricted according to boundary estimate

relationship does not exist for the probability of responding in the second category. The relationship between this probability and the latent variable is not necessarily increasing, decreasing or non-monotone.

Clogg proposed still other latent class models for the analysis of Likert-type data. A second model links the three categories of the manifest variables to three latent classes perfectly by constraining the relevant conditional probabilities to be equal to 1 or 0. A fourth latent class is then added to account for all individuals that cannot be located in one of the first three latent classes. This model is similar to the model presented by Goodman for data which are assumed to follow a Guttman scale. For the example presented in this section this model does not make very much sense. Only three of a total of $3^5 = 243$ response patterns are mapped on the three error-free latent classes. As only a relatively small fraction of all respondents answered with one of these three response patterns, the great majority of the individuals is assigned to the fourth latent class. Therefore this model will not be applied to the data in the present example.

Recently Clogg proposed other latent class models for measurement problems (Clogg, 1988). As these models have a close connection to certain latent trait models, the treatment of these models will be deferred until the next chapter.

### 2.4.3 Ordinal latent class analysis

The restricted latent class models that were reviewed in Sections 2.4.1 and 2.4.2 make use of fixed-value constraints and equality restrictions on the conditional response probabilities. In this section, attention is focussed on ordinal latent class analysis, a technique proposed by Croon (1990). This model links an ordinal latent variable to ordinal manifest indicators by imposing certain *inequality* constraints on the response probabilities. The idea of latent class models for ordered latent classes can be illustrated very easily for the situation with dichotomous manifest indicators.

Suppose there are $T$ different latent classes, numbered $1, \ldots, T$, which are ordered according to some criterion. With a proper ordering of the latent classes, the probability of a "positive" response (which will be indicated with $p_{j1|\theta_t}$) should increase with the class number, so:

$$p_{j1|\theta_{t+1}} \geq p_{j1|\theta_t}.$$

The same idea can be used to model ordered relations between latent and manifest variables for polytomous indicators. However, for polytomous items there are a number of ways to define a "positive response". When an item has $(m+1)$ different response categories, and $g$ and $h$ denote, as before, arbitrary categories, the following set of inequalities with respect to the response probabilities could hold:

$$\sum_{h=g}^{m} p_{jh|\theta_{t+1}} \geq \sum_{h=g}^{m} p_{jh|\theta_t}, \tag{2.38}$$

for $1 \leq t \leq T-1$ and for $1 \leq g \leq m$. Because within each latent class the sum of the conditional response probabilities is equal to one, the restriction can also be written as:

$$\sum_{h=0}^{g-1} p_{jh|\theta_{t+1}} \leq \sum_{h=0}^{g-1} p_{jh|\theta_t}, \tag{2.39}$$

now for $1 \leq t \leq T-1$ and for $0 \leq g \leq m-1$. Equation 2.39 states that the probability for a "negative response" decreases with increasing class number. Combining Equations 2.38 and 2.39 gives the following result:

$$\frac{\sum_{h=g}^{m} p_{jh|\theta_{t+1}}}{\sum_{h=0}^{g-1} p_{jh|\theta_{t+1}}} \geq \frac{\sum_{h=g}^{m} p_{jh|\theta_t}}{\sum_{h=0}^{g-1} p_{jh|\theta_t}}. \tag{2.40}$$

Because the ratio in Equation 2.40 is a monotonically increasing function of class number $t$, the logarithm of this ratio will also increase with increasing class number. These logarithms are called *cumulative logits*. As Croon (1991a) has pointed out, modeling ordinal latent classes with cumulative logits is just one way to handle the problem of polytomous data. Croon, following Agresti (1984), distinguishes four types of logits that could be used:

- the *cumulative* logits

$$\ln\left[\frac{\sum_{h=g+1}^{m_j} p_{jg|\theta_t}}{\sum_{h=0}^{g} p_{jg|\theta_t}}\right], \tag{2.41}$$

  for $g = 0, \ldots, m_j - 1$.

- the *adjacent-categories* logits

$$\ln\left[\frac{p_{jg+1|\theta_t}}{p_{jg|\theta_t}}\right], \tag{2.42}$$

  for $g = 0, \ldots, m_j - 1$.

- the *continuation* logits of the *first type*

$$\ln\left[\frac{p_{jg+1|\theta_t}}{\sum_{h=0}^{g} p_{jg|\theta_t}}\right], \tag{2.43}$$

  for $g = 0, \ldots, m_j - 1$.

- the *continuation* logits of the *second type*

$$\ln\left[\frac{\sum_{h=g+1}^{m_j} p_{jg|\theta_t}}{p_{jg|\theta_t}}\right]. \tag{2.44}$$

It can be proven that the operationalization of the concept of "ordered" relations in terms of these four types of logits will, in general, not lead to the same results (Croon, 1991a). The strongest condition is defined by the adjacent-categories logits. If this condition is satisfied, the response probabilities themselves will be monotonically related to the latent classes. When the set of inequalities implied by the criterion of adjacent-categories logits is fulfilled, sets of inequalities implied by any of the other three criteria will also be satisfied. The weakest condition is given by the cumulative logits. The response probabilities do not have to be monotonically related to the latent classes if the condition of cumulative logits is satisfied. The ratio of the sum of all response probabilities above a certain category to the sum of the remaining response probabilities is, however, a monotonically increasing function of the latent class number. The other two criteria take a medium position between the stronger adjacent-categories logits and the weaker cumulative logits. The two variants of the continuation logits can not be ordered amongst themselves. More details can be found in Croon (1991a).

The latent class model suggested by Croon uses only information concerning the ordering of the categories of the manifest items. Furthermore, the functional relationship between the latent variable and the manifest indicators is not given explicitly (i.e., in algebraic terms). The only requirement about the relation is monotonicity, not an exact functional form such as linearity. In this sense the latent class model developed by Croon is definitely different from other models for the analysis of "ordered data". Many of these other models take into account the ordered character of the data by putting linearizing constraints upon the latent and/or the manifest variables. These linear constraints were already reviewed in Chapter 1. In the next section, latent class models that use this kind of restriction will be studied in more detail. A drawback of these models, as compared to Croon's model, is that they all make explicit algebraic assumptions about the functional relationship between the latent and manifest variables. As a consequence, the latent and/or manifest variables are treated as variables on an interval level. However, these models have fewer parameters than Croon's model, and therefore they offer a more parsimonious description of the latent structure.

### 2.4.4 Latent class models with linear restrictions upon the log-linear parameters

All restricted latent class models that were reviewed so far impose restrictions on the conditional response probabilities. These restrictions can be fixed-value constraints, equality restrictions or inequality restrictions. In this section attention is focussed on latent class models that use restrictions on the log-linear parameters. These models, which were introduced briefly in Chapter 1, will be explored in more depth. In addition, a number of these models will be illustrated by analyzing the data on women's liberation once again.

The log-linear parameterization of the latent class model as given in Equation 2.23, can be used to derive expressions for the conditional response probabilities in terms of log-linear parameters (see Equation 2.26):

$$p_{jg|\theta_t} = \frac{\exp(u_{jg} + u_{jg\theta_t})}{\sum_g \exp(u_{jg} + u_{jg\theta t})}.$$

The one-variable parameters pertaining to category $g$ of the manifest item $j$ are denoted with $u_{jg}$. The two-variable interactions describing the relation between the manifest indicator $j$ and the latent variable $\theta$ are indicated with $u_{jg\theta_t}$. The relation between latent and manifest variables can also be expressed using logits, i.e., the natural logarithm of the ratio of two conditional response probabilities. The logit in which the probabilities for responding in categories $g$ and $g'$ to item $j$ are compared can be written as:

$$\ln\left[\frac{p_{jg|\theta_t}}{p_{jg'|\theta_t}}\right] = (u_{jg} - u_{jg'}) + (u_{jg\theta_t} - u_{jg'\theta_t}). \qquad (2.45)$$

The categories of item $j$ are denoted with successive integers $0, \ldots, g, \ldots, m_j$. Since item $j$ has $(m_j + 1)$ distinct categories, $\theta_t$ $[(m_j + 1) \times m_j]/2$ different logits can be defined for each latent class, in which all possible pairs of response categories are compared. However, for each latent class, there are only $m_j$ non-redundant logits. In Chapter 1 it was suggested that these non-redundant logits be defined by comparing the probability of responding in category $g$ with the probability of responding in category 0; category 0 was defined as being the reference category, i.e., for this category both the one-variable parameters and the two-variable interactions are constrained equal to 0. This simplifies the expressions for the logits to:

$$\ln\left[\frac{p_{jg|\theta_t}}{p_{j0|\theta_t}}\right] = u_{jg} + u_{jg\theta_t}. \tag{2.46}$$

These logits indicate the tendency of responding in category $g$ rather than in category 0 of item $j$. Of course, the value of the logits depends on the latent class. When the number of latent classes is equal to $T$, there are $T$ different logits that can be defined for each category $g$ of item $j$.

It is now possible to formulate restricted latent class models by assuming specific contrasts with respect to these $T$ different logits. If the latent classes are assumed to be ordered properly along a given latent continuum, the $T$ different logits can be hypothesized to be linearly related to the latent variable. This assumption can be formalized by imposing the following restriction on the two-variable log-linear parameters:

$$u_{jg\theta_t} = u^*_{jg} \cdot \theta_t. \tag{2.47}$$

This assumption specifies a linear contrast between the $T$ different logits. By doing so, the latent variable is considered to be measured on a metrical scale. When this linear restriction is applied to the two-variable interactions, a very simple expression for the adjacent-categories logits is obtained:

$$\ln\left[\frac{p_{jg+1|\theta_t}}{p_{jg|\theta_t}}\right] = (u_{jg+1} - u_{jg}) + (u^*_{jg+1} - u^*_{jg}) \cdot \theta_t. \tag{2.48}$$

This expression for the adjacent-categories logits makes several things clear. Firstly, the relation between these logits and the latent variable is linear. The slopes are equal to the difference between successive restricted parameters $u^*_{jg}$, while the intercept is equal to the difference between two successive unrestricted one-variable parameters $u_{jg}$. Secondly, this relation is not necessarily monotonically increasing, because there are no inequality restrictions applied to the parameters $u^*_{jg}$. The relation will be monotonically increasing only if the weights $u^*_{jg}$ increase with category number $g$, i.e., if for each category $g = 0, \ldots, m_j - 1$:

$$u^*_{jg+1} \geq u^*_{jg}.$$

Thus, for some adjacent-categories logits the relation with $\theta$ can be monotonically increasing, while at the same time it is monotonically decreasing for other logits.

In log-linear analysis for manifest variables, there are a number of well-known models with linear restrictions. Depending on whether the row variable or the column variable in a two-way classification is considered as a metrical variable, these models are named *row-association* or *column-association* models. Examples of these models can be found in Goodman (1981, 1983, 1984). These models have not been applied very frequently in the context of latent class analysis, (see, for example, Heinen et al., 1988; Clogg, 1988; and McCutcheon, 1993). The latent class model with the restriction on the log-linear parameters as given in Equation 2.47, is applied to the data on women's liberation later in this section. This model is indicated in the tables reporting results for these analyses as model 1. Firstly, however, some alternative models using linear restrictions will be reviewed.

The restrictions given in Equation 2.47 linearize the relation between the latent variable $\theta$ and the manifest indicators, but the slopes of these linear relations are item- and category-specific. This means that the relation between an adjacent-categories logit and the latent variable will be stronger for one item than for other items. It may be worthwhile to investigate whether the relation between the logits and the latent variable is identical for a particular set of variables. This means that the log-linear two-variable parameters would be restricted for all variables belonging to this set as follows:

$$u_{jg\theta_t} = u_g^* \cdot \theta_t. \tag{2.49}$$

In Table 2.17, it is assumed in model 2 that the parameters $u_g^*$ are invariant for all five manifest indicators.

The two-variable parameters $u_{jg\theta_t}$ can be restricted still further. In model 1, the relation between adjacent-categories logits and the latent variable $\theta$ is linearized by considering $\theta$ as a metrical variable. However, as was stated before, the relation may be monotonically decreasing for some logits because no inequality restrictions are imposed on the parameters $u_{jg}^*$. These parameters are further restricted in model 2 by stating that the logits for a particular pair of categories are identical over (a subset of) items. This means that in model 2 the regression lines formed by regressing the adjacent-categories logits for categories $g$ and $g+1$ on the latent variable have equal slopes for all manifest items (or for all items belonging to the subset). However, within each item the linear relations pertaining to different pairs of categories can, of course, have different slopes.

Model 3 restricts the two-variable parameters by treating both the latent and the manifest variables as metrical variables. The restrictions made in model 3 can be formalized as:

$$u_{jg\theta_t} = u_j^* \cdot g \cdot \theta_t. \tag{2.50}$$

With this restriction, the linear relations for all adjacent-categories logits that can be formed for item $j$ will have equal slopes. The general expression for this adjacent categories logit is given by:

$$\ln\left[\frac{p_{jg+1|\theta_t}}{p_{jg|\theta_t}}\right] = (u_{jg+1} - u_{jg}) + u_j^* \cdot \theta_t. \tag{2.51}$$

Thus, the parameter $u_j^*$ serves as the common slope parameter for all linear relations pertaining to logits for item $j$. Hence, all the "regression" lines will run parallel, and as a consequence, they will all be either monotonically increasing or monotonically decreasing. The indeterminacy at this point, found for models 1 and 2, is eliminated in model 3. In log-linear analysis, models that use restrictions as expressed in Equation 2.50 are called *Uniform Association* models because the relation between the two variables can be described by one single parameter (i.e., $u_j^*$). When the scores $g$ for the manifest variable and $\theta_t$ for the latent variable are equally spaced, this model is equivalent to the *linear-by-linear interaction* model. The two-variable interactions are linearized in both relevant variables (i.e., in the present example item $j$ and the latent variable $\theta$).

Model 4 is basically the same model as model 3, with the additional restriction that the regression coefficient $u_j^*$ is invariant over (a subset of) items. This means that the two-variable parameters are restricted by:

$$u_{jg\theta_t} = u^* \cdot g \cdot \theta_t. \tag{2.52}$$

Model 3 combines the assumption of metrical latent and manifest variables, based on the assumption of an equally strong association between the latent variable and the manifest indicators.

The four models defined by the restrictions given above were tested on the data concerning women's liberation. For this aim, the set of data with three categories, i.e., the data as described in Appendix B, was used. Because it was not possible to decide in advance how many latent classes would be necessary to provide an acceptable fit, the models were estimated with three and four latent classes. All four restricted latent class models considered here required that the metrical values $\theta_t$ for the latent variable be specified a priori. Use was made of equidistant scoring on the interval that is in practice relevant for a standard normal distribution, i.e., the interval within the range $-3$ to $+3$. Hence, three and four points respectively were chosen and equally spaced on the interval $-3$ to $+3$. The scoring system for models using three and four latent classes is reported in Table 2.16.

Table 2.16: Scoring system for three and four latent classes resp.

| $T = 3$ | $T = 4$ |
|---|---|
| $-1.5$ | $-2.1$ |
| $0.0$ | $-0.7$ |
| $+1.5$ | $+0.7$ |
| | $+2.1$ |

For models 1 and 2, the restricted parameters $u^*_{jg}$ resp. $u^*_g$ had to be further restricted in order to make the set of parameters identifiable. This was done by using a dummy-coding scheme in which the parameters pertaining to the first category of the manifest items were set equal to zero, i.e., $u^*_{j0} = 0$ resp. $u^*_0 = 0$.

The models 3 and 4 also required that the numerical category values of the manifest variables also be specified in advance. For the three categories of the manifest items, a simple equidistant scoring system was applied, i.e., the values 0, 1 and 2 were assigned to the three categories.

The results in terms of the test statistics did not differ very much, so the results from the more parsimonious three latent classes analyses were used. The values for the test statistics as well as the number of degrees of freedom for the four models are given in Table 2.17. The results are presented for the total subsample and also for the male and female respondents separately. As before, individuals that did not meet the educational criteria formulated earlier were excluded from the analyses. All computations were performed using Haberman's program

Table 2.17: Testing results for three latent class models with restrictions on the log-linear parameters

| | | Model 1 | Model 2 | Model 3 | Model 4 |
|---|---|---|---|---|---|
| Male | $G^2$ | — | 219.36 | 194.75 | 220.25 |
| | Pearson $\chi^2$ | — | 331.27[a] | 255.71 | 325.82[a] |
| Female | $G^2$ | 163.22 | 200.10 | 193.68 | 204.78 |
| | Pearson $\chi^2$ | 199.64 | 283.23[a] | 256.43 | 292.42[a] |
| Total | $G^2$ | 220.83 | 277.55[a] | 247.86 | 280.19[a] |
| | Pearson $\chi^2$ | 219.44 | 329.91[a] | 276.24[a] | 329.96[a] |
| | *df* | 220 | 228 | 225 | 229 |

[a] $p$-value $< 0.05$

NEWTON. In some cases it was hard to find good initial estimates. Therefore, another program (DILTRAN, a program that will be presented in the next chapter and that uses an EM algorithm) was used to find satisfactory initial estimates for the parameters. The design matrices needed by NEWTON were computed with a special utility program written by Hummelman and Hagenaars called DESMAT (see Hagenaars, 1990 p. 312). Problems were encountered while estimating model 1 for the male subsample. These problems arose because a number of estimated conditional response probabilities approached their boundary values, making the estimated values for the log-linear parameters either too large or too small. These difficulties arose in both analyses (thus with both three and four latent classes). An even more restricted model (i.e., $T = 2$) did not really resolve the problems because latent class model 1 with just two latent classes is identical to an ordinary unrestricted latent class model with two latent classes. The results for that model were already presented in an earlier section. Therefore, no results are available for model 1 in the male subsample.

One of the other problems that developed during these analyses involved the use of test statistics in sparse tables. The problems that arose in this situation were clearly reflected by the discrepancy between the values of the log-likelihood ratio and those of the Pearson statistic. While for the male and the female subsamples separately all models that were tested were acceptable according to the likelihood ratio test, the

Pearson test statistic only supported the first model and, on the edge of significance, model 3. The use of model-selection criteria based on information theory also did not provide decisive indications. Akaike's information criteria supports model 1. The other two criteria ($BIC$ and $CAIC$) favor model 3, and sometimes model 4. This clearly shows that the two modified criteria ($BIC$ and $CAIC$) put higher penalties on using less parsimonious models than $AIC$ does. In summary, all four models seem to be able to reproduce the data to a satisfactory level, but it is difficult to name one that is obviously superior.

A final remarkable result that can be found in Table 2.17 is that the models gave a better fit for the male and female samples separately than for the entire sample. This finding is not entirely surprising as the analysis of the dichotomous data already showed that there was a difference in the difficulty level of the five items between males and females. These differences were also present in the data with three categories, as can be seen from the marginal distributions in Appendix B. A closer examination of the estimated parameters for the two subsamples revealed that the spacing of the categories of some items was clearly different for male then for female respondents. These results as well as the estimated parameters will be inspected more closely in Chapter 3, which deals with latent trait models.

The four models using the linear restrictions introduced earlier have two characteristics in common. They all treat the latent variable as a metrical variable and they all assume a linear relation between this metrical latent variable and the adjacent-categories logits defined for the manifest indicators. The differences between these models lie in whether the manifest variable is also considered a metrical variable, as well as with the possibility of imposing equality restrictions on the slope parameters over different items. Another latent class model that uses linearizing restrictions, but that differs from the four models considered so far, was proposed by Rost (1985, 1988a and 1988b).

Rost's latent class model is an attempt to take into account the ordered information in polytomous manifest items. The model was developed using the threshold approach developed earlier by Masters (1982). In the following it is assumed that the manifest variables have an equal number of response categories. With $(m + 1)$ ordered response categories there are $m$ disparate thresholds between successive categories.

The response process of an arbitrary individual who is confronted with such an item can be described as follows: an individual tries to pass these thresholds, beginning with the threshold separating category 0 from category 1, and so on. When that individual responds in category $g$, he/she is said to have passed $g$ different thresholds.[29] The threshold approach to the response process was suggested by Andrich (1978a). The latent class models developed by Rost lean heavily upon the method used by Andrich. Although Rost's model is a genuine latent class model because it uses a discrete latent variable of which the categories are not necessarily ordered, the model is derived first by assuming a latent continuum reflecting the ability or attitude that a manifest item tries to tap. Each individual occupies a position on this continuum that indicates the tendency for individual $i$ to agree with item $j$. Using Rost's original notation, this tendency is indicated by $\lambda_{j|\theta_i}$. On the same continuum, $m$ different points can be demarcated that correspond to the $m$ distinct thresholds. The value of the threshold between category $(g-1)$ and category $g$ for item $j$ is denoted by $\tau_{jg}$. If the continuum is conceived of as going from "difficult to agree with" to "easy to agree with", an individual must have a relatively large probability of passing the threshold between category $(g-1)$ and category $g$, if $\lambda_{j|\theta_i} > \tau_{jg}$. Also, when $\lambda_{j|\theta_i} < \tau_{jg}$, the probability of individual $i$ passing the threshold between categories $(g-1)$ and $g$ will be relatively small. In other words, the probability of passing the threshold between categories $(g-1)$ and $g$ must be some function of the difference between $\lambda_{j|\theta_i}$ and $\tau_{jg}$.

Following Masters (1982), Rost formalized these ideas by introducing the concept of transition probabilities. Such a probability indicates the likelihood that an individual $i$ will respond to item $j$ with category $g$ rather than with category $(g-1)$. When these transition probabilities are symbolized by $\gamma_{jg|i}$, they can be expressed in terms of the original response probabilities by:

$$\gamma_{jg|i} = \frac{p_{jg|\theta_i}}{p_{j,g-1|\theta_i} + p_{jg|\theta_i}}, \tag{2.53}$$

where $p_{j,g-1|\theta_i}$ indicates the conditional response probability for category

[29] Of course, the implicit assumption that is made here is that the thresholds themselves are also ordered along a given continuum. This problem will be returned to in Chapter 3.

$(g-1)$. Note that the transition probability as defined by Equation 2.53 can be interpreted as the conditional probability that subject $i$ will respond to item $j$ with category $g$, given that item $j$ is responded to with category $g$ or category $(g-1)$. Given this definition of the transition probability $\gamma_{jg|i}$, it can easily be derived that:

$$p_{jg|\theta_i} = p_{j,g-1|\theta i} \cdot \frac{\gamma_{jg|\theta_i}}{1-\gamma_{jg|\theta_i}}. \tag{2.54}$$

From this equation it can be observed that if $\gamma_{jg|\theta_i} > 0.5$, the response probability $p_{jg|\theta_i}$ will be greater than the probability $p_{j,g-1|\theta_i}$, while the reverse is true if $\gamma_{jg|\theta_i} < 0.5$.

Here, as stated before, these transition probabilities should be a function of the difference between the individual $i$'s tendency to agree with item $j$ (so $\lambda_{j|\theta_i}$) and the threshold value $\tau_{jg}$. Rost has proposed modeling the transition probabilities with the logistic function:

$$\gamma_{jg|i} = \frac{\exp(\lambda_{j|\theta_i} - \tau_{jg})}{1+\exp(\lambda_{j|\theta_i} - \tau_{jg})}. \tag{2.55}$$

This logistic function guarantees that $\gamma_{jg|i}$ will be a monotonically increasing function of $\lambda_{j|\theta_i}$ and that $\gamma_{jg|i}$ will never fall outside the interval $[0,1]$ (see Figure 2.4 for an illustration of this logistic function).

The original conditional response probabilities can now also be expressed as a function of the parameters $\lambda_{j|\theta_i}$ and $\tau_{jg}$:

$$p_{jg|\theta_i} = \frac{\exp\left[g \cdot \lambda_{j|\theta_i} - \sum_{x=0}^{g} \tau_{jx}\right]}{\sum_{h=0}^{m} \exp\left[h \cdot \lambda_{j|\theta_i} - \sum_{x=0}^{h} \tau_{jx}\right]}. \tag{2.56}$$

This model can be transformed into a latent class model by imposing a restriction upon the $\lambda_{j|\theta_i}$-parameters. The number of different $\lambda$-parameters is set equal to the number of latent classes and individuals that belong to the same class have the same value $\lambda_{j|\theta_t}$.[30] With this restriction, the expression of the response probabilities in Rost's latent class model becomes:

$$p_{jg|\theta_t} = \frac{\exp\left[g \cdot \lambda_{j|\theta_t} - \sum_{x=0}^{g} \tau_{jx}\right]}{\sum_{h=0}^{m} \exp\left[h \cdot \lambda_{j|\theta_t} - \sum_{x=0}^{h} \tau_{jx}\right]}. \tag{2.57}$$

[30] It is possible to place other restrictions upon the lambda-parameters that lead to other (latent trait) models. This topic will be further explored in Chapter 3.

Some identification restrictions should be imposed on the threshold parameters. This is, however, not of interest for the present discussion.

For this latent class model only one class-specific parameter $\lambda_{j|\theta_t}$ is estimated for each item, while in the unrestricted latent class model, $m$ different independent parameters are estimated per item. Furthermore, a set of item-specific threshold parameters which are the same for all latent classes are estimated.

It can be shown that if the log-linear parameterization of the latent class model is taken as a starting-point, Rost's model can be derived by imposing certain restrictions upon the two-variable interactions. More specifically, these interactions are given by:

$$u_{jg\theta_t} = g \times \lambda_{j|\theta_t}. \tag{2.58}$$

Each two-variable interaction is equal to the product of a class-specific item weight and the category number involved. The fact that the weights are multiplied by the *category numbers* implies that the manifest variables are seen as variables measured on an interval level, as otherwise these restrictions would not be meaningful. On the other hand, no restrictions are put on the class-specific item parameters $\lambda_{j|\theta_t}$, which indicates that the different latent classes can be seen as categories of a variable on a nominal level. The fact that each item has its own set of class-specific difficulty parameters and class-invariant threshold parameters also illustrates clearly that the latent variable is nominal. Rost's model combines a latent variable on a nominal level with manifest items that are assumed to be measured on an interval scale. In this sense the Rost model differs essentially from the four models using linear restrictions, which were presented earlier in this section.

The starting point for these four models (and the linear restrictions that are fundamental to these models) was the idea of expressing the logits for responding in category $g$ rather than in category $g'$ as a function of the log-linear parameters. In the same way, Rost's latent class model can also be reformulated using logits based on the probability that an individual belongs to a particular latent class $\theta_t$ instead of latent class $\theta_{t'}$, given that this individual has responded in category $g$ of item $j$. It can be shown that this logit can be expressed as:

$$\ln\left[\frac{p_{\theta_t|jg}}{p_{\theta_{t'}|jg}}\right] = \text{ intercept } + (\lambda_{j|\theta_t} - \lambda_{j|\theta_{t'}}) \cdot g. \tag{2.59}$$

Thus, this logit is a linear function of the score $g$ on the manifest variable $j$, which is supposed to be a metrical variable. Furthermore, the slope is a function of the restricted two-variable interactions. So far, the analogy with the common base of the four models reviewed earlier in this section is clear. However, the intercept is no longer solely a function of the one-variable terms $u_{\theta_t}$. All one-variable and two-variable log-linear parameters pertaining to manifest items other than item $j$ also appear in the expression of the intercept.

Rost has further elaborated this latent class model by introducing further constraints upon the threshold parameters. Details can be found in Rost (1988a, 1988b).

One of the attractive properties of the Rost model is the fact that no firm assumptions are made with respect to the measurement level of the latent variable. Thus, Rost's latent class model for ordered data is a genuine latent class model in the sense that the latent variable is a nominal variable. Croon's LCA model for ordinal data assumes that the latent classes can be ordered, i.e., that the latent variable is measured on an ordinal scale. It should be noted that the idea of ordered data is operationalized differently in Rost's and Croon's models. As stated before, in the Rost model the transition probabilities are monotonically positively related to the class-specific item parameters $\lambda_{j|\theta_t}$. However, when attention is focussed on the response probabilities instead of the transition probabilities, the latent variable need not necessarily be related to the manifest item in a monotone way in the Rost model. This is, of course, the consequence of considering the latent variable to be a variable on a nominal scale.[31]

The number of parameters that has to be estimated is less than in ordinary unrestricted latent class models. However, the model does not guarantee that the threshold parameters will be ordered properly, although a proper arrangement of the categories is assumed to be present in the development of the model. In this sense, the model has the same deficiencies as model 1. Also, a strong assumption is made concerning the measurement level of the manifest variables. Finally, because of the inherently nominal character of the latent variable in this model, it has

[31]The latent classes can be ordered within one single item according to the estimated $\lambda_{j|\theta_t}$ coefficients. However, this ordering can be different for each manifest item owing to the fact that the latent variable is a nominal variable.

limited usefulness for analyzing sets of data where multiple indicators are used to tap one specific latent variable. It is for this reason that the Rost model cannot be related directly to latent trait models; the four models studied earlier in this section do, however, have direct relations to some well known latent trait models. Therefore, these models are of more importance for the present study than Rost's model.

## 2.5 Latent class models with several latent variables

In all the latent class models presented thus far it is assumed that the relationships between the manifest items can be explained by one discrete latent variable, of which the categories may or may not be ordered. It is possible to generalize the latent class model in order to account for situations in which more than one latent variable is assumed to exist. Goodman (1974a and 1974b) already showed how parameters in latent class models with two latent variables can be estimated. The parameterization chosen by Goodman is in terms of conditional response probabilities. For latent class model with several latent variables, certain restrictions must be imposed on the conditional response probabilities in order to define the relations between the latent variables and the manifest indicators in a correct fashion. This can be illustrated using the example that was presented in Chapter 1 in the discussion of local independence. In the model in this example it is assumed that there are two latent variables $\theta^{(1)}$ and $\theta^{(2)}$ and four manifest items, so $j = 1, \ldots, 4$. Figure 2.5 shows the relations that are assumed to exist between the manifest items and the latent variables.

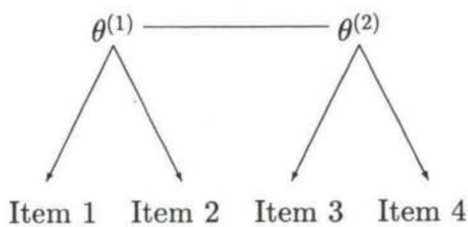


Figure 2.5: Relations between 2 latent and 4 manifest variables

Item 1 and item 2 depend on latent variable $\theta^{(1)}$ and items 3 and 4 depend on latent variable $\theta^{(2)}$. The probability of responding to item 1 or item 2 in a given specific category $g$ is determined by the membership of a given latent class on latent variable $\theta^{(1)}$. The conditional response probabilities for the items 3 and 4 depend only on the latent classes of variable $\theta^{(2)}$. The fact that conditional response probabilities depend on just one latent variable implies that these probabilities must be subjected to certain equality constraints. When an arbitrary latent class that belongs to latent variable $\theta^{(1)}$ is denoted by $r$ and $r = 1, \ldots, T$ and an arbitrary class belonging to $\theta^{(2)}$ is denoted by $s$ and $s = 1, \ldots, T'$, the following set of equality constraints must be imposed on the conditional response probabilities for the example presented here:

$$\text{for } j = 1 \text{ or } j = 2 \quad \text{and} \quad \text{for } r = 1, \ldots, T:$$

$$p_{jg|\theta_r^{(1)};\theta_1^{(2)}} = \ldots = p_{jg|\theta_r^{(1)};\theta_s^{(2)}} = \ldots = p_{jg|\theta_r^{(1)};\theta_{T'}^{(2)}},$$

$$\text{for } j = 3 \text{ or } j = 4 \quad \text{and} \quad \text{for } s = 1, \ldots, T':$$

$$p_{jg|\theta_1^{(1)};\theta_s^{(2)}} = \ldots = p_{jg|\theta_t^{(1)};\theta_s^{(2)}} = \ldots = p_{jg|\theta_T^{(1)};\theta_s^{(2)}}.$$

The first equation states that the probability of responding to item 1 or item 2 with category $g$ is the same for all latent classes $s$ of $\theta^{(2)}$, provided that latent variable $\theta^{(1)}$ is held constant. The second equation states the same for items 3 and 4, but now the roles of the two latent variables are interchanged.

Hence, the problem with just two latent variables can be resolved by combining the two variables $\theta^{(1)}$ and $\theta^{(2)}$ into one new latent variable with $T \times T'$ latent classes. When the equality constraints defined earlier are imposed on this new joint latent variable in the correct manner, the latent class model with two latent variables can be estimated as an ordinary restricted latent class model. This procedure has been implemented in the programs MLLSA and LCAG.

It is also possible to parameterize the model with two latent variables in terms of a log-linear model with some variables that are not observed directly. The Goodman notation with variables as superscripts and categories as subscripts will be used again in order to simplify comparisons with expressions for the log-linear parameterization of the latent class model given earlier in this chapter. The four manifest variables are denoted by $A$, $B$, $C$ and $D$. Categories for these four variables are indicated by $g$, $k$, $\ell$ and $\jmath$ respectively. When the items $A$ and $B$ depend on $\theta^{(1)}$, and items $C$ and $D$ serve as indicators for latent variable $\theta^{(2)}$, the following expression is obtained for the logarithm of the joint latent×manifest proportions in the population:

$$\begin{aligned}\ln p^{ABCD\theta^{(1)}\theta^{(2)}}_{gk\ell\jmath r\ s} &= u + u^{A}_{g} + u^{B}_{k} + u^{C}_{\ell} + u^{D}_{\jmath} + u^{\theta^{(1)}}_{r} + u^{\theta^{(2)}}_{s} \\ &\quad + u^{A\theta^{(1)}}_{gr} + u^{B\theta^{(1)}}_{kr} + u^{C\theta^{(2)}}_{\ell s} + u^{D\theta^{(2)}}_{\jmath s} + u^{\theta^{(1)}\theta^{(2)}}_{r\ s} \quad (2.60)\end{aligned}$$

The latent class models with two latent variables are written out as log-linear models in which the two-variable interactions indicate which manifest items depend on which latent variable. Furthermore, the term $u^{\theta^{(1)}\theta^{(2)}}_{r\ s}$ reflects the strength of the association that exists between the two latent variables.

The latent class model with more than one latent variable presented here can be extended in a number of ways.

Firstly, it is possible to incorporate more than two latent variables. The formulation in terms of log-linear parameters is straightforward. However, if the procedure outlined above is applied, all latent variables are assumed to interact with each other. In terms of the log-linear model, this means that a saturated model is postulated for the marginal table formed by the (unobserved) joint latent frequencies. This is not always attractive. Figure 2.6 shows a latent class model with three latent variables. Each latent variable is related to two manifest indicators. The model assumes that all three latent variables are related to each other, but there is no three-variable interaction. Therefore, an unsaturated model is specified for the latent variables.

Hagenaars (1990) has indicated how latent class models with more than two latent variables for which unsaturated models are defined can be estimated. In the EM algorithm in his LCAG program, this is obtained by fitting a specified unsaturated log-linear model to the joint latent frequencies at the end of each M-step. The expected latent frequencies that result from this unsaturated log-linear model are used as input for the next E-step. A number of latent class models can thus be estimated easily. However, if a specific causal ordering is specified for the latent variables, a modified path-analysis approach may be needed. This method was developed by Goodman (1973b) in the context of log-linear models with directly observed variables. Hagenaars (1988) and Hagenaars et al. (1980) applied this approach to latent class models

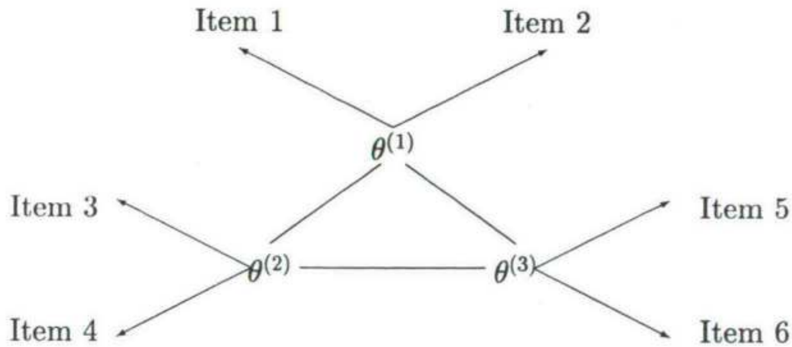


Figure 2.6: Relations between 3 latent and 6 manifest variables

with several latent variables. The details of this procedure are given in Chapter 5.

A second method by which the latent class model with more than one latent variable can be generalized is to allow manifest indicators to be dependent on several latent variables. Figure 2.7 shows three manifest items that are dependent on two latent variables, the second item being influenced by both latent variables.

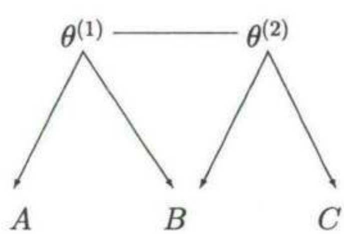


Figure 2.7: Manifest variables dependent on more than one latent variable

How such models can be accounted for within the framework of log-linear models is clear. The log-linear expression of the latent class model in Figure 2.7 is:

$$\begin{aligned}\ln p_{gk\ell r\ s}^{ABC\theta^{(1)}\theta^{(2)}} &= u + u_g^A + u_k^B + u_\ell^C + u_r^{\theta^{(1)}} + u_s^{\theta^{(2)}} \\ &\quad + u_{gr}^{A\theta^{(1)}} + u_{kr}^{B\theta^{(1)}} + u_{ks}^{B\theta^{(2)}} + u_{\ell s}^{C\theta^{(2)}} + u_{r\ s}^{\theta^{(1)}\theta^{(2)}}. \quad (2.61)\end{aligned}$$

This extension can also easily be incorporated into the expression for the conditional response probabilities. For manifest items that depend on two different latent variables, the following holds:

$$p_{jg|\theta_r^{(1)};\theta_s^{(2)}} = \frac{\exp(u_{jg} + u_{jg\theta_r^{(1)}} + u_{jg\theta_s^{(2)}})}{\sum_h \exp(u_{jh} + u_{jh\theta_r^{(1)}} + u_{jh\theta_s^{(2)}})}. \qquad (2.62)$$

The latent class model in which manifest items are dependent on several latent variables is similar to the factor analysis model. Although this type of latent class model can easily be grasped within the framework of log-linear parameterization, it is not as easy to postulate these models using the conceptualization advocated by Goodman, i.e., by using conditional response probabilities. Because this approach (i.e., the parameterization of the latent class model in terms of latent proportions and conditional response probabilities) has become very popular among researchers, the use of latent class models like the one shown in Figure 2.7 has not yet become common practice in applied research. Hagenaars (1990), however, showed how models like these can be estimated using the Goodman parameterization.

The latent class models with two or more latent variables can also be subjected to restrictions on the two-variable parameters in order to treat only the latent variables or both the latent and the manifest variables as variables on an interval scale. The models resuling from this type of restriction are related to the *full-information item factor analysis* developed by Bock et al. (1988). This model and a number of related models will be studied more extensively in Chapter 5. In the next section, attention is given to a problem closely connected with the subject of this section, the relation between latent variables and manifest external variables.

## 2.6 Latent class models with external manifest variables

Although the analysis of measurement models itself deserves special attention, given the great importance of building bridges between theory and observation, in many situations it is desirable to do more than just analyze a measurement structure. It may be necessary to investigate the relation between the theoretically relevant latent variables and a number of other external variables. When these external variables are latent, they become part of the latent structure model and models like the ones that were briefly outlined in the previous section may be used for the analysis of the dependencies or causal relations between the latent variables. If, however, the external variables are manifest, a somewhat different approach is needed. Although this topic is dealt with in depth in Chapter 5, the contours of a strategy that may be relevant for this type of problems is sketched in this section. Two main lines are present in this strategy.

Firstly, one can estimate *latent scores*. In the context of ordinary latent class analysis one can, for instance, use the allocation probabilities to allot individuals to the latent class for which this allocation probability is maximal. When the latent variable is treated as a variable on an interval scale, other possibilities arise. Estimating latent scores has traditionally been one of the chief aims of latent trait theory. The object of this approach is obvious. Once a latent score has been estimated for each individual, these latent scores can be used to investigate the relations between the latent variables and other, external variables. However, intuitively appealing this procedure of estimating latent scores may seem, a number of serious problems may arise with these procedures. In essence these problems are much like those encountered in factor analysis owing to the indeterminacy of factor scores. Hagenaars (1985) addressed this problem as the problem of *identifiability of the individual scores on the latent variable* and showed that, generally speaking, there is no one single method for determining individual level scores on the latent variable using the observed frequencies and the estimated model parameters. It is possible to find different sets of scores on the latent variable that are all in agreement with both the observed frequencies and the estimated model parameters. These different sets may, however, show very different relations with the same external variables. For this reason, it is often better not to investigate the relationship between latent variables and external manifest variables on the individual level, but to incorporate these external variables in the latent structure model.

This leads to a second strategy of analyzing the association between the latent variables and external variables. When log-linear parameterization is used, external variables can easily be incorporated into the latent class model. This bypasses the problem of estimating the latent scores and the accompanying problem of the identifiability of the latent scores on the individual level. A simple illustration of this is given in

Figure 2.8. In this figure, a latent variable is measured through four manifest indicators, denoted as $A$, $B$, $C$ and $D$. The external variable is denoted as $E$.

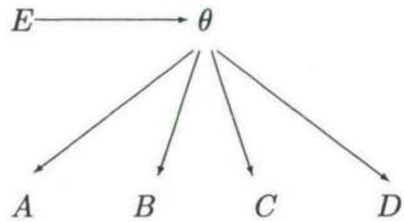


Figure 2.8: Relating an external variable to the latent trait $\theta$

The logarithms of the joint probabilities can be expressed as:

$$\begin{aligned} \ln p_{gk\ell jte}^{ABCD\theta E} &= u + u_g^A + u_k^B + u_\ell^C + u_m^D + u_t^\theta + u_e^E \\ &\quad + u_{gt}^{A\theta} + u_{kt}^{B\theta} + u_{\ell t}^{C\theta} + u_{jt}^{D\theta} + u_{et}^{E\theta}. \end{aligned} \tag{2.63}$$

It should be noted that although $E$ has a conceptually different status than the other manifest variables (i.e., $E$ is not an indicator of $\theta$), the coefficient $u_{et}^{E\theta}$ is technically equivalent to the other two-variable interactions, such as $u_{gt}^{A\theta}$.

An interesting situation occurs when the incorporated external variable is a stratifier or group-defining variable. In that case, the problem takes the form of a simultaneous analysis in several groups. Clogg and Goodman addressed this problem in a series of articles (1984, 1985, 1986). The method proposed by Hagenaars (1988), which is somewhat more general, is based on the idea of introducing quasi-latent variables. These methods are not described in detail here. Instead, a number of models that may be relevant in the context of simultaneous analyses of several groups are indicated. Again, the data on women's liberation is used, but this time the dichotomous variables that were also analyzed in Section 2.3 are employed. In that section the five dichotomous items were subjected to an unrestricted[32] latent class analysis, separately for

[32]The term unrestricted is meant to signify that there were no special sets of equality constraints as in the case of response scale patterns, nor were polynomial restrictions used in order to impose an interval scale on the latent or manifest variables. There were, however, restrictions used to account for estimated parameters that were on the boundary of the parameter space.

the male and female subsample. The results, given that it was a model containing three latent classes, were quite satisfactory, as was reported in Table 2.3.

A model similar to the one shown in Figure 2.8 can now be analyzed for this data. In this latent class model it is assumed that the relations between the variable gender and the five manifest indicators are mediated through the latent variable. In other words, the relations between the latent variable and the manifest items were the same for men and women, but the latent distribution for males may differ from the latent female distribution. This model can easily be tested by adding the variable gender as a sixth indicator. The log-likelihood ratio for this model yielded a value of 83.11 (Pearson $\chi^2 = 80.49$) which, with 43 degrees of freedom, is clearly significant. The postulated hypothesis that the measurement model is the same for the male and the female subsample, therefore, had to be rejected. If the test for this hypothesis had resulted in a non-significant result, a still more restrictive model could have been tested, in which complete independence between the latent variable and the variable gender would have been assumed. This model can also easily be tested by restricting the conditional probabilities for belonging to the male or the female subsample to be equal to the observed marginal proportions.

A less restrictive model can be tested by taking the sum of the values of the log-likelihood ratios and the degrees of freedom. This tests the hypothesis that a latent class model with three latent classes gives an adequate fit to the data for both males and females. However, both the latent distribution and the conditional response probabilities may now vary freely over the two groups (males and females). The sum of the two log-likelihood ratios is equal to 32.67 which, with 32 degrees of freedom, is clearly not significant. The hypothesis that a latent class model with three latent classes can explain the observed pattern of frequency counts for both males and females can be accepted.

Hagenaars (1990) described a number of alternative models that can also be relevant to simultaneous group comparisons. For example, one might consider the hypothesis that the relation between latent and manifest variables is the same for both groups but allow for some direct relations between the grouping variable and the manifest indicators, thereby allowing for the fact that some responses are more popular in one group than they are in another. These more complex models for group comparison lie beyond the scope of this chapter, but they are considered in more detail in Chapter 5.

## 2.7 Evaluation

In this chapter, a large number of different latent class models were reviewed. The latent class model can be parameterized in a number of different ways. The parameterization in terms of conditional response probabilities and latent proportions has the advantage that the parameters being used have a intuitively clear interpretation. It is possible to formulate restricted latent class models by imposing equality or fixed-value restrictions on the conditional response probabilities and/or the latent proportions. This was illustrated by a number of models that are probabilistic versions of the Guttman scale. Some of these models use error-rates to allow for the occurrence of response patterns that are not legitimate according to the Guttman scale model. Other models try to cope with these non-Guttman type response patterns by including an unscalable class.

Models with error-rates have some practical value in the analysis of attitudinal data when the variables are supposed to constitute a cumulative scale. A major problem which remains is that the ordering of the items is often based on the marginal distributions because there is no a priori ordering on theoretical grounds. Moreover, models incorporating error-rates were developed primarily for the analysis of dichotomous data.

Models with an unscalable class do not seem to have much practical value because in many cases the estimated proportion of unscalable individuals is extremely high. Also, the assumption that one segment of the population responds with legitimate response patterns, while the remaining segment responds to the items in a random fashion, does not seem to be very meaningful. More interesting models can be developed by including two unscalable classes, one class of respondents who have a tendency to answer the items positively and one class with high probabilities for negative answers. However, the problem that a relatively large part of the population is essentially unscalable remains.

Clogg (1977, 1981) has proposed some latent class models for Likert-type data that are also based on parameterization in terms of conditional response probabilities. These models have not became very popular, mainly because the restrictions imposed on the response probabilities in these models are very strict. Another more promising method of analyzing these types of data was proposed by Croon (1990). He used certain inequality restrictions on the conditional response probabilities in order to develop an ordinal latent class analysis model.

An alternative parameterization of the latent class model is based on the log-linear model. Using this parameterization, it becomes possible to develop models with linear relations between the latent and the manifest variables. By using these types of restrictions, metrical variables can be included in the latent class models. A number of models in which the latent variable is treated as a metrical variable can be shown to be equivalent to some discretized latent trait models. The log-linear parameterization with linear restrictions can also be used to handle metrical manifest variables. When the latent variables are also metrical, the resulting models can again be related to specific latent trait models.

However, when the latent variable is treated as nominal, as was done in Rost's model, the link between latent class and latent trait models can no longer be established. Although Rost's model was developed using concepts borrowed from item response theory, it is a genuine latent class model and, therefore, it cannot be related to latent trait models in the same way as the other latent class models with linear restrictions. Another consequence of the restrictions that Rost proposed is that in the resulting measurement model the latent variable is measured on a lower level than the manifest indicators. It is not clear how one should justify that the underlying theoretical concept cannot be measured on at least the same level as the manifest variables indicating this theoretical variable. This characteristic also makes Rost's model less attractive for the analysis of attitudinal data than the other latent class models using linear restrictions.

# Chapter 3

# Latent Trait Models

## 3.1 Introduction

The methodology of latent trait models is also called item response theory. First, some basic latent trait models are presented in order to clarify the elementary concepts (Section 3.2). After the fundamentals of latent trait models for dichotomous data and polytomous data have been explained, a short outline will be given of the estimation methods that can be used in the context of latent trait models thus clarifying the terminology necessary for understanding the examples presented in the remainder of this chapter. A more profound and extensive treatment of estimation methods for latent trait models is given in Chapter 4.

A number of interesting latent trait models will be studied in Section 3.3. A typology of latent trait models is presented in Section 3.3.1. The Nominal Response model, proposed by Bock (1972), can be shown to be the most general latent trait model in this typology. This model is studied in detail in Section 3.3.2. Many other known latent trait models can be derived from the Nominal Response model by imposing certain restrictions on the parameters. The two major types of restrictions that can be used will be examined separately in Section 3.3.3 (restrictions on the discrimination parameters) and Section 3.3.4 (restrictions on the difficulty parameters). The various models that are presented in this chapter are illustrated using the same data that was analyzed in the previous chapter with latent class analysis. The parameters in the latent trait models are estimated by discretizing the latent trait. When this estimation method is used, the resulting latent trait models can be shown to be equivalent to certain restricted latent class models that were presented in Chapter 2.

## 3.2 Some basic latent trait models

Latent trait models specify a particular relationship between a latent variable and the manifest responses. Because these models have their roots in the psychometric literature, the latent variable is called a latent *trait*. The relationship between the latent trait and the manifest response is operationalized by modeling the probability that an individual with latent score $\theta_i$ will respond in category $g$ of item $j$. Hence, it is the conditional probability $p_{jg|\theta_i}$ that is modeled in a certain fashion. Up till this part, latent trait models have not been distinguished from the latent class models in the former chapter. In Chapter 1, it was already stated that latent trait models get their identity from the type of variable that is used as a latent trait. The relevant aspects that are often mentioned to illustrate the difference between latent class and latent trait models can be reduced to the difference between the types of latent variables. In item response theory, the latent trait is considered to be a *continuous* variable.

This continuous character of the latent variable has an important consequence. Because the latent variable is continuous and quantitative, the relationship between the latent variable and the probability of responding to a manifest variable with some specific category is modeled using a parametric distribution. For instance, when the observed variables are dichotomous (i.e., a wrong or a correct response can be given to the item), it is common practice to model the probability that an individual with some latent score $\theta_i$ will give the correct answer. This probability is denoted by $p_{j1|\theta_i}$ because the correct answer for item $j$ is scored 1. The probability that individual $i$ will respond to item $j$ with the wrong answer is denoted by $p_{j0|\theta_i}$ as the wrong category is scored as 0. When the latent trait stands for the ability, the probability $p_{j1|\theta_i}$ should increase monotonically for higher latent scores. Simple linear relations between that probability and the latent variable cause problems in many situations because the probability cannot be larger than 1 or lesser than 0.

When both items and individuals are located on the same unidimensional continuum, and the latent position of individual $i$ is denoted by

$\theta_i$, while the latent position of item $j$ is indicated as $b_j$, the probability $p_{j1|\theta_i}$ should increase monotonically by the difference $(\theta_i - b_j)$. This can be accomplished by using certain cumulative distribution functions to relate $p_{j1|\theta_i}$ to the difference $(\theta_i - b_j)$. So, in general:

$$p_{j1|\theta_i} = \Phi(\theta_i - b_j).$$

For the function $\Phi$, the following properties should hold:

$$\begin{aligned} 0 \leq \Phi(u) \leq 1 \quad &\text{for} \quad -\infty \leq u \leq \infty, \\ u_1 \leq u_2 \quad &\Rightarrow \quad \Phi(u_1) \leq \Phi(u_2). \end{aligned}$$

Usually, it is assumed that $\Phi(u)$ is continuous in $u$. It is now possible to use the logistic distribution. The probability $p_{j1|\theta_i}$ is then modeled by

$$\begin{aligned} p_{j1|\theta_i} &= \frac{1}{1 + \exp[-(\theta_i - b_j)]} \\ &= \frac{\exp(\theta_i - b_j)}{1 + \exp(\theta_i - b_j)}. \end{aligned} \tag{3.1}$$

A brief look at the expression for the probability $p_{j1|\theta_i}$ in (3.1) shows a close resemblance with the formulas given for the latent class model in Chapter 2, equation (2.26). The important difference between the two expressions is that in the expression for $p_{j1|\theta_i}$ in (3.1), the quantities $\theta_i$ and $b_j$ lie on a continuous interval scale. The expression, therefore, pertains to a logistic distribution. Equation (2.26) has the same mathematical form, but the argument of the exponential function is no longer a continuous variable. In that case the probability $p_{jg|\theta_t}$ does not follow a cumulative logistic distribution.

This point of difference has been stressed in distinguishing between latent trait models and latent class models. The relationship between the response probabilities on the one hand and the latent variable on the other hand is said to be parameterized in the case of latent trait models, and to be non-parameterized in the case of latent class models. As was stated before, this distinction between the two types of models depends on the difference between the natures of the latent variables.

Having thus made clear the basic distinction between latent class and latent trait models, some attention must be given to latent trait models for dichotomous and polytomous data in order to introduce some basic terminology which will facilitate the explanation of the estimation of parameters.

### 3.2.1 Latent trait models for dichotomous data

Latent trait models for dichotomous data are often denoted by the number of item parameters that are used. In the literature one may encounter one-, two-, three-, or four-parameter models. The one-parameter logistic was presented in the previous section (Equation 3.1). Figure 3.1 shows the item characteristic curve (ICC) or item response function (IRF) for this one-parameter logistic model. The only item pa-

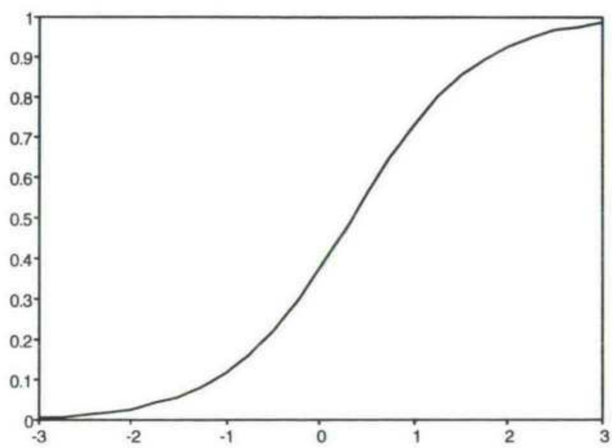


Figure 3.1: Item response function for a one-parameter logistic latent trait model

rameter which appears in Equation 3.1, i.e., $b_j$, is denoted as the item difficulty. This is rather obvious because $b_j$ marks off a point on the latent continuum for which the probability of giving a correct answer is exactly 0.5. Below this point, the probability $p_{j1|\theta_i}$ is less than 0.5 and above that point greater than 0.5. The log-odds for giving the correct answer to the dichotomous item $j$ is equal to

$$\ln \frac{p_{j1|\theta_i}}{p_{j0|\theta_i}} = \theta_i - b_j.$$

The log-odds are a simple linear function of the individual ability and the item difficulty.

The one-parameter latent trait model as described above uses the cumulative logistic distribution function to model the relationship between the latent variable and the response probabilities for the correct answer. As an alternative one could use the normal ogive. This has not been done in the context of the one-parameter model, but theoretically it is possible to do so. The logistic model is more tractable because it is mathematically simpler. Furthermore, parameter estimates obtained with the two functions (i.e., the cumulative logistic and the normal ogive) do not differ much, as was shown by Haberman (1974b).

The one-parameter logistic model can be seen as the simplest, i.e., the most restricted latent trait model for dichotomous data. This model is identical to the model developed by Rasch (1960). Rasch developed this model departing from certain requirements measurement models normally must fulfill. These requirements have become known under the heading *specific objectivity*. This subject was touched upon briefly in the previous chapter; it is elaborated on in Chapter 4. The main idea behind the principle of specific objectivity is that the measurement of individual latent abilities should not depend on the specific items used in a test. Also, the item difficulties should be assessed independent of the specific individuals in the sample. The idea of specific objectivity is worked out technically by requiring that the item response model has known sufficient statistics for both the person and the item parameters.

A restriction of the one-parameter model is that only the difficulty parameters may vary. All items have the same discrimination parameter, hence, the one-parameter model assumes that all items will discriminate in the same way between individuals with different abilities. Because the items do not differ with respect to the discrimination parameter, the trace lines for the various items run parallel, i.e., they never intersect.

The one-parameter model assumes that the relation between the latent variable and the manifest indicators is equally strong for all indicators. In many applications, this is not a realistic assumption. Variability with regard to the way in which items can discriminate can be introduced by incorporating a second item parameter:

$$p_{j1|\theta_i} = \frac{\exp[a_j(\theta_i - b_j]}{1 + \exp[a_j(\theta_i - b_j]}. \tag{3.2}$$

Figure 3.2 shows the trace lines for two items with different discrimination parameters (i.e., $a_j$). The greater the value of the discrimination

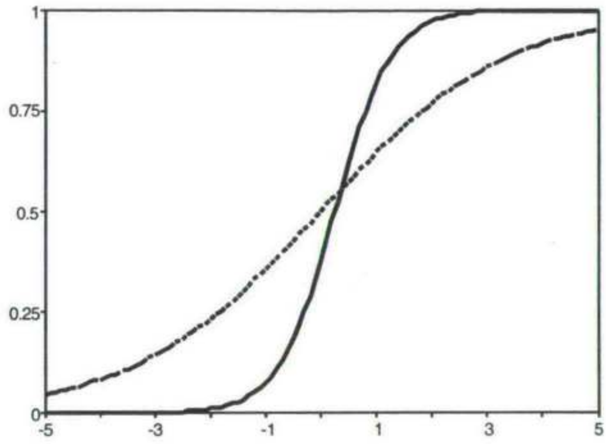


Figure 3.2: Item response functions for the two-parameter logistic latent trait model

parameter $a_j$, the steeper the trace line will be, and, therefore, the more discriminatory power the item will have. The model in Equation 3.2 is a two-parameter logistic latent trait model and it was studied extensively in the chapters written by Birnbaum in Lord and Novick (1968). The two-parameter model was originally formulated using the normal ogive and it has been examined by, among others, Lawley (1943), Tucker (1946), Lord (1952) and Bock and Lieberman (1970). Again, the model using the normal ogive is mathematically more complex, and so the logistic model has become more popular. An advantage of the model based upon the normal ogive is that it allows the parameters of the latent trait model to be interpreted in the context of Thurstonian scaling (Bock and Jones, 1968; Reiser, 1981).

The three- and four-parameter models extend the number of item parameters in order to deal with guessing behavior for low-ability individuals and mistaking behavior for high-ability individuals respectively. The three-parameter model provides one additional parameter as compared to the two-parameter model, thus the left asymptote can be greater than 0. In this fashion it is possible to cope with the eventuality that low-ability individuals will still be able to provide the correct answer by guessing. This is, of course, particularly important when multiple choice items are used to measure an aptitude. The four-parameter model, proposed by McDonald (1967), makes it possible to have a right asymptote which is smaller than 1. The rationale behind this model is that high-ability individuals will not always have a probability of giving the correct answer which approaches 1. It is always possible for them to give a wrong answer owing to errors, carelessness, etc.

The relevance of the three- and four-parameter models for the analysis of ability is clear. It is possible that these models may also be relevant for the analysis of attitudinal data. Certain response categories will be avoided by individuals if, for example, the response to certain items depends on social desirability. Whether the three- and four-parameter models can be used in such situations is, however, a question that falls outside the scope of this study. Therefore, these models will not be discussed any further.

In order to facilitate comparisons between latent trait and latent class models, a slight shift in notation is necessary. The item parameters are denoted in the remainder of this study by $c_j$. This parameter is defined as

$$c_j = -a_j \times b_j,$$

which leads to the following expression for the two-parameter logistic model:

$$p_{j1|\theta_i} = \frac{\exp(a_j \cdot \theta_i + c_j)}{1 + \exp(a_j \cdot \theta_i + c_j)}. \tag{3.3}$$

This short introduction to latent trait models for dichotomous data will suffice for the moment. In the following, attention is focussed on latent trait models for polytomous data.

### 3.2.2 Latent trait models for polytomous data

The purpose of this section is not to describe in great detail all latent trait models that have been proposed for the analysis of polytomous data. Rather, some lines will be drawn in order to categorize the various latent trait models. A number of these models are studied in greater detail in other sections of this chapter. The effort made here to discover some order among the variety of latent trait models leans heavily on Thissen and Steinberg (1986).

A first major point of distinction which can be made is between what Thissen and Steinberg call *difference models* and *divide-by-total models*. Models in the first category are not attempts to model the usual response probabilities (i.e., the probability that individual $i$ will respond in category $g$ of item $j$), but focus instead on modeling the probability of responding in category $g$ or above. In Chapter 1, these probabilities were denoted by $p^*_{jg|i}$. Thus, the cumulative probabilities (or the complements of these cumulative probabilities) take a central position in these difference models. In the context of latent trait models, the most common difference model is the *Graded Response model* proposed by Samejima (1969). The essentials of this model were discussed in Chapter 1. The term *difference models* refers to the fact that the usual response probabilities can be derived by taking the difference between the complements of two cumulative probabilities.

#### 3.2.2.1 Divide-by-total models

Difference models use trace lines for dichotomous items to model the probabilities $p^*_{jg|i}$, i.e., the probabilities of responding in category $g$ or above. The divide-by-total models also use trace lines for binary data, but they do not focus on the cumulative probabilities or their complements, but rather on the conditional probability that an individual is responding n category $g$, given that the response is in category $g$ or category $g-1$. The idea of using these conditional probabilities as the building blocks for a latent trait model for polytomous data was first proposed by Masters (1982). When the probability for responding in category $g$, given that the response is in category $g$ or category $g-1$, is assumed to be governed by a binary logistic trace line and the threshold formulation proposed by Masters is used, the following expression is obtained:

$$\frac{p_{jg|\theta_i}}{p_{j,g-1|\theta_i} + p_{jg|\theta_i}} = \frac{\exp(\theta_i - \delta_{jg})}{1 + \exp(\theta_i - \delta_{jg})}. \tag{3.4}$$

From this expression, a formulation in terms of the response probabilities $p_{jg|\theta_i}$ can readily be derived. This topic will be returned to in Section 3.3.3.2 in which Masters' Partial Credit model is described in more detail. A central role in the Partial Credit model is played by the threshold parameter $\delta_{jg}$. This threshold parameter denotes the value of $\theta$ for which $p_{j,g-1|\theta_i}$ equals $p_{jg|\theta_i}$. In other words, the threshold parameter $\delta_{jg}$ equals the value of $\theta$ for which the category specific curves for categories $g$ and $g-1$ intersect. For individuals with latent scores smaller than $\delta_{jg}$, $p_{j,g-1|\theta_i}$ will be greater than $p_{jg|\theta_i}$. The reverse is true for individuals with latent scores greater than $\delta_{jg}$. For these individuals $p_{jg|\theta_i}$ will be greater than $p_{j,g-1|\theta_i}$.

Masters and Wright (1984) were the first to make it clear that the Partial Credit model can be shown to be a model from which a number of other known latent trait models can be derived by imposing some additional restrictions upon the threshold parameters $\delta_{jg}$. In other words, the principle introduced by Masters (1982), i.e., the idea of using trace lines for binary data to model the conditional probabilities $p_{jg|\theta_i}/(p_{jg-1|\theta_i} + p_{jg|\theta_i})$, defines a whole class of latent trait models. It is the contribution of Thissen and Steinberg to have shown that the Partial Credit model itself can be seen as a special case of the more general Nominal Response model proposed by Bock (1972):

$$p_{jg|\theta_i} = \frac{\exp(a_{jg} \cdot \theta_i + c_{jg})}{\sum_{h=0}^{m_j} \exp(a_{jh} \cdot \theta_i + c_{jh})}. \tag{3.5}$$

The Nominal Response model is the most general model in the class of divide-by-total models. It was briefly introduced in Chapter 1 as an example of a latent trait model with a latent variable on an interval scale and manifest indicators on a nominal level. The item characteristic curves or item response functions resemble those of the Graded Response model. However, in contrast to this Graded Response model, the Nominal Response model leaves open some questions regarding the ordering of the categories of the manifest indicators.

The Nominal Response model is particularly relevant to this study because it has a close link to a restricted latent class model that was introduced in Chapter 2, i.e., the latent class models with linear restrictions. It may be recalled that for this latent class model the two-variable interaction parameters $u_{jgt}$ were restricted by

$$u_{jg\theta_t} = u^*_{jg} \cdot \theta_t,$$

which leads to the following expression for the response probabilities:

$$p_{jg|\theta_t} = \frac{\exp(u^*_{jg} \cdot \theta_t + u_{jg})}{\sum_{h=0}^{m_j} \exp(u^*_{jh} \cdot \theta_t + u_{jh})}. \tag{3.6}$$

When this expression is compared with that of the Nominal Response model, it becomes clear that the two models are very similar, with the important difference that in the Nominal Response model the latent variable is continuous, while this variable is discrete in the linear restricted latent class model. By establishing this link, it becomes easy to gain insight into the connections between the other restricted latent class models described in the previous chapter and the latent trait models which fall under the heading "divide-by-total". The category of divide-by-total models is thus the most closely affiliated to the latent class models.

Within the category of divide-by-total models, there is still one important distinction that must be made. The most general model, the Nominal Response model, and some models that can be derived from this model use both slope parameters (i.e., $a_{jg}$) and parameters that are related to the difficulty of the categories (i.e., $c_{jg}$). Conceptually, these models show a clear resemblance to the two-parameter Birnbaum model for dichotomous data.

Some other models (the Partial Credit models and all models that are derived from this model) use only one type of item parameters, namely, category thresholds (or linear functions of these thresholds). It is possible to bypass slope parameters in these models by (implicitly) assuming that the categories of the manifest indicators are equidistant, i.e., the manifest variables are treated as variables on an interval scale. These models are true divide-by-total models as defined by Masters' (1982). As such these models are closely related to the one-parameter Rasch model for dichotomous data.

The question whether or not latent trait models use discrimination parameters in addition to category parameters is relevant with respect to the estimation procedures that can be applied, as is discussed in the next section.

Thissen and Steinberg (1986) distinguished a third category of latent trait models which they called *left-side added models*. This category contains models which try to correct for guessing by adjusting the left asymptote. This adjustment is really nothing more than adding something to the probability of a correct response at the left side of the latent continuum. Hence the name "left-side added models". The three-parameter model for dichotomous data mentioned in the previous section belongs to this category. This type of correction for guessing can also be carried out in the context of polytomous data. This suggests that the latent trait models for polytomous data presented in this section should be generalized in order to include a left-side added component. The two most widely known examples of this were proposed by Samejima (1969) and by Thissen and Steinberg (1984).

A brief introduction to the estimation procedures that can be used for estimating the parameters in latent trait models is presented in the following section.

### 3.2.3 A brief outline of estimation procedures

The discussion in the previous section made clear that there are two sorts of parameters used in the parameterization of latent trait models: person parameters $\theta_i$ and item parameters (i.e., item difficulties $b_j$, discrimination parameters $a_j$, threshold parameters $\delta_{jg}$, or functions of these threshold parameters $c_{jg}$). The estimation of the parameters used in a specific latent trait model can be performed in various ways.

Firstly, it is possible to estimate both the person parameters and the item parameters simultaneously. This method is called *joint maximum likelihood estimation* (JML). It is not used in this study because this method has some serious shortcomings. This is explained in more detail in Chapter 4.

Another possibility is to estimate only the item parameters using maximum likelihood procedures. This requires the elimination of the person parameters in the likelihood function. There are two methods for eliminating these person parameters. Firstly, there are known sufficient statistics for the person parameters of a number of latent trait models.

By conditioning on these sufficient statistics, a likelihood function is obtained in which only the item parameters appear. This is called *conditional maximum likelihood* (CML). A second method for getting rid of the person parameters is to make some assumptions regarding the distribution of the latent variable $\theta$. If the proper assumptions are made, the person parameters can be integrated out of the likelihood function. This latter method is known as *marginal maximum likelihood.*

Conditional maximum likelihood can only be used if sufficient statistics for the person parameters are known. This is true for latent trait models that do not use discrimination parameters.[1] When the only item parameters that are used are threshold parameters (or functions of these threshold parameters), sufficient statistics for the person parameters are known. Likewise, conditional maximum likelihood can be employed for the Partial Credit model and all models that are derived from this model by imposing additional restrictions on the threshold parameters.

When both threshold parameters and slope parameters are used, these parameters can be estimated using marginal maximum likelihood (MML). In order to apply this method, certain restrictions must be made with respect to the distribution of the latent variable $\theta$. The most far-reaching assumption states that the distribution of $\theta$ is completely known. This is called *parametric MML.* Most examples presented in this chapter use less rigorous assumptions. It is assumed that the distribution of $\theta$ can be approximated by a discrete distribution. The latent probabilities in this discretized distribution can be estimated along with the item parameters. In most cases, it is also assumed that the values of the discretized latent variable (i.e., the latent nodes) are known. This is known as *semi-parametric MML estimation.* For certain models, however, both the values of the categories of the discretized latent distribution and the latent probabilities can be estimated. This procedure is known as *fully semi-parametric MML estimation.*

In all of the examples presented in this chapter, semi-parametric MML with fixed latent nodes was used. Thus, the values of the discretized latent variable were considered known and the latent probabil-

[1] Conditional maximum likelihood can be used in models with discrimination parameters, provided that the values of these parameters are known so that they can be imputed. This is done, for example, in the OPLM program (Verhelst, 1992).

ities were estimated along with the item parameters.

## 3.3 Discretized latent trait models

All discretized latent trait models discussed here belong to what Thissen and Steinberg (1986) called the class of divide-by-total models. A typology of these models is presented in Section 3.3.1. The most general model in this typology is Bock's Nominal Response model. This model is discussed extensively in Section 3.3.2.

Because Bock's model contains slope as well as difficulty parameters, restricted models can be derived by putting certain constraints on these two categories of parameters. After the Nominal Response model has been introduced, a number of models in which only the discrimination parameters are restricted are reviewed (Section 3.3.3). Finally, in Section 3.3.4, a number of models are discussed in which certain assumptions are also made concerning the difficulty parameters. Some of these models deserve attention in their own right; other models are interesting because they allow specific hypotheses to be tested by comparing these restricted models with more general models.

### 3.3.1 A typology of latent trait models

Within the class of divide-by-total models, the Nominal Response model can be seen as the most general model. A number of more specific latent trait models can be derived from the Nominal Response model by imposing certain restrictions on the discrimination and/or the difficulty parameters.

In order to help the reader obtain an overview of the various restricted models that are discussed in this chapter, Figure 3.3 provides a scheme for the hierarchy of all the models surveyed here. For the sake of an orderly presentation, only the working part of the expression for the response probabilities is denoted in Figure 3.3.

It must be stressed that the arrows in Figure 3.3 only indicate the main relations between a number of latent trait models. If there is no arrow between two models, this does not mean that the one model cannot be derived from the other model by imposing certain restrictions. For example, it is possible to derive model 4 from model 2. Yet, these

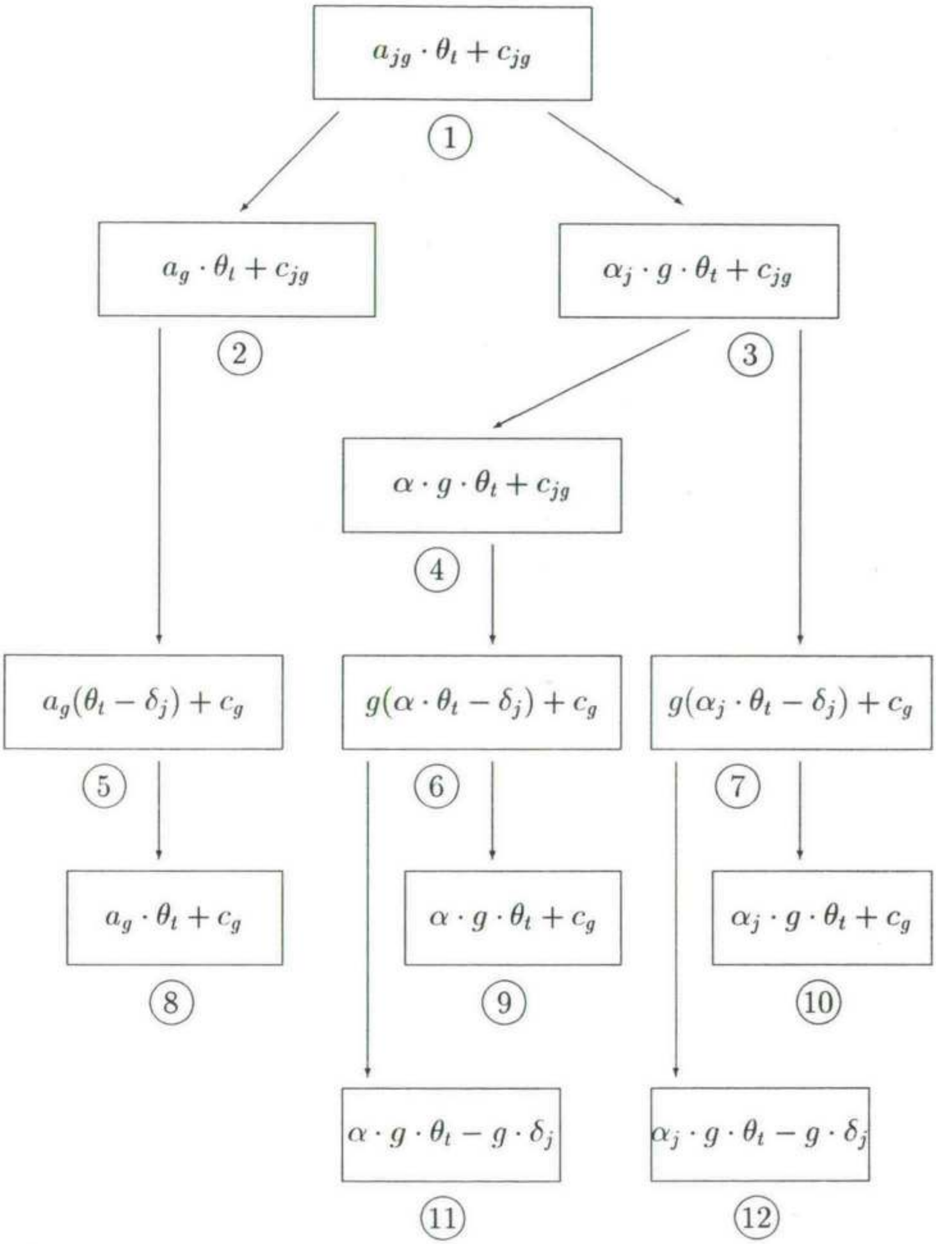


Figure 3.3: Relationships between the latent trait models presented in this chapter

models have not been connected by an arrow in order to indicate that these two models belong to different branches. This will become clear in the following.

**Branch 1: Models with discrimination parameters.** Model 1 is, of course, the Nominal Response model. This model can be restricted along two lines. Firstly, it is possible to impose equality restrictions on the discrimination parameters $a_{jg}$ (model 2). If it is assumed, for example, that the discrimination parameters for items 1 and 2 are equal, the assumption can be denoted by $a_{1g} = a_{2g}$. With such restrictions, a number of characteristics of the Nominal Response model are preserved. The manifest items are not assumed to be measured on an interval scale. There are no assumptions made with respect to the ordering of the categories of the manifest indicators. However, if the categories of the manifest indicators are properly ordered, this will be reflected by the outcome of the analysis. The model itself does not, however, imply a specific ordering. Models 5 and 8 belong to the same branch as model 2. The significance of these models will be discussed at a later point. All of the models that belong to the first branch (thus, models 2, 5 and 8) are parameterized using discrimination parameters that have to be estimated.

**Branch 2 and 3: Linear restrictions on the relation between the latent and manifest variables.** A second line of possible restrictions that can be imposed on the Nominal Response model requires treating the manifest indicators as variables on an interval scale and linearizing the interaction between the indicators on the one hand and the latent variable on the other hand accordingly. These restrictions were also discussed in Chapters 1 and 2. Because both the values of the categories of the manifest items (i.e., $g$) and the values of the latent "node- points" (i.e., $\theta_t$) are fixed in advance, it is necessary to introduce a scaling parameter. This parameter is denoted by $\alpha_j$ when it varies over items. All models that use item-specific scaling parameters belong to the third branch in Figure 3.3, depicted on the right-hand side of Figure 3.3. The use of item-specific scaling parameters (or item-specific association parameters) is the most general method of linearizing the relation between the latent and the manifest variables. When the scaling parameter is assumed to be the same for all items, it is simply denoted by $\alpha$. This leads to the second branch of models in Figure 3.3, represented on the right-hand side of Figure 3.3.

Models 3 and 4 introduce a number of important new characteristics in comparison to models 1 and 2.

1. As was already noted, the manifest variables are treated as variables measured on an interval scale.

2. If both latent and manifest variables are treated as variables on an interval scale, the resulting (discretized) latent trait models are related closely to the Partial Credit model and other latent trait models that can be derived from the Partial Credit model. In the latter group of latent trait models, known sufficient statistics for the ability parameter are available. These models are considered important for two reasons. It is possible to estimate the parameters in these models using CML. Furthermore, models with known sufficient statistics for the ability parameter belong to the Rasch family of latent trait models, which are highly valued by a number of researchers because they satisfy the demand for specific objectivity.

3. Because the indicators are considered interval-level variables, an ordering principle for the categories of the indicators is introduced in model 3. Therefore, models that use linearizing restrictions always consider the categories of the manifest items as ordered, in contrast to models that use estimable discrimination parameters (i.e., models that belong to branch 1).

4. Model 3 and all models that are derived from this model have a number of symmetric properties. These properties are discussed in Section 3.3.3.

5. The $\delta_{jg}$ thresholds that were discussed earlier in the context of the Partial Credit model depend only on the $c_{jg}$ parameters; as will be shown later on, if the latent trait is discretized, these $c_{jg}$ parameters are identical to the log-linear one-variable parameters in latent class models with linear restrictions. For models that use

discrimination parameters and difficulty parameters (such as the Nominal Response model), the thresholds are a function of both types of parameters (thus, both the $a_{jg}$ and the $c_{jg}$ parameters).

**Restrictions on the thresholds.** Models 2, 3 and 4 each define a line along which further restricted models can be developed. Each of these models can be restricted further by introducing a distinction between a "mean" item difficulty and a set of category parameters that is invariant over items. This can easily be illustrated for models 3 and 4.

The conditional response probabilities for models 3 and 4 can be expressed in terms of the $c_{jg}$ parameters (and, of course, the discrimination parameters $\alpha_j$ and $\alpha$ respectively). But it is also possible to express these probabilities in terms of the threshold parameters $\delta_{jg}$. This threshold parameter denotes the value of $\theta$ for which $p_{j,g-1|\theta_i}$ equals $p_{jg|\theta_i}$. The restriction that leads to models 6 and 7, respectively, can be expressed as

$$\delta_{jg} = \delta_j + \tau_g.$$

The threshold is separated in a item difficulty parameter $\delta_j$ and a set of category parameters $\tau_g$ that are constant over items. This restriction implies that the spacing of the thresholds around the mean item difficulty is the same for all items and, thus, depends only on the specific response categories in question. The distance between the threshold for categories $g-1$ and $g$ and the threshold for categories $g$ and $g+1$ is equal to $\tau_{g+1} - \tau_g$, for each item $j$.

As is shown in Section 3.3.3.2, a simple relation exists between the the threshold parameters $\delta_{jg}$ and the $c_{jg}$ parameters for models 3 and 4:

$$\delta_{jg} = c_{j,g-1} - c_{jg}.$$

Using this equation, the restriction in terms of the thresholds can easily be translated into a restriction on the $c_{jg}$ parameters:

$$c_{jg} = c_g - g \cdot \delta_j,$$

in which $c_g$ is defined as

$$c_g = -\sum_{h=1}^{g} \tau_h.$$

Models 6 and 7 are called the Rating Scale model without and with item-specific scaling parameters, respectively. The difference between these two models is that in model 6 the spacing of the thresholds around the item difficulty is exactly the same for all items, while in model 7 this spacing is *proportionally* the same for all items. Model 7, therefore, allows for stretching or shrinking of the latent scale in order to compensate for differences in the discriminatory power of the various items.

A similar restriction can be imposed on model 2. For this model, the values of the thresholds depend on both the $a_g$ and the $c_{jg}$ parameters:

$$\delta_{jg} = \frac{c_{j,g-1} - c_{jg}}{a_g - a_{g-1}}.$$

The restriction in terms of the $c_{jg}$ parameters that leads to a separation of the item difficulty and a set of item-invariant category parameters for model 2 is expressed as

$$c_{jg} = c_g - a_g \cdot \delta_j.$$

This restriction leads to model 5, which is known as the Unidimensional Polychotomous Rasch model. With this restriction on the $c_{jg}$ parameters, the thresholds for model 2 can be rewritten as

$$\delta_{jg} = \frac{c_{g-1} - c_g}{a_g - a_{g-1}} + \delta_j.$$

The first part on the right side of this equation depends only on the categories $g$ and $g-1$. The items differ only in the item difficulties $\delta_j$.

**Equality restrictions with respect to the item difficulties or the category parameters.** Models 5, 6 and 7 can be further restricted in two ways. First, it is possible to test the assumption that certain items are equally difficult, i.e., have the same $\delta_j$ parameter. In that case, the result is models 8, 9 and 10. The most flexible of these three models is model 10 because that model preserves some item-specific information.

Another method of restricting models 5, 6 and 7 is to assume that the $c_g$ parameters are zero.[2] The assumption is that the distance between

[2]Although this type of restriction can be imposed on all three models (i.e.. models 5, 6 and 7), Figure 3.7 shows these restrictions only for the two Rating Scale models, and not for the Unidimensional Polychotomous Rasch model (UPRM), the reason being that the UPRM, restricted in this way, is not implemented in the DILTRAN program.

the thresholds and the mean difficulty is zero, and thus the thresholds all collapse to the point at which $\theta = \delta_j$. This assumption can be of some use when the manifest items have three categories, and there are two thresholds. If these two thresholds collapse, the response probability for the middle category will always be smaller than the probability for one of the other two categories. This is, in other words, a boundary case. When $\tau_1 < \tau_2$ there are some $\theta$-values for which responding in the middle category has the highest probability. When $\tau_1 > \tau_2$, responding in the middle category is always less probable than responding in one of the two extreme categories. $\tau_1 = \tau_2 = 0$ borders between these two situations.

In the following sections, the models presented in Figure 3.3 are studied in more detail. Firstly, the most general item response model in Figure 3.3, i.e., the Nominal Response model, is discussed. Restrictions on the discrimination parameters are examined in Section 3.3.3. This leads to models 2, 3 and 4, which, as noted before, are the parent models for the three branches of latent trait models in Figure 3.3. Finally, in Section 3.3.4 attention is given to restrictions on the $c_{jg}$ parameters.

### 3.3.2 A general latent trait model: the Nominal Response model

Bock developed the Nominal Response model (denoted by model 1 in Figure 3.3) as a latent trait model that can be used when there are a number of alternative response categories to an item and there is no a priori ordering for the "wrong" answers. In Bock's (1972) formulation, the model is based on measurement theoretical ideas developed in the context of choice models (see, for instance, Bock and Jones, 1968 and Luce, 1959). The relationships between these various measurement models will not be studied here. Nor is attention given to Andersen (1983), who developed the same model as Bock in an attempt to formulate a general latent structure model.[3] Instead, attention is focussed

[3]Andersen, however, did not actually analyze data with this model because he preferred a CML procedure for estimating the parameters. The discrimination parameters were only introduced in the theoretical presentation of the model, but they were fixed in his examples for analyzing data.

on the properties of the model by studying the category characteristic curves and the expressions for the conditional response probabilities. To start with the latter, it was already mentioned that the probability that an individual $i$ with latent score $\theta_i$ will respond to item $j$ with category $g$ is written as

$$p_{jg|\theta_i} = \frac{\exp(a_{jg} \cdot \theta_i + c_{jg})}{\sum_{h=0}^{m_j} \exp(a_{jh} \cdot \theta_i + c_{jh})} \text{ with } g = 0, \ldots, m_j. \tag{3.7}$$

The number of categories for item $j$ is equal to $(m_j + 1)$. Note that the notation $\theta_i$ is used to indicate the location of individual $i$ on the latent continuum. When the latent trait is discretized, only certain node-points on this continuum play a role in the formalization of the model and the values of these latent nodes are denoted by $\theta_t$. For estimation purposes, certain constraints must be imposed on both the discrimination parameters and the difficulty parameters. Bock proposed setting the sum of both sets of parameters over the categories equal to 0, i.e., $\sum_h a_{jh} = \sum_h c_{jh} = 0$, but in this study it is preferred to set one of the parameters in each set equal to zero.

Although the Nominal Response model does not require an a priori ordering of the categories, it is assumed here that the categories are ordered in such a manner that the following inequalities hold:

$$a_{j0} < a_{j1} < \ldots < a_{j,m_j-1} < a_{jm_j}. \tag{3.8}$$

It should be noted that this ordering is considered here for merely explanatory purposes. The discussion of the form of the trace lines is greatly facilitated when the categories are ordered such that the inequalities among the discrimination parameters hold. It is clearly not the objective to suggest that the categories of the manifest item are ordered according to substantive criteria, i.e., that the indicators may be measured on a nominal level. Furthermore, it is assumed that the identification restrictions are made for the "lowest" category. As this category is assigned the value 0, the restrictions made are $a_{j0} = c_{j0} = 0$.

When the results of an analysis result in one or more $a_{jg}$ values that are less than 0, the categories can always be rearranged so that the inequalities mentioned above with regard to the $a_{jg}$ parameters will hold. The same remarks can be made with respect to the "highest" category. If, for a given category $g'$, the estimated $a_{jg'}$ parameter is greater than $a_{jm_j}$, category $g'$ should be taken as the "highest" category. In the following, it is assumed that the categories are arranged properly so that the inequalities between the $a_{jg}$ parameters hold.

It can be proven that the category characteristic curve for category 0 is a monotonically decreasing function of $\theta$, whereas the curve for category $m_j$ is monotonically increasing. As will be shown later on, the category characteristic curves for all intermediate categories are unimodal, for there is only one value for which the first derivative of $p_{jg|\theta_i}$ (or equivalently: $\ln p_{jg|\theta_i}$) to $\theta$ is zero. An example of the category characteristic curves for an item with four categories is depicted in Figure 3.4.

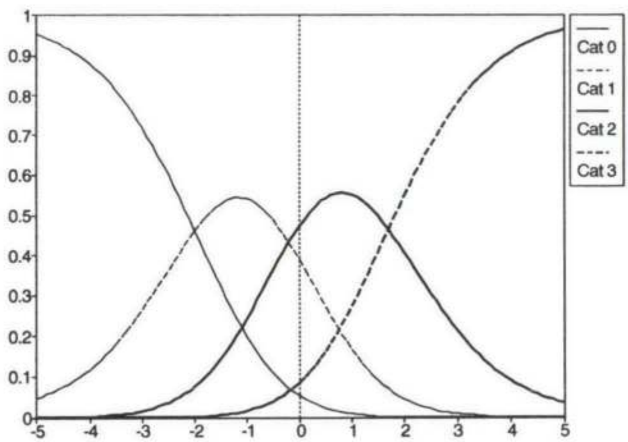


Figure 3.4: Nominal Response model: category characteristic curves

The curves for adjacent categories intersect at the threshold value $\delta_{jg}$. The value of this threshold is, in the case of the Nominal Response model, equal to

$$\delta_{jg} = -\frac{(c_{j,g+1} - c_{jg})}{(a_{j,g+1} - a_{jg})}.$$

Bock notes, with regard to the location of the maxima, that they

> *... are functions of all the parameters of the item and appear difficult to specify ...*

but that is not entirely true. From Equation 3.7 it can be derived that

$$\frac{\partial p_{jg|\theta_i}}{\partial \theta_i} = p_{jg|\theta_i}\left[a_{jg} - \sum_{h=0}^{m_j} a_{jh} \cdot p_{jh|\theta_i}\right].$$

The derivatives of $p_{jg|\theta_i}$ and of $\ln p_{jg|\theta_i}$ are equal to zero for the same $\theta$-values. Because the derivative of $\ln p_{jg|\theta_i}$ is easier to work with, this derivative will be used in the remainder of this section. It can easily be shown that

$$\begin{aligned}\frac{\partial \ln p_{jg|\theta_i}}{\partial \theta_i} &= a_{jg} - \sum_{h=0}^{m_j} a_{jh} \cdot p_{jh|\theta_i} \\ &= a_{jg} - f(\theta_i),\end{aligned} \tag{3.9}$$

where the function $f(\theta_i)$ is defined as

$$f(\theta_i) = \sum_{h=0}^{m_j} a_{jh} \cdot p_{jh|\theta_i}.$$

Because of the assumption that

$$0 = a_{j0} < a_{j1} < \ldots < a_{j,m_j-1} < a_{jm_j},$$

$f(\theta_i)$ will always be greater than or equal to zero. For the derivative of $f(\theta_i)$ to $\theta$, the following results can be obtained:

$$\begin{aligned}f'(\theta_i) &= \sum_h a_{jh} \cdot \frac{\partial p_{jh|\theta_i}}{\partial \theta_i} \\ &= \sum_h a_{jh} \cdot p_{jh|\theta_i}\left(a_{jh} - \sum_{k=0}^{m_j} a_{jk} \cdot p_{jk|\theta_i}\right) \\ &= \sum_h a_{jh}^2 \cdot p_{jh|\theta_i} - \left(\sum_h a_{jh} \cdot p_{jh|\theta_i}\right)^2,\end{aligned}$$

using Jensen's inequality. Thus, the function $f(\theta_i)$ is monotonically increasing.

A number of interesting results can be derived for the category characteristic curves. First, these curves can be examined for the "lowest" and the "highest" category. Afterwards, results can be obtained for the category characteristic curves for the middle categories.

1. When $g = 0$, the discrimination parameter $a_{jo}$ is set equal to zero, resulting in

$$
\begin{aligned}
\frac{\partial \ln p_{j0|\theta_i}}{\partial \theta_i} &= -\sum_h a_{jh} \cdot p_{jh|\theta_i} \\
&= -f(\theta_i) \\
&< 0.
\end{aligned}
$$

   Therefore, the category characteristic curve for $g = 0$ is monotonically decreasing.

   Furthermore, by inspecting the expression for the conditional response probability $p_{j0|\theta_i}$, it can easily be verified that the left asymptote for this curve is equal to one, while the right asymptote is equal to zero.

2. For the highest category (i.e., $g = m_j$), it is true that

$$a_{jm_j} > \sum_h a_{jh} \cdot p_{jh|\theta_i}$$

   because $a_{jm_j}$ is greater than all other discrimination parameters when the categories are ordered properly (i.e., when the inequalities in Equation 3.8 are satisfied). Thus:

$$\frac{\partial \ln p_{jm_j|\theta_i}}{\partial \theta_i} > 0,$$

   from which it can be concluded that the category characteristic curve for category $m_j$ is monotonically increasing. The left asymptote is equal to zero, and the right asymptote is equal to one.

3. For all middle categories, i.e., the categories $g$ for which $0 < g < m_j$, the following three results can be obtained for the first derivative of $\ln p_{jg|\theta_i}$ to $\theta$:

   - The left asymptote of this derivative can be expressed as

$$\lim_{\theta \to -\infty} \frac{\partial \ln p_{jg|\theta_i}}{\partial \theta_i} = a_{jg} > 0.$$

- The expression for the right asymptote of this derivative is

$$\lim_{\theta \to +\infty} \frac{\partial \ln p_{jg|\theta_i}}{\partial \theta_i} = a_{jg} - a_{jm_j} < 0.$$

- The first derivative of $\ln p_{jg|\theta_i}$ to $\theta$ is a monotone decreasing function.

The first of these three results states that the left asymptote of the first derivative of $\ln p_{jg|\theta_i}$ to $\theta$ is positive, while the second result expresses that the right asymptote is negative. Because, as the third result says, this first derivative is a monotonically decreasing function of $\theta$, it can be concluded that there is exactly one value of $\theta$, for which this derivative is equal to zero. The category characteristic curves for the middle categories are, therefore, unimodal.

Furthermore, the modi for the middle categories are ordered in the same fashion as the categories. If the categories are ordered so that the inequalities in Equation 3.8 hold (i.e., $a_{jg} < a_{jg+1}$ for all $g = 0, \ldots m_j - 1$), the modi for the categories 1 through $m_j - 1$ will be ordered along the latent continuum according to the ordering of the $a_{jg}$ values. This can be shown as follows. Suppose that the trace line for category $g$ reaches its maximum for a $\theta$-value equal to $\hat{\theta}_g$, and that the mode for category $(g+1)$ is equal to $\hat{\theta}_{g+1}$, it can be concluded from Equation 3.9 that

$$\begin{aligned} a_{jg} - f(\hat{\theta}_g) &= 0 \\ a_{jg+1} - f(\hat{\theta}_{g+1}) &= 0. \end{aligned}$$

Because the categories are ordered such that

$$a_{jg} < a_{jg+1},$$

it must be true that

$$f(\hat{\theta}_g) < f(\hat{\theta}_{g+1})$$

and because $f(\theta)$ is monotonically increasing:

$$\hat{\theta}_g < \hat{\theta}_{g+1}.$$

The results that were obtained for the first derivative of $\ln p_{jg|\theta_i}$ to $\theta$ can be summarized graphically as is done in Figure 3.5, which displays the functions $(\ln p_{jg|\theta_i})$ for an item with four categories. This figure clearly illustrates that the first derivative for category 0 is always negative, which means that the response probability for category 0 is monotonically decreasing. The reverse is true of category 3. The first derivatives for the two middle categories are zero for only one specific value of $\theta$, indicating that the category characteristic curves are unimodal. Furthermore, the value of $\theta$ for which the first derivative of category 2 is zero is located to the right of the corresponding $\theta$-value for category 1. The two modi for the categories 1 and 2 are, therefore, ordered along the latent continuum.

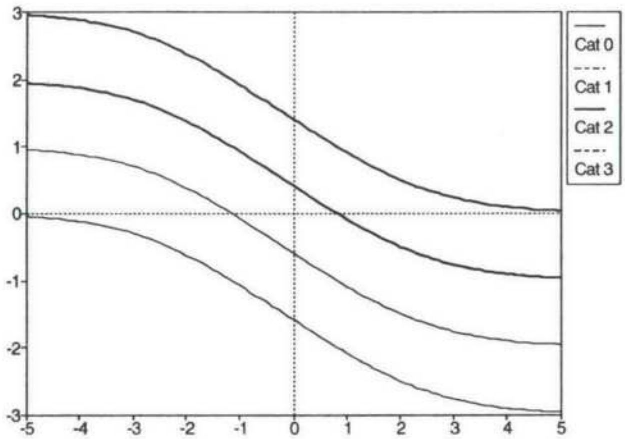


Figure 3.5: Nominal Response model: First derivatives of $\ln p_{jg|\theta_i}$

It must be stressed that the results on the category characteristic curves, as obtained here, depend on a certain ordering of the categories. However, the categories need not be ordered at all. Even when the categories can be ordered on substantive grounds, the estimated $a_{jg}$ parameters may not show the order that would be predicted in advance on the basis of these substantive grounds. In this sense, the Nominal Response model is truly a model for nominal data. However, when the model is used for the analysis of ordered data, the results achieved in this section can be used to facilitate the interpretation of the estimated parameters.

When $m_j = 2$ and, therefore, item $j$ has 3 categories, it is possible to calculate the $\theta$-value for which $p_{j1|\theta_i}$ will be maximum by:

$$\theta\text{max} = \frac{1}{a_{j2}} \left[ \ln \left( \frac{a_{j1}}{a_{j2} - a_{j1}} \right) - c_{j2} \right]. \tag{3.10}$$

When the number of categories is greater than three, there are no closed-form solutions for the $\theta$-values at which $p_{jg|\theta_i}$ takes its maximum.

The $a_{jg}$ parameters not only provide information as to the ordering of the categories, they also bear information with regard to the question how different two response categories are for people with latent scores below or above the threshold value for the two categories. This can be seen by focussing on the adjacent categories logits that were discussed in Chapter 2. For the Nominal Response model these logits can be expressed as

$$\ln \left[ \frac{p_{j,g+1|\theta_i}}{p_{jg|\theta_i}} \right] = (a_{j,g+1} - a_{jg}) \cdot \theta_i + (c_{j,g+1} - c_{jg}). \tag{3.11}$$

The adjacent logits are a linear function of the latent variable $\theta$ and the slope is equal to the difference between the two discrimination parameters. The greater this difference is, the larger the effect will be of a certain shift in the latent position on the adjacent categories logits, and, therefore, the sharper the distinction will be between persons with latent scores below and above the threshold value.

The parameters in the Nominal Response model can only be estimated with CML or fully semi-parametric MML if there are known sufficient statistics for the ability parameters. These sufficient statistics are equal to the weighted total-scores for the subjects, for which the weights are identical to the $a_{jg}$ parameters. These discrimination parameters are supposed to be known and it is necessary to make explicit assumptions with regard to the values of these parameters. An example of this approach can be found in Andersen (1983). Bock (1972) described a JML procedure for estimating the parameters, but currently the MML procedures are more popular.

One possibility is to use parametric MML. A common choice for the latent distribution is to assume a normal distribution function. For attitudinal data, this has the clear disadvantage that the assumption is often not very appropriate with regard to the tails of the population distribution. With attitudinal data, it is not uncommon to register response patterns with relatively low or relatively high total-scores; i.e., individuals clearly object or favor certain points of view. Therefore, semi-parametric MML would appear to be more suited to the analysis of attitudinal data. The five items on women's lib that were analyzed in the previous chapter were again analyzed using semi-parametric MML.[4] These semi-parametric MML analyses yielded the same results as the restricted latent class models that were discussed in Chapter 2, Section 2.4.4. The Nominal Response model with a discretized latent trait is equivalent to the row- or column-association model. The values of the test statistics were identical to the ones obtained in Chapter 2 (see Table 2.17). Again, problems were encountered fitting this model to the data for the male subsample. As was noted in Chapter 2, these problems were caused by the fact that certain estimated conditional response probabilities had reached their boundary values. For the female subsample, the same difficulties arose in estimating the parameters.

It should be noted that when these kinds of problems (i.e., estimated probabilities near 0 or 1) occur and the parameters are estimated using procedures such as Haberman's NEWTON, the algorithm will probably fail to converge. With EM (which is used in DILTRAN), the procedure often converges despite the problems associated with the boundary values. These problems must then be detected by inspecting the estimated conditional response probabilities. In these cases the estimated standard errors of the parameter estimates will also attain unreasonably high values. When EM procedures are used, these standard errors should always be estimated and inspected, because they give very useful information on the identifiability of the parameters. These standard errors can be estimated using the expected information matrix that is described in Appendix A.[5] As is clear from equation A.3, the first derivatives of $p_\nu$ to the model parameters are needed in order to calculate this expected information matrix. These derivatives, for the Nominal Response model as well as for the other latent trait models reviewed in this chapter, are given in Appendix C.

[4] All calculations were performed with the DILTRAN program. See Heinen and Vermaseren (1992).

[5] This is, in fact, the procedure that is followed in the DILTRAN program.

Because of the problems in estimating the parameters of the Nominal Response model for both the female and the male subsample, this model was only tested for the total sample. Obviously, this analysis was not entirely adequate because of the differences between male and female respondents concerning the evaluation of the five items that were clearly present in the analyses in Chapter 2. Fitting the model for the total sample is nonetheless still instructive. Figure 3.6 shows the category characteristic curves for the five items that were obtained from this analysis. The estimated values for the discrimination and difficulty parameters are given in Table 3.1. The standard errors of the estimated parameters are given in parentheses.

Table 3.1: Estimated parameters for the Nominal Response model; semi-parametric estimates; $T = 3$

| | $a_{j1}$ | | $a_{j2}$ | | $c_{j1}$ | | $c_{j2}$ | |
|---|---|---|---|---|---|---|---|---|
| Item 1 | 0.85 | (0.13) | 1.56 | (0.15) | 0.56 | (0.11) | 0.83 | (0.14) |
| Item 2 | 1.09 | (0.22) | 2.83 | (0.34) | 1.07 | (0.22) | 2.21 | (0.24) |
| Item 3 | 1.39 | (0.29) | 2.94 | (0.37) | −0.39 | (0.13) | −0.60 | (0.22) |
| Item 4 | −0.33 | (0.29) | 1.42 | (0.23) | 0.03 | (0.32) | 3.33 | (0.23) |
| Item 5 | 0.99 | (0.16) | 1.93 | (0.18) | −0.81 | (0.12) | −0.62 | (0.16) |

The category characteristic curves clearly indicate that the probability of responding in the middle category is nearly always smaller than the probability of responding in the lowest or highest category. From both the figures and the values of the estimated parameters it is obvious that item 4 is highly unusual. The values of the discrimination parameters do not conform with the inequality restrictions that were discussed earlier because $a_{41} < 0$. This indicates that for item 4 the categories 0 and 1 cannot be ordered. This anomalous pattern can readily be explained by the fact that item 4 had a highly skewed distribution: over 90 % of all respondents fell within the highest response category.

The values of the latent nodes were fixed at $-1.5$, 0 and 1.5. A question that naturally arises is whether this equidistant scoring of the discretized latent trait is adequate and whether a different spacing between the latent nodes would not have led to other results. Before dealing with this question, the estimated latent proportions for the model with three equidistant latent nodes are presented: The distribution of

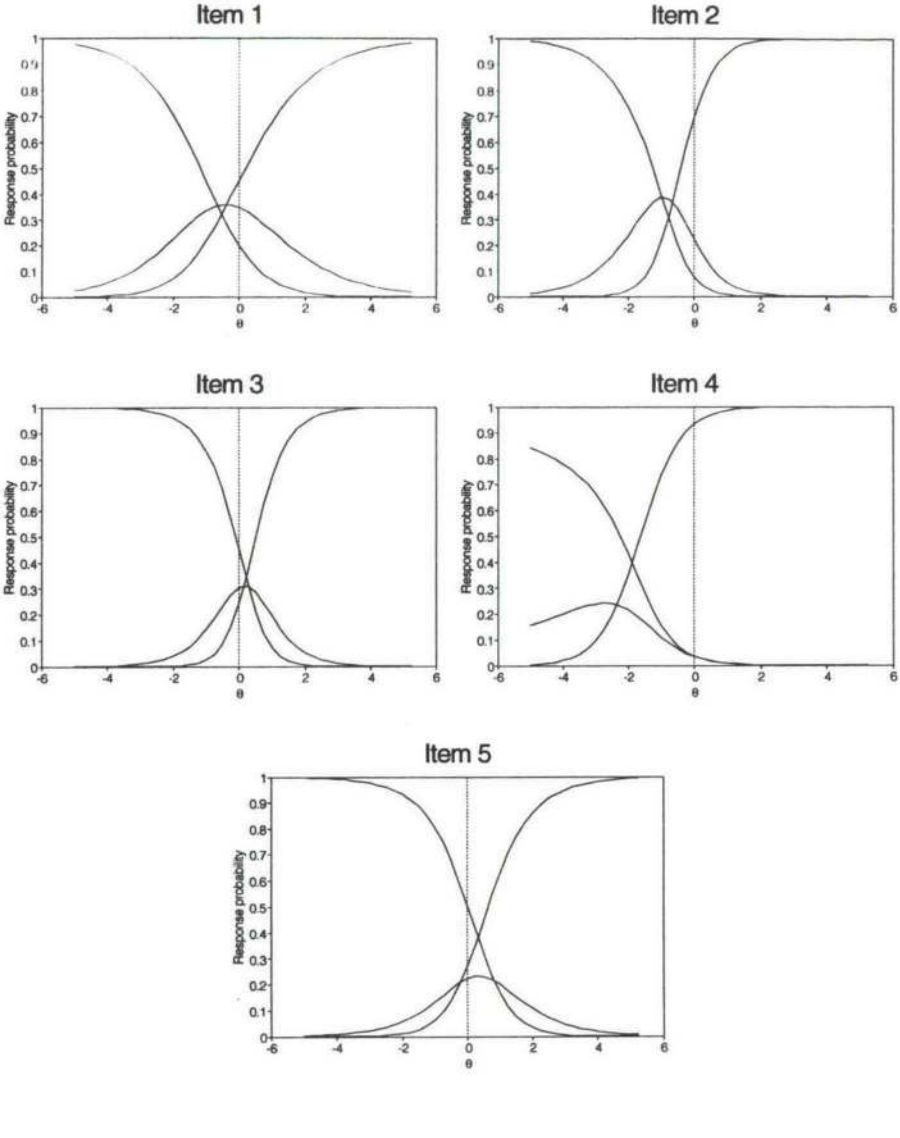


Figure 3.6: Nominal Response model: category characteristic curves for the total sample

Table 3.2: Estimated latent proportions for the Nominal Response model; semi-parametric estimates; $T = 3$

| $\theta_t$ | $p_{\theta_t}$ | |
|---|---|---|
| $-1.5$ | 0.12 | (0.15) |
| 0.0 | 0.45 | (0.08) |
| $+1.5$ | 0.44 | (0.09) |

the latent variable is obviously positively skewed. Because only three latent nodes were employed in this analysis, the choice of the values of the two extreme nodes only served the purpose of fixing the latent scale. The question remains whether the choice for the value of the middle node was satisfactory. In order to obtain information with respect to the problem at hand, the Nominal Response model was tested for the five items with a number of different choices for the middle node-point. The following table gives the values of the log-likelihood ratio statistic and the Pearson statistic for varying choices for the value of the middle node. The results in Table 3.3 show that a minimum value of the log-

Table 3.3: Values for test statistics obtained by different spacing of the latent nodes; $T = 3$

| Value for the middle latent node | $G^2$ | Pearson's $\chi^2$ |
|---|---|---|
| $-0.4$ | 233.41 | 228.77 |
| $-0.3$ | 230.28 | 225.62 |
| $-0.2$ | 226.49 | 222.33 |
| $-0.1$ | 223.18 | 220.15 |
| 0.0 | 220.83 | 219.44 |
| 0.1 | 219.59 | 220.34 |
| 0.2 | 219.52 | 222.71 |
| 0.3 | 220.60 | 226.31 |
| 0.4 | 222.76 | 230.93 |

likelihood ratio statistic is reached when the value of the middle latent node point is fixed at about 0.20. Varying values in the neighborhood of 0.20 made clear that the minimum value for the log-likelihood ratio was 219.40 and that this value would be reached when the middle latent class was given a value of 0.16. The numbers presented in the table also show that minimizing the Pearson statistic leads to another choice for the value of this latent node, but within the framework of maximum likelihood estimation, minimizing $G^2$ is the obvious choice. The fact that the values of Pearson $\chi^2$ and $G^2$ depend on the choice of the value of the middle latent node raises the question whether it is possible to estimate the value of this latent node. Further research on this topic is needed.

When the value of the middle latent node is fixed at 0.16, the spacing of the latent nodes is in concordance with the positively skewed latent distribution that was estimated in the preliminary analysis. It would appear from this example that the spacing of the latent nodes is optimal when it conforms to the global shape of the latent distribution. Starting with equidistant spacing is, therefore, an obvious choice because with equidistant spacing no a priori assumptions are made regarding the form of the latent distribution. The results in Table 3.3 also indicate that choosing a spacing for the latent nodes that does not correspond to the estimated skewness in the latent distribution leads to values for the test statistics that are too high. Of course, this discussion of the choice of an appropriate spacing of the latent nodes is greatly facilitated by the fact that there are only three nodes. When the number of latent nodes is greater than three, searching for an optimal spacing is more intricate. It is also not clear whether the results reported here can be generalized. In other words, is it always advisable to choose a spacing for the latent nodes that corresponds to the global shape of the estimated latent distribution? This topic also needs further study.

Finally, it should be noted that a slight shift in the spacing of the latent nodes, leads to changes in the estimates of the difficulty parameters, so the values for the $c_{jg}$ parameters reported in Table 3.1 should be interpreted with care. Despite these changes in the estimated values for the difficulty parameters, the category characteristic curves are not greatly affected by a change in the spacing of the latent nodes. Changing the value of the middle node from 0.0 to 0.16 in the example presented above has the impact of shifting the majority of the category characteristic curves a little bit to the right. The basic shapes for all the curves are retained, however.

When the manifest items are all dichotomous, the Nominal Response model is equivalent to the Lord-Birnbaum model. It proved interesting to investigate this model using the dichotomous data. Because the model was introduced here merely for explanatory purposes, only the data for the male subsample were analyzed. Firstly, the parameters of the Lord-Birnbaum model were estimated using the semi-parametric MML for models with the number of latent nodes ranging from 2 through 5. The values of the latent nodes were spaced equidistantly on the interval from $-3$ to $+3$. Table 3.4 presents the values of the $G^2$ and the Pearson $\chi^2$-statistic along with the significance levels and the number of degrees of freedom.

Table 3.4: Values for test statistics for the male subsample; Lord-Birnbaum model

| Number of latent nodes | $G^2$ | $p$ | Pearson's $\chi^2$ | $p$ | $df$ |
|---|---|---|---|---|---|
| $T = 2$ | 42.20 | 0.003 | 40.65 | 0.004 | 20 |
| $T = 3$ | 21.77 | 0.296 | 21.75 | 0.297 | 19 |
| $T = 4$ | 16.79 | 0.537 | 15.84 | 0.604 | 18 |
| $T = 5$ | 16.81 | 0.467 | 15.79 | 0.539 | 17 |

With two latent nodes, the results were identical to those obtained with an unrestricted latent class analysis (see Table 2.3 in the previous chapter). The fit increased rapidly for three and four latent nodes. With $T = 5$, however, the fit could not be improved. Moreover, with five latent classes the standard errors for some estimated $\hat{p}_{\theta_t}$ parameters were very large. These results seem to indicate that the model with $T = 5$ is not identified. This suspicion was supported by inspecting the estimated expected information matrix. As was noted in Chapter 2, a sufficient condition for local identifiability is that this matrix is of full rank. The DILTRAN program was used to calculate the eigenvalues of this matrix. Because of numerical rounding errors, some of these eigenvalues are not exactly equal to 0 when the model is not identified, but at least one eigenvalue is very small. This was indeed the case for the analysis with $T = 5$.

A more restricted model can be tested by imposing equality restrictions on the slope parameters. The resulting model is a Rasch model with a discretized latent variable. The hypothesis that all five discrimination parameters are equal can be tested by taking the difference in the $G^2$ values for this Rasch model and the Lord-Birnbaum model with $T = 4$. This difference was equal to 26.19 which, with four degrees of freedom[6] is significant to the 0.1%-level. Thus, at least one slope parameter was different from the others. Table 3.5 presents the values of the estimated parameters for the Lord-Birnbaum model. From these

Table 3.5: Estimated parameters for the Lord-Birnbaum model in the male subsample; $T = 4$

| | $a_j$ | | $c_j$ | | Difficulty=$(-c_j/a_j)$ |
|---|---|---|---|---|---|
| Item 1 | 0.87 | (0.18) | −0.95 | (0.30) | +1.10 |
| Item 2 | 1.35 | (0.32) | −0.29 | (0.45) | +0.22 |
| Item 3 | 2.99 | (0.72) | −3.71 | (0.68) | +1.24 |
| Item 4 | 1.10 | (0.30) | +1.55 | (0.42) | −1.41 |
| Item 5 | 1.50 | (0.25) | −2.66 | (0.51) | +1.77 |

results it is immediately obvious that item 1 is the least discriminatory, while item 3 has the greatest discriminatory power. The other three items have estimated slope parameters that do not differ much. This can also be seen in a graphic display of the item characteristic curves as shown in Figure 3.7. Besides the estimated values for the $a_j$ and the

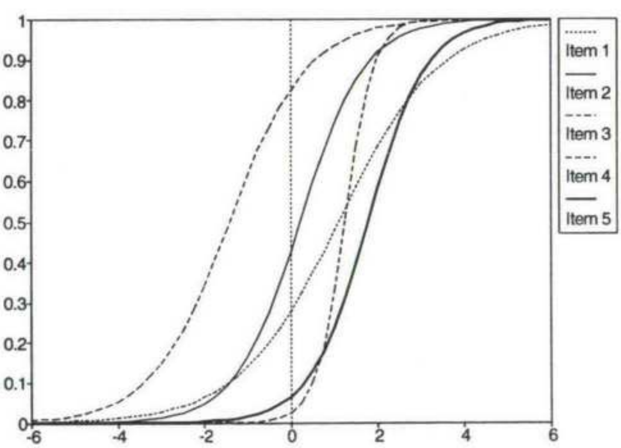


Figure 3.7: Lord-Birnbaum model: item characteristic curves for the male subsample

$c_j$ parameters, Table 3.5 also gives values for the item difficulties. The

[6]Instead of one common slope parameter, there are four additional discrimination parameters in the Lord-Birnbaum model that have to be estimated.

value of $\theta$ at which the probability of answering in category 1 changes from less than 0.5 to more than 0.5 is equal to $(-c_j/a_j)$. The values for these estimated item difficulties show the same ordering as the percentages responding in category 1 of the five items in Table 2.1.

The estimated (discretized) latent distribution can be characterized by looking at the estimated latent proportions. These results are given in Table 3.6. This estimated latent distribution again shows that the

Table 3.6: Estimated latent proportions for the Lord-Birnbaum model; male subsample

| $\theta_t$ | $p_{\theta_t}$ | |
|---|---|---|
| −2.1 | 0.08 | (0.50) |
| −0.7 | 0.01 | (0.47) |
| +0.7 | 0.50 | (0.12) |
| +2.1 | 0.41 | (0.17) |

distribution of the latent variable is negatively skewed. Once more the question can be raised whether the fit of the model can be improved by choosing the values of the latent nodes in accordance with this estimated skewness. With equidistantly spaced latent nodes, the value for $G^2$ was 16.79. By varying the values of the latent classes on the right-hand side of the continuum, this $G^2$ value could only be diminished by less than 0.1 point. There would seem to be little gain in fit by choosing other values for the latent nodes. However, when a number of latent nodes were concentrated on the left side of the latent continuum (which is discordant with the idea of a negatively skewed distribution), the model fit deteriorated rapidly. Again, it is clear that a uniformly spaced set of latent nodes is a good starting point for the analysis, and that when the values of the latent nodes are not chosen equidistantly spaced, the spacing used should be congruent with the estimated skewness for the latent distribution.

The lack of fit of the Rasch model with the data for the male subsample can thus be explained by noting that both item 1 and item 3 had different slope parameters than the other three items. To test this more specific hypothesis, a model was tested in which the Lord-Birnbaum model was specified for the items 1 and 3, while for the other three items, it is assumed that they are characterized by a common slope parameter. The results for this model were $G^2 = 17.83$, $df = 20$ and $p = 0.599$. Again, four equidistant latent nodes were used. This mixed model provided an excellent fit. The values for $G^2$ and the number of degrees of freedom that pertain to the three different models that were tested against the dichotomous data for the male subsample are given in Table 3.7 in order to facilitate the comparison between these models. Because these three models are hierarchically nested, differences

Table 3.7: Values for test statistics for the male subsample; three different models; $T = 4$

| | $G^2$ | $df$ |
|---|---|---|
| Rasch model | 42.98 | 22 |
| Mixed model | 17.83 | 20 |
| Lord-Birnbaum model | 16.79 | 18 |

between the three log-likelihood ratios can be used to test specific hypotheses. The difference between the Lord-Birnbaum model and the mixed model shows that the restriction that the items 2, 4 and 5 have a common slope does not lead to a significantly worse fit. The gain of two degrees of freedom amply compensates for the increase of the log-likelihood by 1.04 points. However, the assumption that all five items were characterized by the same discrimination parameter (which is the discretized Rasch model) led to a further increase of the log-likelihood ratio by 25.15 points which is very significant with two degrees of freedom. The mixed model is, therefore, the most parsimonious model, with an acceptable fit for the dichotomous data in the male sample.

### 3.3.3 Restrictions on the discrimination parameters

Three types of restrictions with regard to the discrimination parameters are surveyed. The first type of restriction is very straightforward. In the Nominal Response model, each item is characterized by both discrimination and difficulty parameters. Both sets of parameters are category specific and are allowed to vary over items. When the results in Table 3.1 are inspected closely, it is clear that for some items the estimated discrimination parameters are very much alike. Thus, a more parsimonious model can be obtained by restricting some sets of category-specific discrimination parameters to be equal. This leads to model 2 in Figure 3.3.

Secondly, the Partial Credit model, which was presented somewhat superficially in previous sections and chapters, is dealt with in greater detail below. In this model it is assumed that the categories of the manifest items are equally spaced and so the manifest indicators are assumed to be measured on an interval scale. The parameters for this model can be estimated by, for example, CML or semi-parametric MML methods. If the latter procedure is employed and the values of the latent nodes are fixed, a scaling parameter has to be introduced. The Partial Credit model parameterized by such a scaling parameter is denoted by model 4 in Figure 3.3.

In this section a third type of restricted model can be generated by letting this scaling parameter vary over items (model 3 in Figure 3.3).

#### 3.3.3.1 Discrimination parameters restricted to be equal over items

In the foregoing, the Lord-Birnbaum model was analyzed using the data from the male subsample. This model for dichotomous data uses two parameters, a discrimination parameter and a difficulty parameter, for each item. When the discrimination parameters are restricted to be equal over items, the resulting model is the Rasch model. It is also possible to specify models in which some but not all discrimination parameters are restricted to be equal.

The same restrictions can be applied to the Nominal Response model. The parameters in this restricted model can be estimated by semi-parametric MML with fixed latent nodes. This model is illustrated using the data from the total sample, the same data that were used to examine the Nominal Response model. The estimated parameters from the Nominal Response model were reported in Table 3.1. It is clear from these results that the estimated $a_{jg}$ parameters for items 2 and 3 were very much the same. The same can be said for the items 1 and 5. As was noted earlier, item 4 had an aberrant pattern of estimated $a_{jg}$ parameters.

Table 3.8 presents the results for a number of models in which particular sets of discrimination parameters were restricted to be equal. Just as with the Nominal Response model, three latent nodes were used. To

Table 3.8: Semi-parametric MML with fixed nodes; Nominal Response model for the total sample with equality restrictions on the $a_{jg}$ parameters

| Items for which $a_{jg}$ are set equal | $G^2$ | $p$ | $df$ |
|---|---|---|---|
| — | 220.83 | 0.472 | 220 |
| 2 3 | 222.00 | 0.487 | 222 |
| 1 5 | 223.92 | 0.451 | 220 |
| 2 3 5 | 233.32 | 0.321 | 224 |
| all items | 277.55 | 0.014 | 228 |

facilitate the comparison between the different models presented in Table 3.8, the results for the Nominal Response model that were obtained earlier in this chapter are also included. The preliminary conclusions that were drawn from the results reported in Table 3.1 can now be based on statistical grounds. The value of the log-likelihood ratio increased just slightly when the $a_{jg}$ parameters for the items 2 and 3 were restricted to be equal. The same conclusion is reached when the discrimination parameters for items 1 and 5 are set equal. The difference between the $G^2$ values for these two models with the $G^2$ value for the Nominal Response model are not significant. All other pairs of items resulted in higher values for the $G^2$ when the discrimination parameters were set equal. Not surprisingly, this was in particular the case for pairs of items in which item 4 was involved. When the parameters for three items were restricted to be equal, the optimal set consisted of items 2, 3 and 5. All other sets of three items yielded $G^2$ values that were much higher. However, the difference between the value of the log-likelihood ratio for the specified model (i.e., the model in which the $a_{jg}$ parameters were set equal for items 2, 3 and 5) on the one hand and the $G^2$ for the Nominal Response model on the other hand, was significant at the 5% level. When the discrimination parameters for all five items were set equal, the model did not fit the data very well. This model is equivalent to a latent class model that was presented in Section 2.4.4 and which was denoted there by "model 2". The value for the log-likelihood ratio in Table 2.17 for model 2 is therefore equal to the value in Table 3.8.

Thus far, the data from the total sample were used to illustrate results for the Nominal Response model with and without equality restrictions upon the discrimination parameters. The reason for this is that the parameters in the Nominal response model could not be estimated for the two subsamples, i.e., the male and the female subsample. It is possible to solve such problems by fixing certain response probabilities a priori to zero. The parameters pertaining to these response probabilities were not estimated in that case. However, at present such options are not implemented in the DILTRAN program. The choice for the total sample is, however, not optimal because, as was shown in the previous chapter, male and female respondents differ with respect to the difficulty of the five items. In the sections to follow, the data from the male subsample are analyzed.

In order to make a number of comparisons between competing models possible, the Nominal Response model with equality restrictions on the discrimination parameters was also analyzed for the male sample. When all five items were assumed to be characterized by the same set of $a_g$ parameters, the model was acceptable according to the $G^2$ ($G^2 = 218.02$, $df = 227$, $p = 0.654$), but the value of the Pearson $\chi^2$-statistic was less satisfying ($\chi^2 = 347.62$, $p = 0.000$). The parameters for the male subsample were estimated assuming four fixed latent nodes with values of $-2.1$, $-0.7$, $0.7$ and $2.1$, respectively. The reason for taking four rather than three latent nodes was that in the models for the male subsample that are presented in the following four latent nodes were used.

Because the two test statistics had such different values, it is not possible to place much confidence in these statistical tests. However, the restricted Nominal Response model does seem to fit much better for the male subsample than for the total sample.

#### 3.3.3.2 The Partial Credit Model

The Partial Credit model was proposed by Masters (1982). The model was developed for the analysis of ordered polytomous data. The basic idea behind the model is that when the ordered categories of an item are numbered from 0 through $m_j$, there are $m_j$ steps that an individual can take in responding to this item. The number of steps that an individual actually takes defines the credit that this individual is given; hence, the name Partial Credit model. It is obvious that this idea pertains particularly to items that are developed to measure some kind of performance. Such items can often be responded to at various levels. Hence, the number of steps that are successfully completed indicates the performance level of an individual. However, the idea of a successive number of thresholds that should be passed in order to reach a particular level can easily be transferred to attitudinal items.

The basic assumption defining the model is that the probability that an individual will pass the threshold between category $g-1$ and $g$ depends on the latent ability of the individual and the location of the threshold. Both the thresholds and the individuals are located on the same latent continuum; in other words, the model is unidimensional. The probability of passing the threshold between category $g-1$ and category $g$ is defined as the probability of scoring in category $g$ rather than in category $g-1$ or category $g$. Furthermore, the relation between this probability on the one hand and the two parameters characterizing the individual ability and the location of the threshold on the other is assumed to follow a logistic distribution. When the latent ability of individual $i$ is denoted again as $\theta_i$ and the value of the threshold between category $g-1$ and $g$ for item $j$ is indicated by $\delta_{jg}$, the model states that

$$\frac{p_{jg|\theta_i}}{p_{j,g-1|\theta_i}+p_{jg|\theta_i}}=\frac{\exp(\theta_i-\delta_{jg})}{1+\exp(\theta_i-\delta_{jg})}. \tag{3.12}$$

As the total number of categories is $m_j+1$, there are $m_j$ of these equations for each item $j$. An item with $m_j+1$ categories is, thus, broken down into $m_j$ hypothetical subitems and for these subitems the common Rasch model is specified. Because it must be true that $\sum_{g=0}^{m_j} p_{jg|\theta_i}=1$, it can be shown that the expression for the probability that individual $i$ will respond in category $g$ is

$$p_{jg|\theta_i} = \frac{\exp\left(\sum_{k=1}^{g}(\theta_i-\delta_{jk})\right)}{1+\sum_{h=1}^{m_j}\exp\left(\sum_{k=1}^{h}(\theta_i-\delta_{jk})\right)} \text{ for } g=1,\ldots,m_j \tag{3.13}$$

$$p_{j0|\theta_i} = \frac{1}{1+\sum_{h=1}^{m_j}\exp\left(\sum_{k=1}^{h}(\theta_i-\delta_{jk})\right)}. \tag{3.14}$$

The model can now be reparameterized in order to facilitate the comparison with restricted latent class models. This reparameterization also shows the restrictions imposed on the Nominal Response model in order to generate the Partial Credit model. The threshold parameters $\delta_{jg}$ are reformulated as

$$\delta_{jg} = c_{j,g-1} - c_{jg}.$$

This definition of the threshold parameters not only allows for a less complicated formulation of the Partial Credit model, it also focuses attention on the fact that the threshold parameter $\delta_{jg}$ not only pertains to category $g$, but also to category $g-1$ (see Molenaar, 1983a). Using the reparameterization in terms of the $c_{jg}$ parameters, the response probabilities can now be rewritten as

$$p_{jg|\theta_i} = \frac{\exp(g \cdot \theta_i + c_{jg})}{\sum_h \exp(h \cdot \theta_i + c_{jh})}. \tag{3.15}$$

As was done before, an identifiability restriction is imposed on the $c_{jg}$ parameters by stating that $c_{j0} = 0$. This formulation of the Partial Credit model makes it clear that the slope parameters in the Nominal Response model are restricted in the Partial Credit model to equal:

$$a_{jg} = g.$$

There are two important facets to this restriction. Firstly, the slope parameters are equal across items. Secondly, the slopes do not have to be estimated, because they equal the category number $g$. This also implies that the manifest items are measured on an interval scale and that the categories for these items are equidistantly spaced.

It is easy to derive a number of interesting properties concerning the category-characteristic trace lines. Firstly, the trace lines for $p_{j0|\theta_i}$ and $p_{jm_j|\theta_i}$ will intersect at a $\theta$-value which is equal to $\theta = -c_{jm_j}/m_j$. This specific $\theta$-value is denoted by $\theta^\star$. Defining two points on the latent scale at equal distances above and below $\theta^\star$ respectively leads to:

$$\begin{aligned} \theta^\circ &= \theta^\star + \varepsilon = -\frac{1}{m_j} c_{jm_j} + \varepsilon, \\ \theta^\bullet &= \theta^\star - \varepsilon = -\frac{1}{m_j} c_{jm_j} - \varepsilon. \end{aligned}$$

Now it can be proven that

$$\begin{aligned} p_{j0|\theta^\circ} &= p_{jm_j|\theta^\bullet}, \\ p_{j0|\theta^\bullet} &= p_{jm_j|\theta^\circ}. \end{aligned}$$

When the trace line for $p_{j0|\theta_i}$ is mirrored around $\theta^\star$, the resulting image is equivalent to the trace line for $p_{jm_j|\theta_i}$, and vice versa. These results are independent of the number of response categories existing for the manifest variables. Therefore, the restriction that $a_{jg} = g$ implies that the trace lines for $p_{j0|\theta_i}$ and $p_{jm_j|\theta_i}$ are mirrored around the point $\theta = -c_{jm_j}/m_j$.

If the manifest items have 3 different categories, it is easy to derive from Equation 3.10 that the $\theta$-value at which the trace line for the middle category reaches its maximum is equal to $(-c_{j2}/2)$. Furthermore, it can easily be shown that the probability of responding in the middle category is equal for $\theta^\circ$ and $\theta^\bullet$. In other words, the trace line for the middle category is symmetric around $\theta^\star$. This symmetric property of the Partial Credit model if the number of categories is equal to 3 does not necessarily hold when the number of categories of the manifest variables is greater than 3.

The $c_{jg}$ parameters in model 3.17 are often estimated using CML. Because the slope parameters equal the category numbers $g$, the person parameters $\theta_i$ can easily be conditioned out of the likelihood. The CML procedure for the Partial credit model was described by Wright and Masters (1982). Thissen and Mooney (1989) showed how CML estimates for these parameters can be obtained by using log-linear models. In his original article, Masters proposed a JML approach for estimating the parameters of the Partial Credit model (Masters, 1982).

Another method of dealing with the estimation problem is to employ MML methods. Glas and Verhelst (1989) proposed a parametric MML procedure. The estimation problem may also be solved by using fully semi-parametric MML. This approach, which was originally proposed for the Rasch model (for dichotomous data), can readily be generalized to the case of polytomous data. Using fully semi-parametric MML, the $c_{jg}$ parameters are estimated together with the values of a specified number of latent nodes as well as the latent proportions for this discretized latent distribution. Finally, the $c_{jg}$ parameters can also be estimated using semi-parametric MML with fixed node-points. It is necessary to include a scaling parameter in the latter approach because both the $g$ values of the categories and the $\theta_t$ values of the node-points are fixed in advance. The formulation for this model (denoted by model 4 in Figure 3.3) becomes

$$p_{jg|\theta_t} = \frac{\exp(\alpha \cdot g \cdot \theta_t + c_{jg})}{\sum_h \exp(\alpha \cdot h \cdot \theta_t + c_{jh})}. \tag{3.16}$$

The scaling parameter $\alpha$ is estimated together with the $c_{jg}$ parameters and the latent proportions $p_{\theta_t}$. This scaling parameter absorbs changes in the choice for the values of the latent nodes. Choices for the $\theta_t$ values that are proportionally different will not change the estimated $c_{jg}$ parameters.

The data on women's liberation for the male subsample were analyzed using the Partial Credit model. Again, only respondents who satisfied the educational requirements were used in the analysis. The Partial Credit model was estimated with three and four (fixed) latent nodes respectively. The values for the log-likelihood ratio and the Pearson $\chi^2$-statistic are given in Table 3.9. The results for $G^2$ in Table 3.9 indicate

Table 3.9: Semi-parametric MML with fixed nodes; Partial Credit model for the male subsample

| Number of latent nodes | $G^2$ | $p$ | Pearson's $\chi^2$ | $p$ | $df$ |
|---|---|---|---|---|---|
| $T = 3$ | 220.25 | 0.649 | 325.82 | 0.000 | 229 |
| $T = 4$ | 218.18 | 0.668 | 347.60 | 0.000 | 228 |

that the hypothesis that the data was sampled from a population for which the Partial Credit model held should be accepted. However, the Pearson $\chi^2$-statistic would seem to suggest a rejection of this hypothesis. The fact that both test statistics differ so clearly reflects the problem of sparse data. As was already noted in the previous chapter, in these cases the test statistics cannot be used without running into problems as it is questionable whether the sampling distribution of the test statistics was still approximately $\chi^2$-distributed. The question of whether the Partial Credit model was an acceptable model for the male population should then be answered by comparing this model with alternative models. The only alternative model that has been studied so far is the Nominal Response model. The parameters of this model could not, however, be estimated for the male sample. Other possible competing models are presented later in this section.

Despite the fact that it was not clear whether the Partial Credit model fit the data for the male subsample, the estimates for the parameters were interpreted substantively as this clarified some characteristics of the Partial Credit model.

As was stated before, the threshold is the value of $\theta$ where the two curves for adjacent categories intersect. In other words, it is the location on the latent scale where the probability for responding in category $g-1$ equals the probability of responding in category $g$. To the left of this $\theta$-value, category $g-1$ is more likely to occur, whereas to the right of this threshold value, subjects are more prone to come up with category $g$. This characteristic of the Partial Credit model can also be seen by examining the expression for the adjacent logits:

$$\begin{aligned}\ln\left[\frac{p_{j,g+1|\theta_i}}{p_{jg|\theta_i}}\right] &= \theta_i + (c_{j,g+1} - c_{jg}) \\ &= \theta_i - \delta_{j,g+1}.\end{aligned}$$

The threshold parameters can be interpreted in a clear-cut fashion in the graphical display of the category-characteristic curves. It is also apparent that the threshold parameters $\delta_{jg}$ are equal to the difference between the two successive parameters $c_{j,g-1}$ and $c_{jg}$. Therefore, the interpretation of the item difficulty parameters $c_{jg}$ is greatly facilitated by calculating the thresholds $\delta_{jg}$ from these parameters. Table 3.10 presents the estimated values both for the $c_{jg}$ and the $\delta_{jg}$ parameters.

It follows from Equations 3.13 and 3.14 that the location of the latent scale is set arbitrarily. Therefore, the comparison of results of different analyses can be hampered if there are differences between the locations of the latent scale. For that reason Table 3.10 also presents the values of the thresholds $\hat{\delta}^*_{jg}$ that are relocated so that the first threshold (i.e., $\hat{\delta}^*_{j1}$) equals 0.

The results in Table 3.10 show that $\delta_{j2} < \delta_{j1}$ for all of the items except item 1. This means that passing the threshold between categories 1 and 2 is easier than passing the threshold between categories 0 and 1. Within the perspective of the threshold approach, the ordering of the categories is somewhat problematic (see, for example, Andrich, 1992). The fact that $\delta_{j2} < \delta_{j1}$ for four out of five items demonstrates that the

Table 3.10: Estimated category parameters and thresholds for the Partial Credit model; male subsample

| | | Semi-parametric MML with $T = 4$ fixed nodes | | |
|---|---|---|---|---|
| | | $\hat{c}_{jg}$ | $\hat{\delta}_{jg}$ | $\hat{\delta}^*_{jg}$ |
| Item 1 | Cat. 1 | 0.07 | −0.07 | 0.00 |
| | Cat. 2 | −0.75 | 0.82 | 0.89 |
| Item 2 | Cat. 1 | 0.04 | −0.04 | 0.00 |
| | Cat. 2 | 0.47 | −0.43 | −0.39 |
| Item 3 | Cat. 1 | −1.24 | 1.24 | 0.00 |
| | Cat. 2 | −1.79 | 0.55 | −0.69 |
| Item 4 | Cat. 1 | 0.83 | −0.83 | 0.00 |
| | Cat. 2 | 2.61 | −1.78 | −0.95 |
| Item 5 | Cat. 1 | −1.98 | 1.98 | 0.00 |
| | Cat. 2 | −2.91 | 0.93 | −1.05 |

middle category is not very popular.

When the parameters in the Partial credit model are estimated using semi-parametric MML and fixed latent nodes, the model is closely related to a restricted version of the *Uniform Association model* discussed by Haberman (1978) and Goodman (1981, 1983, 1984). The restriction imposed in the Partial Credit model is that the association must not only be uniform within the cross-classifications for the latent variable and a specific indicator, but also uniform across these cross-classifications.

Despite this similarity, the two models (i.e., the Partial Credit model and the Uniform Association model) stress different aspects. Within the framework of the Partial Credit model, a crucial role is played by the threshold parameters. These thresholds, defined as the difference between successive log-linear one-variable parameters, have never received much attention from researchers involved with the Uniform Association model.

Within the context of this latter model, interest has been focused on the $\alpha$ parameter. From the measurement theoretical perspective that dominates the Partial Credit model, this $\alpha$ parameter is merely a scaling parameter that compensates for different choices for the values of the latent nodes. In the Uniform Association model, it is the only parameter that describes the association between item $j$ and the latent variable $\theta$. This can easily be illustrated as follows. Assuming as before that the categories for the manifest item are scored from 0 to $m_j$, the number of categories of item $j$ is equal to $(m_j + 1)$. Furthermore, for convenience and without loss of generality, it can be assumed that the latent nodes are scored with successive integers: $1, \ldots, t, \ldots, T$. As before, the number of latent classes is equal to $T$. Because of the assumption of local independence, all manifest items are related only to the latent variable $\theta$. Following the well known collapsibility theorem, the relationship between item $j$ and the latent variable $\theta$ can then be studied in the two-way $[(m_j+1)\times T]$-table. Consider a $2\times 2$ subtable for adjacent categories $g$ and $(g+1)$ of the manifest item $j$ and categories $t$ and $(t+1)$ for the latent variable $\theta$. Under the assumptions made, it can easily be demonstrated that the log-odds for this subtable is equal to

$$\ln \frac{p_{gt} \cdot p_{(g+1)(t+1)}}{p_{(g+1)t} \cdot p_{g(t+1)}} = \alpha.$$

Because this is true for each $2 \times 2$ subtable made up of adjacent categories, the $\alpha$ parameter has been termed the parameter of *uniform association* (Goodman, 1981). Within the context of linear by linear association models, the $\alpha$ parameter is not merely a scaling parameter, but rather a parameter which summarizes the degree of association for the complete $[(m_j + 1) \times T]$-table. As this $\alpha$ parameter in the Partial Credit model does not depend upon the specific item involved, this single parameter describes the association in all complete two-way tables made up of the manifest items on the one hand and the latent variable on the other.

Models with less restrictions can now be formulated by allowing the $\alpha$ parameter to vary over the manifest items. Such models, while preserving the threshold approach that characterizes the Partial Credit model, are more flexible because not all manifest items have to be equally "good indicators" of the latent trait, if the quality of an indicator is measured from the degree of association between the manifest and the latent variables. An example of such a model is discussed later on.

The intrinsic relevance of the $\alpha$ parameter becomes obvious when different groups measured on the same indicators are compared. The five items concerning women's liberation, for instance, were also measured in the female subsample. When the same scaling for the four latent nodes was used (i.e., $-2.1$, $-0.7$, $0.7$ and $2.1$), the estimated $\alpha$ parameter for the male sample was equal to 1.07, while the estimated value of this parameter in the female sample was 1.21. Therefore, it could be concluded that the indicators were better for the female sample than for the male sample. The greater the value of $\alpha$, the steeper the category characteristic curves were, and the clearer the distinction was between the categories of the manifest item. The interpretation of $\alpha$ as merely a scaling factor ignores the intrinsic relevance of this parameter.

Both the Partial Credit model and the Uniform Association model stress different, but equally important, aspects of the same model. Establishing a link between latent trait models and restricted latent class models is fruitful, both because of the more profound interpretation of the parameters that is made possible and because the formulation of relevant competing models is made easier.

One such competing model was introduced briefly in Chapter 2. This model will now be considered again by placing it in a latent trait perspective.

#### 3.3.3.3 Item-specific scaling parameters

Previously, the data for the male subsample were analyzed using the Partial Credit model. Although the value of the log-likelihood ratio indicated an acceptable fit, the value of the Pearson $\chi^2$-statistic cast some doubt on the adequacy of this model for describing this data. It is possible that the fit can be improved by relaxing the assumptions concerning the discrimination parameters.

As with all latent trait models discussed in this chapter, there is a linear relationship between the latent variable and the adjacent logits. For the Partial Credit model, the slopes of these linear functions do not depend on the specific item involved. When the categories of the manifest items are scored with successive integers, these slopes are all equal to $\alpha$ with semi-parametric MML with fixed latent nodes. The Partial Credit model does not account for differences in the discriminatory power of different items.

There are several ways of allowing for variation in the discrimination parameters. One method is, of course, to apply the Nominal Response model. This would, however, completely negate the philosophy underlying the Partial Credit model. The manifest items could no longer be seen as equally spaced variables, measured on an interval scale. Moreover, the idea of breaking up an item into $m_j$ hypothetical subitems and specifying simple logistic models for these subitems would be abandoned by returning to the Nominal Response model. In other words, the elegance of the threshold approach, so characteristic of the Partial Credit model, would be marred.

However, it is possible to specify a model which retains this threshold approach and at the same time is less restrictive with regard to the discrimination parameters by allowing this scaling parameter to vary over items. One such latent trait model was proposed by Heinen and Croon (1992). By introducing item-specific scaling parameters, the variation of discrimination parameters *among* items is accounted for. Within each item, however, there is just one slope parameter governing the adjacent logits for successive categories. The expression for the response probabilities for this Partial Credit model with item-specific scaling parameters (model 3 in Figure 3.3) is

$$p_{jg|\theta_t} = \frac{\exp(\alpha_j \cdot g \cdot \theta_t + c_{jg})}{\sum_h \exp(\alpha_j \cdot h \cdot \theta_t + c_{jh})}. \tag{3.17}$$

The item-specific scaling parameters $\alpha_j$ can be interpreted in several ways. One interpretation is that these $\alpha_j$ parameters allow for variation in the latent scale for different items. When an item has more discriminatory power than the other items, it is necessary to stretch the latent scale in order to make this item comparable to the other less discriminatory items. The value of $\alpha_j$ then gives a clear indication of the discriminatory power of an item. The higher the value of this scaling parameter, the better the item discriminates. Before returning to two other possible ways of interpreting the $\alpha_j$ parameters, some results will be presented for the application of this model to the data from the male subsample. The item-specific scaling parameters estimated with semi-parametric MML and 4 fixed latent nodes are given in Table 3.11.

These results illustrate the strong discriminatory power of item 3, a fact that was observed before when the data for the total sample were analyzed using the Lord-Birnbaum model. The values of the latent nodes were equally spaced. For $T = 4$, these values were equal to those

Table 3.11: Estimated item-specific scaling parameters for the male subsample

| | $\hat{\alpha}_j$ | |
|---|---|---|
| Item 1 | 0.66 | (0.10) |
| Item 2 | 0.96 | (0.15) |
| Item 3 | 2.78 | (1.18) |
| Item 4 | 0.78 | (0.14) |
| Item 5 | 0.81 | (0.13) |

mentioned in Table 3.6. The variation in the scaling parameters that is allowed for in this model only implies that the distances between successive categories can vary among the five items. The latent nodes are still equally spaced within each item.

The model with item-specific scaling parameters also provided a better fit than the original Partial Credit model. An interesting question is whether this improvement is largely due to the fact that item 3 was more discriminatory than the other four items, or the fact that these other four items also differed significantly in their slopes. In order to test this assumption, a model was fitted in which only item 3 was allowed to have a different slope parameter. The other four items were assumed to have equal slopes. The values of the two test statistics as well as the number of degrees of freedom are given in Table 3.12. The results are presented for models with $T = 3$ and $T = 4$ fixed latent nodes. For $T = 3$ the values of the latent nodes were chosen as reported in Table 3.2. To facilitate the comparisons between the models, the results for the original Partial Credit model already given in Table 3.9 are also included.

It can be seen from these results that item 3 had a different slope parameter than the other four items. When $\alpha_3$ was set free, the value of both the $G^2$ and the Pearson $\chi^2$ statistic decreased significantly. When the slopes for the other four items were also set free, the $G^2$ values decreased further, but not significantly so. With $T = 3$ latent nodes, the value of the Pearson statistic actually increased in this situation. This rather strange effect can probably be explained by the fact that the amount of data dealt with was relatively small. The overall conclusion is that item 3 had a different slope parameter than the other four items.

Table 3.12: Semi-parametric MML with fixed nodes; Partial Credit model for the male subsample with and without item-specific scaling parameters

| Number of nodes | | $G^2$ | $p$ | Pearson's $\chi^2$ | $p$ | $df$ |
|---|---|---|---|---|---|---|
| | Partial Credit model | 220.25 | 0.649 | 325.82 | 0.000 | 229 |
| $T = 3$ | $\alpha_3$ set free | 201.49 | 0.896 | 250.46 | 0.147 | 228 |
| | All $\alpha_j$ set free | 194.75 | 0.928 | 255.71 | 0.078 | 225 |
| | Partial Credit model | 218.18 | 0.668 | 347.60 | 0.000 | 228 |
| $T = 4$ | $\alpha_3$ set free | 193.53 | 0.948 | 265.87 | 0.039 | 227 |
| | All $\alpha_j$ set free | 189.76 | 0.953 | 254.43 | 0.079 | 224 |

These latter items conformed rather neatly to the original Partial Credit model.

For $T = 3$, the results for the Partial Credit model and the model with item-specific scaling parameters for all items were equivalent to those presented in Table 2.17. This is quite logical considering the Partial Credit model with item-specific scaling parameters is equivalent to the restricted latent class model ("Model 3") presented in section 2.4.4. As was mentioned before, the Partial Credit model is equivalent to "Model 4" in section 2.4.4. For "Model 3", the two-variable parameters are restricted by

$$u_{jg\theta_t} = u_j^* \cdot g \cdot \theta_t.$$

The equivalence between this restricted latent class model and the Partial Credit model with item-specific scaling parameters, provides a new perspective on the $\alpha_j$ parameters. Within the latent class framework, the $\alpha$ parameters are regarded as association coefficients. In the original Partial Credit model, there was just one $\alpha$ parameter and this parameter described the degree of association within all $n$ $[(m_j+1)\times T]$-tables formed by the manifest items on the one hand and the latent variable on the other. With item-specific scaling parameters, each $[(m_j + 1) \times T]$-table is characterized by its own coefficient of uniform association. There is, therefore, just one relevant parameter describing the degree of association for each table. This parameter, however, is allowed to obtain different values for the various tables. Thus, the estimated $\alpha_j$ parameters for the male subsample indicated that the association between the latent variable and item 3 was much stronger than the association between $\theta$ and the other indicators. Hence, the model with item-specific scaling parameters was equivalent to the original Uniform Association model.

Finally, a third way of interpreting the $\alpha_j$ parameters becomes obvious by noting that the model with item-specific scaling parameters preserves the threshold approach in the Partial Credit model. It must be kept in mind that the Partial Credit model was developed by focussing on the probability that an individual will respond in category $g$ rather than in category $g-1$. The probability that this threshold between categories $g$ and $g-1$ will be passed was expressed earlier as a Rasch model (see Equation 3.12). It can now be easily derived from Equation 3.17 that

$$\frac{p_{jg|\theta_t}}{p_{j,g-1|\theta_t} + p_{jg|\theta_t}} = \frac{\exp(\alpha_j \cdot \theta_t - \delta_{jg})}{1 + \exp(\alpha_j \cdot \theta_t - \delta_{jg})}. \tag{3.18}$$

Thus the threshold approach is still adequate for the Partial Credit model with item-specific scaling parameters, although the probability of passing a given threshold is now modeled by the Lord-Birnbaum model instead of the Rasch model. This also illustrates clearly why this model accounts for variation in the discriminatory power of the various indicators. The $\alpha_j$ parameters function as the slopes of the curves describing the probability of passing a threshold within item $j$. For each item $j$, all of these "threshold characteristic" curves run parallel as there is just one $\alpha_j$ parameter for each item. For different items, however, these curves may vary in the slopes.

The idea of generalizing the Partial Credit model by using the Lord-Birnbaum instead of the Rasch model for modeling the probability of passing a threshold was suggested by Heinen and Croon (1992). Independently of these authors, Muraki (1992) proposed the same model calling it the *generalized Partial Credit model.* However, Muraki estimated the parameters in this model using parametric MML. Because he did not discretize the latent trait, the relation between this generalized Partial Credit model and the Uniform Association model was lost.

### 3.3.4 Restrictions on the difficulty parameters

A number of models with restrictions placed upon the difficulty parameters have been proposed within the context of latent trait analysis. Until recently, this type of restriction rarely attracted the attention of researchers using latent class analysis (see, for example, Formann, 1985), because within the latent class framework, the $c_{jg}$ parameters (i.e., the log-linear one-variable parameters) are not considered to be inherently interesting.As was pointed out in the previous section, this is not entirely true. After all, the $c_{jg}$ parameters are the building blocks for the threshold parameters of the models that use linear restrictions on the relation between latent and manifest variables, and these latter parameters are quite interesting from a measurement-theoretical point of view. If difficulty parameters are studied from this threshold perspective, a number of interesting models come into play.

Firstly, the possibility of applying the type of restriction used in the previous section to the difficulty parameters can be examined. One of the restrictions reviewed in Section 3.3.3 involved the equality of the set of $a_{jg}$ parameters over a number of items. Analogously, it is possible to set the restriction that the $c_{jg}$ parameters for some (or all) items must be equal to each other ($c_{jg} = c_g$). In the Nominal Response model, the adjacent logits are a linear function of the latent variable. Because the intercept of this linear function depends on the $c_{jg}$ parameters, this restriction, when used in the context of the Nominal response model, will restrict the intercepts for corresponding linear functions to be equal over items. It is not clear what the substantive significance of such a restriction is.

Another possibility is to restrict both the $a_{jg}$ parameters and the $c_{jg}$ parameters to be equal over a number of items. The response probabilities are then expressed as

$$p_{jg|\theta_i} = \frac{\exp(a_g \cdot \theta_i + c_g)}{\sum_{h=0}^{m} \exp(a_h \cdot \theta_i + c_h)} \text{ for } g = 0, \ldots, m. \tag{3.19}$$

When semi-parametric MML with fixed latent nodes is used, $\theta$ is indexed by $t$ rather than by $i$. When both the $a_{jg}$ and the $c_{jg}$ parameters are restricted to be equal over a set of items, all item-specific information is lost and the category-characteristic curves for these restricted items are identical. It is clear that these restrictions are strict and may appear to only be of practical value for investigating whether the trace lines for two (or more) items differ significantly. In this manner, hypotheses about parallel items ("parallel" in the psychometric sense) can be tested by comparing the log-likelihood ratio values for the two relevant models. In Figure 3.3 this model is numbered model 8.

As an example, the data for the total sample were again used. From the results in Table 3.8 it was concluded that the discrimination parameters for items 2 and 3 could be restricted to be equal without a detrimental loss in the goodness of fit. The same conclusion could be drawn with regard to items 1 and 5. The question arose whether these two sets of items could be restricted any further by assuming that all parameters (both $a_{jg}$ and $c_{jg}$) were equated. When this assumption was made with respect to items 2 and 3, the $G^2$ value became 547.49 which, with 224 degrees of freedom, was highly significant. This result was not surprising as item 2 was one of the easier items, while item 3 was a rather difficult item. When all of the parameters for items 1 and 5 were equated, the results were likewise unsatisfactory, though they were less dramatic. The $G^2$ value was 368.51 ($df = 224$).

Of course, the same types of assumptions can also be tested within the framework of the Partial Credit model, with or without item-specific scaling parameters. The expressions for the response probabilities are

$$p_{jg|\theta_t} = \frac{\exp(\alpha_j \cdot g \cdot \theta_t + c_g)}{\sum_h \exp(\alpha_j \cdot h \cdot \theta_t + c_h)}, \tag{3.20}$$

$$p_{jg|\theta_t} = \frac{\exp(\alpha \cdot g \cdot \theta_t + c_g)}{\sum_h \exp(\alpha \cdot h \cdot \theta_t + c_h)}. \tag{3.21}$$

The first of these two expressions preserves some item-specific information; the other does not. In Figure 3.3, these models are shown as models 10 and 9, respectively.

The manner in which these restricted models can be used is illustrated by the data of the male subsample. In the previous section these data were analyzed using, among others, the Partial Credit model with item-specific scaling parameters. The results in Table 3.12 showed that this model fitted the data quite satisfactory. In order to convert it into a more parsimonious model, the model with item-specific scaling parameters was restricted in two ways.

First, as was also reported in Table 3.12, it is possible to assume that a number of scaling parameters are equal over items. In that case, equality restrictions are imposed on some of the $\alpha_j$ parameters but *not* on the $c_{jg}$ parameters. When all scaling parameters are assumed to be equal, the resulting model is the original Partial Credit model. It was found that item 3 was more discriminating than the other four items so a model in which the scaling parameters for these other four items were restricted to be equal, gave acceptable results (see Table 3.12). When scaling parameters are assumed to be equal for a number of items, the units of measurement of the latent scale underlying these items are made equal. Items are, however, still allowed to differ from each other by the $\theta$ values at which the thresholds are located because the $c_{jg}$ parameters are not restricted.

A second means of creating a model which is more parsimonious than the Partial Credit model with item-specific scaling parameters is to restrict the threshold values rather than the scaling parameters. When the $c_{jg}$ parameters are restricted to be equal over a number of items but the $\alpha_j$ parameters are set free, it is assumed that the threshold values for this subset of variables are *proportionally* equal. The distances between the thresholds are thus proportionally equal, with "proportionally" signifying a proportionality relative to $\alpha_j$. Differences in the spacing of the thresholds are due solely to the shrinking or stretching of the latent scale.

From a measurement-theoretical point of view, the difference between the two types of restriction is straightforward. In the first case, in which some of the $\alpha_j$ parameters are restricted, the latent scale is "measured" in the same units for all items involved in the restricting assumptions. But the thresholds for these items can be located anywhere on the latent scale. In the second case, only some $c_{jg}$ parameters are restricted. The distances between the threshold values are now proportionally equal; proportionally, because the $\alpha_j$ values are free to vary and these parameters are determined by the degree the latent scale is shrunk or stretched for the particular items.

When these assumptions are looked at from the "association" perspective, the substantive relevance of the restrictions is not always obvious. This association perspective focuses on the linear relationship between the latent variable and the adjacent logits. If the $\alpha_j$ parameters are restricted, the slopes of these linear relationships will be equal across items. In other words, the degree of association between the latent variable on the one hand and the manifest items on the other is constant. This assumption clearly has substantive significance. If, however, the $c_{jg}$ parameters are restricted to be equal, as opposed to the $\alpha_j$ parameters, this means that the intercepts are restricted to be equal, which in turn, has no clear theoretical meaning. Assumptions concerning the $c_{jg}$ parameters appear, therefore, to bear more substantive relevance when these assumptions are interpreted in the context of the threshold approach.

Table 3.13 contains some of the results pertaining to the Partial Credit models with item-specific scaling parameters in which the $c_{jg}$ parameters were restricted for some items. In Table 3.13 only certain

Table 3.13: Semi-parametric MML with $T = 4$ fixed nodes; Partial Credit model for the male subsample with item-specific scaling parameters and equality restrictions on the $c_{jg}$ parameters

| Items for which $c_{jg}$ are restricted | $G^2$ | $p$ | Pearson's $\chi^2$ | $p$ | $df$ |
|---|---|---|---|---|---|
| — | 189.76 | 0.953 | 254.43 | 0.079 | 224 |
| 1 2 | 198.95 | 0.902 | 248.15 | 0.149 | 226 |
| 1 3 | 197.80 | 0.912 | 258.02 | 0.071 | 226 |
| 3 4 | 202.38 | 0.869 | 357.11 | 0.000 | 226 |
| 3 5 | 201.72 | 0.876 | 396.14 | 0.001 | 226 |
| 2 3 5 | 240.83 | 0.267 | 321.80 | 0.000 | 228 |

subsets of those items with restricted $c_{jg}$ parameters which gave an acceptable fit are reported. These results illustrate that it is possible to assume that two items have proportionally the same threshold values. If, however, the $c_{jg}$ parameters are equated for 3 or more items, the fit deteriorates rapidly. Remarkably enough, relatively difficult items (for example, item 3) and relatively easy items (i.e., item 4) can both be present in subsets of restricted items.

The model can be further restricted by assuming that, for a subset of items, both the $\alpha_j$ and the $c_{jg}$ parameters are submitted to equality constraints. This leads to the model in Equation 3.21. If items are restricted according to this model, all item-specific information is lost. Items which belong to such a restricted subset are characterized by the same set of trace lines. Clearly, when these assumptions are applied to pairs of variables in which one variable is relatively difficult and the other item relatively easy, the model will provide a bad fit. Table 3.14 provides the results for the pairs of variables that were also reported in Table 3.13.

Table 3.14: Semi-parametric MML with $T = 4$ fixed nodes; Partial Credit model for the male subsample with equality restrictions on both the scaling parameters and the $c_{jg}$ parameters

| Items for which $\alpha_j$ and $c_{jg}$ are restricted | $G^2$ | $p$ | Pearson's $\chi^2$ | $p$ | $df$ |
|---|---|---|---|---|---|
| 1 2 | 259.20 | 0.070 | 299.54 | 0.001 | 227 |
| 1 3 | 268.40 | 0.031 | 365.28 | 0.000 | 227 |
| 3 4 | 562.04 | 0.000 | 650.50 | 0.000 | 227 |
| 3 5 | 252.20 | 0.120 | 314.40 | 0.000 | 227 |

As these results show, there is some evidence that items 3 and 5 have identical relationships to the latent variable $\theta$. Note that this model provides a reasonable fit, although it has only 5 degrees of freedom less than the model of total independence.

Obviously, it does not make much sense to apply these restrictions to all manifest items because the resulting model would be even more parsimonious than the model for total independence between the indicators,[7] i.e., a log-linear model in which only one-variable parameters for the manifest indicators appear. As this latter model is generally seen as a base-line model, it would not be profitable to investigate models that are even more parsimonious.

Besides imposing equality restrictions on the $c_{jg}$ parameters, it is also possible to make the difficulty parameters a linear function of the category number $g$, i.e., $c_{jg} = c_j \cdot g$. It would seem natural to make this assumption only if the same kind of restrictions are also imposed on the discrimination parameters. In other words, the restriction that $c_{jg} = c_j \cdot g$ would only be made for the Partial Credit model with or without item-specific scaling parameters. The expressions for the

[7]The resulting model would be more parsimonious than the model for total independence, unless the number of latent nodes is quite large. In that case, the model would probably not be identifiable.

response probabilities would become

$$p_{jg|\theta_t} = \frac{\exp(\alpha_j \cdot g \cdot \theta_t + c_j \cdot g)}{\sum_h \exp(\alpha_j \cdot h \cdot \theta_t + c_j \cdot h)}, \quad (3.22)$$

$$p_{jg|\theta_t} = \frac{\exp(\alpha \cdot g \cdot \theta_t + c_j \cdot g)}{\sum_h \exp(\alpha \cdot h \cdot \theta_t + c_j \cdot h)}. \quad (3.23)$$

In Figure 3.3 these models are denoted as model 12 and model 11, respectively. It should be noted that the $c_j$ parameters in Equations 3.22 and 3.23 are replaced by $-\delta_j$ in Figure 3.3. The reason for this shift in notation is that the models in Equations 3.22 and 3.23 can also be regarded as restricted Rating Scale models; in these models the item difficulty is denoted by $\delta_j$.

The model in Equation 3.22 allows for item-specific information. It can easily be verified that with this assumption imposed on the $c_{jg}$ parameters, all threshold values are equal to $-c_j$. The model implies that all category-characteristic curves intersect at one point. The $\theta$-value at which this occurs is equal to $-c_j$ and this is also the $\theta$-value at which the category-characteristic curve for the middle category of item $j$ reaches its maximum, provided that the manifest items have 3 categories. A distinguishing characteristic of this model is that for all $\theta$-values but one, one of the two most extreme categories has the highest probability of occurring. The probability of responding in one of the other categories is always lower. Only for $\theta = c_j$ are the response probabilities equal for all categories.

The model in 3.22 was estimated for the male subsample. The obtained values for the test statistics as well as the number of degrees of freedom and the probability levels are reported in Table 3.15.

Table 3.15: Semi-parametric MML with $T = 4$ fixed nodes; model 3.22 and 3.23 for the male subsample

| | $G^2$ | $p$ | Pearson's $\chi^2$ | $p$ | $df$ |
|---|---|---|---|---|---|
| Model 3.22 | 224.80 | 0.566 | 519.80 | 0.000 | 229 |
| Model 3.23 | 276.31 | 0.027 | 506.19 | 0.000 | 233 |
| Model 3.23 with item 3 set free | 249.74 | 0.202 | 410.96 | 0.000 | 232 |

According to $G^2$, this model fits very well. The value of the Pearson $\chi^2$-statistic is, however, highly significant. When the restriction

$c_{jg} = c_j \cdot g$ was applied to the Partial Credit model without item-specific scaling parameters and the parameters for model 3.23 were thus estimated, the resulting model yielded values for the test statistics which were much higher. As it is known that this phenomenon can be explained by the fact that item 3 was more discriminatory than the other four items, a mixed model in which item 3 was allowed to have its own scaling parameter was also estimated. Although this mixed model fit better than model 3.23, it still provided a worse fit than model 3.22.

#### 3.3.4.1 The Rating Scale model

The restrictions on the difficulty parameters discussed so far either equated some $c_{jg}$ parameters over a set of items or linearized these parameters in terms of the categories of the manifest variables, i.e., $c_{jg} = c_j \cdot g$. Another method of restricting these parameters is to separate $c_{jg}$ into one component that pertains to the *item* difficulty and another component that deals with the specific *category*.

This idea of separating the difficulty parameters into two components was first suggested by Andrich (1978a). Actually, Andrich suggested such a separation with regard to the threshold parameters and not to the $c_{jg}$ parameters themselves. The Rating Scale model proposed by Andrich (1978a) can be derived in the same fashion as the Partial Credit model, although Andrich used another method that is less compelling and straightforward, as is shown later on. In fact, the Rating Scale model can be seen as a restricted Partial Credit model. It should be remembered that in the Partial Credit model the thresholds only depend on the $c_{jg}$ parameters. Therefore, restricting these thresholds by separating an item component from a category-specific component can be translated directly into comparable restrictions with respect to the $c_{jg}$ parameters. The restriction that leads to the Rating Scale model is

$$\delta_{jg} = \delta_j + \tau_g. \tag{3.24}$$

In the Rating Scale model, items differ only by the value of $\delta_j$. This parameter can be regarded as a kind of mean location for the category-specific trace lines that belong to a certain item. The thresholds themselves are located at distance $\tau_g$ from this mean location. These latter distances are, of course, category-specific, but they do not vary over items. The spacing of the thresholds is the same for all items. Items only differ in the location of these thresholds. It is obvious that these restrictions only make sense if the number of response categories is equal for all items.

The restriction in terms of the threshold $\delta_{jg}$ can easily be translated into one that pertains to the $c_{jg}$ parameters. Within the context of the Partial Credit model, thresholds can be expressed as

$$\delta_{jg} = c_{j,g-1} - c_{jg},$$

from which it can easily be inferred that

$$c_{jg} = -\sum_{h=1}^{g} \delta_{jh} \text{ for } g = 1, \ldots, m.$$

For the Rating Scale model, the following results are obtained:

$$\begin{aligned} c_{jg} &= -\sum_{h=1}^{g} (\delta_j + \tau_h) \\ &= -g \cdot \delta_j - \sum_{h=1}^{g} \tau_h \\ &= -g \cdot \delta_j + c_g \text{ where } c_g = -\sum_{h=1}^{g} \tau_h. \end{aligned} \tag{3.25}$$

So, for the response probabilities, the expression now becomes

$$\begin{aligned} p_{jg|\theta_i} &= \frac{\exp(g \cdot \theta_i - g \cdot \delta_j + c_g)}{\sum_h \exp(h \cdot \theta_i - h \cdot \delta_j + c_h)} \\ &= \frac{\exp\left[g(\theta_i - \delta_j) + c_g\right]}{\sum_h \exp\left[h(\theta_i - \delta_j) + c_h\right]}. \end{aligned} \tag{3.26}$$

Andrich (1978a) derived this model in the following fashion. If it is supposed that an item has three categories and therefore two thresholds, each threshold will have an associated variable which can take on just two values: the value 0 if the threshold is not passed and the value 1 if the threshold is passed, denoted as variables $y_1$ and $y_2$. Andrich also assumed that the two thresholds were ordered along with the category numbers, thus $\tau_1$ was located to the left of $\tau_2$. Next, it is assumed that the probabilities of passing these two thresholds can be expressed as simple Rasch models, taking into account the above mentioned restriction on $\delta_{jg}$:

$$p_{y_1=1|\theta_i} = \frac{\exp(\theta_i - (\delta_j + \tau_1))}{1 + \exp(\theta_i - (\delta_j + \tau_1))},$$
$$p_{y_2=1|\theta_i} = \frac{\exp(\theta_i - (\delta_j + \tau_2))}{1 + \exp(\theta_i - (\delta_j + \tau_2))}.$$

In order to obtain expressions for the response probabilities pertaining to the original variable, it is necessary to combine the probabilities for passing the thresholds. Andrich made the assumption that the variables $y_1$ and $y_2$ were independent. Theoretically, four outcomes can be observed for the joint variable obtained by combining $y_1$ and $y_2$. The probabilities for observing these four outcomes can be expressed easily from the probabilities for passing the thresholds given above, owing to the assumption of independence. However, given the assumption that $\tau_1$ is located to the left of $\tau_2$, it is not possible that $y_1 = 0$ while $y_2 = 1$. The probability that corresponds to this outcome for the joint variable must be zero, and the other three probabilities must be normalized again to insure that their sum equals 1. All this requires is some simple algebra. The resulting equation is Equation 3.26.

The derivation of the Rating Scale model as given by Andrich leads to the following observations. Firstly, the threshold approach on which the Rating Scale is built incorporates the assumption that the thresholds are ordered along the latent continuum in the same way as the ordered categories of the manifest variable. Andrich (1992) defended this assumption on measurement-theoretical grounds. However, the assumption that the thresholds are ordered along with the category numbers is not translated into the model. This means that it is possible to get an estimated value for $\tau_2$ that is smaller than the estimated value for $\tau_1$. Within the threshold perspective, such results are not allowed. An inappropriate ordering of the thresholds can indicate multidimensionality of the latent variable or an inappropriate ordering of the categories of the manifest variable. However, if the marginal frequencies for the middle categories of the indicators are relatively small (as is often the case for attitudinal data), the estimated $\tau$ parameters are often not ordered along with the categories of the indicator.

Secondly, the assumption that the variables pertaining to the thresholds ($y_1$ and $y_2$) are mutually independent seems hard to defend. It seems reasonable to assume that it is only possible to pass the threshold with the highest value after the easier threshold has been passed. The manner in which the Rating Scale model was derived by Andrich is not logical in this sense. Therefore, it is preferable to derive the Rating Scale model in the same way the Partial Credit model was developed. This means that the probability of responding in category $g$ rather than in category $(g-1)$ of item $j$ can be expressed as a simple Rasch model, in which the Rating Scale restriction with respect to the parameter $\delta_{jg}$ is included:

$$\frac{p_{jg|\theta_i}}{p_{j,g-1|\theta_i} + p_{jg|\theta_i}} = \frac{\exp\left[\theta_i - (\delta_j + \tau_g)\right]}{1 + \exp\left[\theta_i - (\delta_j + \tau_g)\right]}. \tag{3.27}$$

In the same manner the Partial Credit model was developed, this expression can be used to derive Equation 3.26. The parameters in the Rating Scale model, as presented in Equation 3.26, can be estimated in a number of ways. In one of his earlier articles, Andrich (1978b) proposed JML. However, the parameters in this Rating Scale model can also be estimated using CML because the sufficient statistic for $\theta_i$ is equal to the total-score for individual $i$ (and therefore known), provided that the manifest variables are scored with successive integers. In other articles, in which certain variants of the Rating Scale model were developed, Andrich (1979, 1982) used sufficient statistics to estimate threshold-related parameters. Finally, MML can also be employed to estimate the parameters in this model.

The use of semi-parametric MML with fixed latent nodes is illustrated at a later point in this section. For this procedure, the expression for the response probabilities includes a scaling parameter, because the values for the latent nodes are fixed in advance:

$$p_{jg|\theta_t} = \frac{\exp\left[g(\alpha \cdot \theta_t - \delta_j) + c_g\right]}{\sum_h \exp\left[h(\alpha \cdot \theta_t - \delta_j) + c_h\right]}. \tag{3.28}$$

In Figure 3.3, this model has been numbered model 6.

Like the Partial Credit model, the Rating Scale model can also be generalized by including item-specific scaling parameters. This leads to model 7 in Figure 3.3. The expression for the response probabilities in this model is

$$p_{jg|\theta_t} = \frac{\exp\left[g(\alpha_j \cdot \theta_t - \delta_j) + c_g\right]}{\sum_h \exp\left[h(\alpha_j \cdot \theta_t - \delta_j) + c_h\right]}. \tag{3.29}$$

This model can also be generated from the concept that the probability of responding in category $g$, rather than in category $(g-1)$ is a logistic function as expressed in Equation 3.27. However, the building blocks are no longer simple Rasch models but Lord-Birnbaum models with an item-specific discrimination parameter. The model with item-specific scaling parameters is, of course, much more flexible than the original Rating Scale model. The parameters in this extended model can be easily estimated using semi-parametric MML with fixed latent nodes.

Instead of introducing item-specific scaling parameters, one could also generalize the Rating Scale model by including category-specific discrimination parameters in expression 3.27. In that case, the *Unidimensional Polychotomous Rasch model* is obtained (Rasch, 1961; model 5 in Figure 3.3):

$$p_{jg|\theta_t} = \frac{\exp\left[a_g(\theta_t - \delta_j) + c_g\right]}{\sum_h \exp\left[a_h(\theta_t - \delta_j) + c_h\right]}. \tag{3.30}$$

This model was studied by Andersen (1977), who proved that the sufficient statistic for $\theta_i$ is only known and equal to the total-score if the $a_g$ parameters have equidistant values. The model in 3.26 is, therefore, the only variant of the Rating Scale model for which it is true that there is a known sufficient statistic for $\theta_i$. As stated earlier, this variant has the advantage that the item and category parameters can be estimated using CML. This is not, however, the decisive advantage in the view of Andrich, as he proposed the use of JML for estimation purposes. The main reason Andrich prefers the Rating Scale model to the more general Unidimensional Polychotomous Rasch model is a substantive one. When the discrimination parameters $a_g$ are replaced by the category numbers $g$, as is done in Equation 3.26, the model is said to use an integral scoring function, i.e., the categories of the manifest variables are scored with successive integers. In Andrich's view, this integral scoring function is attractive because the sufficient statistic for the ability parameter is equal to the number of thresholds passed. Of course, this statement is only true if the thresholds are ordered in correspondence with the category numbers. However, the estimated $\hat{\tau}$ values may be ordered improperly. Whether the raw total-score equals the number of thresholds that an individual has passed depends on the estimates.

Andrich (1978a) mentions still another advantage of the integral scoring function:

> ... *The second and related point is that the integral scoring function does not arise from any references to distances. Instead, it arises from the number of equally discriminating thresholds passed. This is in contrast to certain traditional formulations where the so called "equidistant" scoring is considered justified only if the distances between successive categories are equal.* ...

Andrich was correct when he claimed that there are no assumptions made with respect to the distances between successive thresholds and that the use of an integral scoring function implies nothing concerning these distances. But that is not the point. Equidistant scoring means that the categories of the manifest variables are considered to be equally spaced. The fact that these fixed distances between categories do not necessarily lead to equal distances between thresholds does not mitigate the assumption concerning the equidistant weights that are given to the successive categories of the manifest indicators in the calculation of the response probabilities in Equation 3.26. In this sense, the Rating Scale model is no less "traditional" than other formulations.

Of course, the Rating Scale model with an integral scoring function is a very simple model and therefore valuable in its own right. But the advantages that Andrich imputed to the integral scoring function are somewhat questionable. The more general Rating Scale models, like the model with item-specific scaling parameters and the Unidimensional Polychotomous Rasch model are applicable in a larger number of practical situations because Andrich' Rating Scale model is highly restrictive. As was noted earlier, the parameters in the model with item-specific scaling parameters can easily be estimated using semi-parametric MML with fixed latent nodes. Within the context of this estimation procedure, this model assumes a natural position. This same procedure can also be used for the estimation of the parameters in the Unidimensional Polychotomous Rasch model. This was already implied by the notation used in Equation 3.30 in which the latent variable $\theta$ was indexed by $t$

and not by $i$.

The data on women's liberation for the male subsample were analyzed with the three models presented in this section. The parameters were estimated using semi-parametric MML with fixed latent nodes. In order to establish comparability with the other discretized latent trait models presented in this chapter, the number of latent nodes was equal to 4. Furthermore, the values for these latent nodes were chosen in the same fashion as in the models introduced earlier.

It should be noted that the estimation of parameters in the Unidimensional Polychotomous Rasch model is numerically more problematic than in any of the other discretized latent trait models discussed so far. The reason for these problems is the fact that the expression for the response probabilities contains the product of the parameters $a_g$ and $\delta_j$. Therefore, this product also appears in the log-likelihood that must be minimized. For all other discretized latent trait models, the log-likelihood function only comprises (functions of) estimable parameters in an additive expression. The estimation of the parameters in the Unidimensional Polytomous Rasch models appeared to be quite sensitive to choices of the initial estimates.

Table 3.16 shows some of the results pertaining to the goodness of fit statistics for the three models applied to the data for the male subsample. It is clear from these results that the Rating Scale model does

Table 3.16: Semi-parametric MML with $T = 4$ fixed nodes; Rating Scale model (RS) with and without item-specific scaling parameters; the Unidimensional Polychotomous Rasch model (UPRM) for the male subsample

| | $G^2$ | $p$ | Pearson's $\chi^2$ | $p$ | $df$ |
|---|---|---|---|---|---|
| RS Model | 269.46 | 0.046 | 421.48 | 0.000 | 232 |
| RS Model with item-specific scaling | 222.39 | 0.592 | 453.07 | 0.000 | 228 |
| UPRM | 268.02 | 0.048 | 437.23 | 0.000 | 231 |

not provide an acceptable fit. This is not surprising as the Rating Scale model is a restricted Partial Credit model and this latter model likewise did not fit the data from the male subsample very well. The other two models (UPRM and RS with item-specific scaling parameters) provide their own generalization for the Rating Scale model and should therefore yield $G^2$ values which are smaller than the $G^2$ for the Rating Scale model.

In the case of the UPRM, this generalization was obtained by introducing discrimination parameters that were category-specific but constant over items. The assumption made was that the probability for responding in category $g$ rather than in category $(g-1)$ was governed not only by the "mean" item difficulty $\delta_j$ and the value for the threshold $\tau_g$, but also by a discrimination parameter $a_g$ which depended on the specific threshold in question. Introducing this parameter had as a consequence that some thresholds were passed more easily than others. This variation in difficulty of the thresholds was identical for all items. A direct result of this parameterization of the threshold difficulties was that the manifest indicators were no longer considered as variables measured on an interval scale because the category numbers $g$ were not used as weights in the calculation of the response probabilities. This means that the same remarks that were made with respect to the Nominal Response model, concerning the ordering of the categories of the manifest items, also apply to the UPRM. If the adjacent logits are a monotone increasing function of the latent position $\theta_i$ (or if the latent variable is discretized, $\theta_t$), then the weights $a_g$ should also be ordered so that

$$a_0 \leq a_1 \leq \ldots \leq a_{m-1} \leq a_m.$$

Again, the model does not insure that these inequalities will hold. If, however, the categories for the manifest indicators are ordered properly, the estimated $\hat{a}_g$-values will show this ordering.

The results in Table 3.16 indicate clearly that the generalization provided by the UPRM did not improve the fit substantially. Allowing for variation in the difficulty of passing the thresholds was not a successful strategy. As the results in previous sections already indicated, this lack of fit was due to the fact that the five items differed with respect to their discriminatory power. The most pronounced item in this respect was item 3. This item was not only one of the more difficult items, it was also the item which had the greatest discriminatory power. Thus, in order for a model to be more successful in the sense that it provides a smaller $G^2$ value it must take this aspect into account.

The Rating Scale model with item-specific scaling parameters is such a model. This generalization of the Rating Scale model does not focus on variations in the difficulty of thresholds (which are constant for all items) as the UPRM does, but instead introduces variation in the discriminatory power of items (which is, within each item, constant for all thresholds). When the fit is evaluated using the $G^2$, this generalized model gives a remarkably good fit. However, when attention is focused on the Pearson $\chi^2$-statistic, this model must also be rejected, just as the other two variants of the Rating Scale model were. Again, the problem of performing adequate statistical tests in the presence of sparse data clearly makes it impossible to draw clear-cut conclusions.

To conclude the discussion of the Rating Scale model, the estimated values of the parameters for the three models presented in this chapter are also reported. The estimated standard errors for the parameters are given in parentheses.

Table 3.17: Semi-parametric MML with $T = 4$ fixed nodes; parameter estimates for Rating Scale model (RS) with and without item-specific scaling parameters and the unidimensional polychotomous Rasch model (UPRM) for the male subsample

| | RS | | RS with item-specific scaling | | UPRM | |
|---|---|---|---|---|---|---|
| $\delta_1$ | $0.0^a$ | – | $0.0^a$ | – | $0.0^a$ | – |
| $\delta_2$ | $-0.70$ | (0.10) | $-0.44$ | (0.13) | $-0.60$ | (0.10) |
| $\delta_3$ | 0.41 | (0.09) | 1.43 | (0.30) | 0.38 | (0.08) |
| $\delta_4$ | $-1.89$ | (0.15) | $-1.68$ | (0.15) | $-1.66$ | (0.15) |
| $\delta_5$ | $-1.00$ | (0.10) | $-1.56$ | (0.19) | $-0.89$ | (0.10) |
| | $\alpha = 1.13$ | (0.08) | $\alpha_1 = 0.58$ | (0.08) | $a_0 = 0.00^a$ | – |
| | | | $\alpha_2 = 1.07$ | (0.14) | $a_1 = 1.26$ | (0.14) |
| | | | $\alpha_3 = 1.97$ | (0.41) | $a_2 = 2.32$ | (0.15) |
| | | | $\alpha_4 = 1.00$ | (0.15) | | |
| | | | $\alpha_5 = 1.30$ | (0.18) | | |
| $c_0$ | $0.0^a$ | – | $0.0^a$ | – | $0.0^a$ | – |
| $c_1$ | $-0.67$ | (0.14) | $-0.03$ | (0.11) | $-0.72$ | (0.18) |
| $c_2$ | $-1.03$ | (0.26) | 0.13 | (0.19) | $-0.99$ | (0.25) |

[a] Value restricted by assumption

All three models clearly showed the differences in difficulties for the five items. Items 3 and 5 were more difficult, while items 2 and 4 were relatively easy. The value of $\delta_1$ was fixed in advance at 0.0, in order to achieve the necessary location restriction. As the estimated values for $a_g$ in the UPRM increased with the category numbers, the adjacent logits were in fact a monotone increasing function of the latent variable $\theta$. The estimated item-specific scaling parameters reflected the fact that item 3 had greater discriminatory power than the other four items. Finally, from the estimated $c_g$ parameters for the Rating Scale models with and without item-specific scaling parameters, one can easily derive the values for the threshold values $\tau_g$. Given the relationship between the $\tau_g$ parameters and the $c_g$ parameters that was presented earlier, it can easily be shown that

$$\begin{aligned} \tau_1 &= -c_1, \\ \tau_2 &= c_2 - c_1. \end{aligned}$$

When these equations were applied to the estimated $c_g$ values in Table 3.17, it became apparent that the estimated value for $\tau_1$ was greater in value that the estimated value for $\tau_2$. This result held for both types of Rating Scale models. This analysis illustrates that the estimated values for the threshold values were not ordered along the latent continuum in correspondence with the category-numbers for the present example. This also makes clear that the raw total-score does not always indicate how many thresholds a subject has passed. The values for the thresholds in the UPRM are a function of both the $c_g$ and the $a_g$ parameters. If, however, these threshold values are calculated from the estimated parameters presented in Table 3.17, this model too shows that the estimated value for $\tau_2$ is less than the estimated value for $\tau_1$.

Andrich has shown that the Rating Scale model can be further restricted in a number of ways. As noted before, the Rating Scale model does not restrict the distances between the thresholds in any way. Andrich (1978a) showed how these distances could be restricted to equal unity. This restriction was generalized in Andrich (1982) where it was indicated how the distances between thresholds can be restricted to be equal to some arbitrary value (which is not necessarily unity). It can be shown that these types of restriction lead to specific constraints upon the $c_g$ parameters. In still another model, the Binomial Trials model (Andrich, 1978c), the passing of a threshold is treated as a binomial process. This model also imposes specific con- straints on the $c_g$ parameters. The details with respect to these restricted Rating Scale models are not discussed here. The interested reader is referred to the literature cited.

### 3.4 Evaluation

A great number of latent trait models were discussed in this chapter. The parameterization of these models was based on a discretized latent trait. Because the values of the discrete node points along the latent continuum were fixed in advance, a scaling parameter had to be included in the parameterization of some of these models.

The most general latent trait model discussed here was Bock's Nominal Response model. If it has a discretized latent trait, this model is equivalent to the rows- or columns-association models presented in Chapter 2. There is, in other words, a direct link between certain restricted latent class models and discretized latent trait models. The Nominal Response model links a metrical latent variable to nominal manifest indicators. No ordering information concerning the categories of the indicators is used in the model. However, if the categories are ordered, the estimated trace lines for the Nominal Response model should reflect this information and a substantive interpretation of the results can be based on the ordering of the categories. Because the Nominal Response model uses both discrimination parameters and difficulty parameters, it is a very flexible model. The drawback is that the results are less easy to interpret because, for example, thresholds depend on both discrimination and difficulty parameters.

Another important latent trait model is the Partial Credit model. This model assumes that both the latent and the manifest variables have been measured on an interval scale. Because the spacing of latent and manifest variables is fixed in advance, a scaling parameter must be introduced. The Partial Credit model with a common scaling parameter is equivalent to a restricted Uniform Association model, a fact that is significant for a number of reasons. Firstly, attention is thus focussed on the possibility of formulating less restricted models such as the Partial Credit model with item-specific scaling parameters (Heinen and Croon, 1992). This model is more flexible than the original Partial Credit model and can, therefore, be vital in the analysis of attitudinal data. Secondly, the parallel with the Uniform Association model makes it clear that the scaling parameters has a substantive meaning because it can be interpreted as an association coefficient of the relation between the latent and the manifest variables. The parameters of the Partial Credit model (with or without item-specific scaling parameters) can be interpreted rather easily using the threshold approach. One problematic aspect is the possibility that thresholds are not ordered properly, as became clear in the illustrating analyses in this chapter.

Using the threshold formulation, still more restricted models can be formulated by separating the threshold parameters into item-specific difficulty parameters and a set of category-specific threshold parameters that are invariant over items. The Rating Scale model proposed by Andrich (1978a) is the most well-known of these models. In the discretized versions of these models scaling parameters are needed. By allowing the scaling parameter to vary over items, a more general Rating Scale model can be obtained. Although the Rating Scale model is a rather direct and logical translation of the threshold approach in latent trait theory, it is a highly restrictive model. Therefore, the practical value of this model in the analysis of attitudinal data is limited.

The models discussed so far can be restricted by imposing equality restrictions on parameters. In this manner, convergence of trace lines over several items can be investigated. Several analyses were performed in this chapter to illustrate these possibilities. It can be stated that by using discretized latent traits highly fine-tuned analyses may be performed in order to gain further understanding of the relations between manifest indicators and the latent variable.

# Chapter 4

# Estimation and Testing in Latent Trait Models

## 4.1 Introduction

In Chapter 3 a number of latent trait models were introduced. All of these models are parameterized in terms of both person parameters (i.e., the latent ability $\theta_i$) and item parameters. The item parameters can be discrimination parameters, threshold parameters or difficulty parameters. These parameters can be estimated using a number of procedures.

In this chapter, three important maximum likelihood procedures that can be used in the context of item response theory are discussed. These methods were briefly mentioned in Chapter 3. The reason for restricting attention to maximum likelihood procedures and bypassing other estimation methods is that only maximum likelihood methods were discussed in Chapter 2. Because a comparison of latent trait and latent class models is the main theme in this study, it would seem reasonable to focus attention in this chapter on ML methods.

Firstly, *Joint Maximum Likelihood* estimation is reviewed (Section 4.2.1). This ML method estimates the person parameters and the item parameters simultaneously. Because the person parameters act as nuisance parameters, this procedure is hampered by a number of problems.

The second method dealt with is *Conditional Maximum Likelihood*. This procedure is discussed in Section 4.2.2. CML was developed in an attempt to overcome the problem of nuisance parameters by conditioning them out of the likelihood function. This can be accomplished if there are sufficient statistics for the person parameters. A natural consequence is that this procedure can only be used if known sufficient statistics for the person parameters exist, and this is not the case for all latent trait models discussed in Chapter 3. Furthermore, it is important to understand that CML is not merely a technical method of ridding the likelihood function of nuisance parameters. This procedure is also closely linked to Rasch's methodological claims of *specific objectivity*.

The third method, *Marginal Maximum Likelihood*, also uses a likelihood function in which the person parameters have been eliminated. MML, which is discussed in Section 4.2.3, is based on a number of assumptions about the distribution of $\theta$ in order to integrate the person parameters out of the likelihood function. If the complete density function of $\theta$ as well as the values of the parameters defining this distribution are specified by assumption, the estimation method is called parametric MML. A less restrictive method is to assume that only the type of the density function is known and to estimate the parameters of this density function from the data. This is likewise called parametric MML. It is, however, also possible to approach the unknown continuous density function by using a discrete distribution. This estimation method, called semi-parametric MML, makes it possible to link latent trait models directly to latent class analysis.

The fit of a specified latent trait model can be tested against empirical data with the same methods that were discussed in Chapter 2 for latent class models. In the context of latent trait models, it may be relevant to test hypotheses concerning specific assumptions. Examples are hypotheses about the sample independence of the item parameters or the unidimensionality of the latent trait. Section 4.3 discusses a number of tests that have been proposed in the literature on item response theory.

## 4.2 Estimation

Estimating the parameters in a latent trait model is a topic that has attracted a good deal of attention in the last 25 years. A number of estimation methods have been proposed. Fischer used a modified minimum chi-square estimation method in which a chi-square statistic based on the number of individuals that respond positively to item $j$ and negatively to item $j'$ was minimized (see Fischer, 1970 and Zwinderman, 1991a). Swaminathan and Gifford (1982) and Mislevy (1986) suggested the use of Bayesian methods (see Engelen, 1989 for a review of these Baysian methods). The most widely used method is maximum likelihood.[1] In this chapter only ML methods are discussed.

At first glance, the problem of estimating the parameters in item response models using ML methods would appear to be more complex than in latent class models because of the presence of ability parameters that may vary over subjects. One may be surprised by the fact that within the framework of ML methods there are at least three different procedures which are all claimed to result in maximum likelihood estimates. It is important to understand in what respect these procedures differ from each other. In the following an attempt is made to clarify the differences, as well as the relations, between these three ML procedures.

In latent class analysis, ML estimates can be obtained by using a number of different numerical procedures, among others, Newton-Raphson, Fisher's scoring algorithm, and the stabilized Newton-Raphson.[2] These procedures differ *numerically* but they all lead to the same estimates as they all maximize the same likelihood. As was shown in Chapter 2, this log-likelihood function can be written as

$$\begin{aligned}\ln L &= \sum_{\nu} f_\nu \ln p_\nu \\ &= \sum_{\nu} f_\nu \ln \left[ \sum_{t=1}^{T} p_{\nu|\theta_t} \cdot p_{\theta_t} \right] \\ &= \sum_{\nu} f_\nu \ln \left[ \sum_{t=1}^{T} \left( \prod_j \prod_g p_{jg|\theta_t}{}^{x_{\nu jg}} \right) \cdot p_{\theta_t} \right]. \qquad (4.1)\end{aligned}$$

The indicator variables $x_{\nu jg}$ indicate whether the response to item $j$ in response pattern $\nu$ falls in category $g$, so:

$x_{\nu jg} = 1$ if, in response pattern $\nu$, item $j$ is responded to with category $g$,
$x_{\nu jg} = 0$ otherwise.

[1] Bayesian methods also use maximum likelihood estimation. The ML methods discussed in this chapter are, however, "traditional", i.e., non-Bayesian methods.

[2] Stabilized Newton-Raphson is used in Haberman's NEWTON program.

If the EM algorithm is used, the likelihood is rewritten in a slightly different fashion because, in the M-step, the log-likelihood can be defined on the complete data-matrix. After all, the complete[3] latent $\times$ manifest data-matrix is estimated in the foregoing E-step. In this case, the expression for the log-likelihood function is simple:

$$\ln L = \sum_{\nu} \sum_{t} e_{\nu\theta_t} \cdot \ln p_{\nu t}, \tag{4.2}$$

in which $e_{\nu\theta_t}$ indicates the estimated number of individuals responding with pattern $\nu$ and belonging to latent class $\theta_t$.

However, EM also leads to the same maximum likelihood-estimates as the other numerical methods that are available in the context of latent class models. The two likelihood functions are equivalent, though they are expressed somewhat differently. Both contain parameters pertaining to the latent distribution, i.e., the $p_{\theta_t}$-parameters. In Newton-Raphson, scoring, or NEWTON, these parameters are estimated simultaneously with the other parameters, i.e., the response probabilities or the log-linear parameters on which the response probabilities are based. In the EM algorithm, the $p_{\theta_t}$-parameters are estimated separately in the M-step.

For latent trait models, the problem is more complex because the three different procedures that have been proposed are all attempts at maximizing a different likelihood function. The differences between these three methods are not merely numerical. This raises some serious questions because it is not clear which of the three methods should be preferred.

In the following sections, a global description of these different maximum likelihood methods is presented. A more detailed review can be found in Baker (1987). The question of which method should be preferred cannot be answered until the relations between the various methods have been explored. Recent work by Holland (1990b) and Lindsay, Clogg, and Grego (1991) has lead to new insights into the relations between the different estimation procedures.

---

[3]The data-matrix is incomplete because the membership of the latent classes is unknown.

### 4.2.1 Joint maximum likelihood

Joint maximum likelihood (JML) is a procedure in which item parameters and ability parameters are estimated simultaneously. The log-likelihood function that is to be maximized is equal to

$$\begin{aligned}\ln L &= \sum_{i=1}^{N} \ln p_{\nu|\theta_i} \qquad (4.3)\\ &= \sum_{i=1}^{N} \ln \left[ \prod_j \prod_g p_{jg|\theta_i}{}^{x_{ijg}} \right].\end{aligned}$$

The conditional probability that individual $i$ will respond with pattern $\nu$, provided that this individual has an ability parameter $\theta_i$, is denoted by $p_{\nu|\theta_i}$. Again, $x_{ijg}$ indicator variables are used in order to write the log-likelihood in a form that allows for polytomous data. In this fashion, the comparison between this log-likelihood function and the log-likelihood function for other maximum likelihood procedures as well as those presented in Chapter 2 is made easier. The $x_{ijg}$ indicator variables take the value 1 if subject $i$ responds in category $g$ of item $j$. Otherwise, the $x_{ijg}$ variable is equal to 0. The number of subjects in the sample is equal to $N$.

Because two individuals who have responded to the $n$ items with exactly the same response pattern $\nu$ have the same estimated ability parameter $\hat{\theta}_\nu$, the log-likelihood function can be written, somewhat more economically, as:

$$\begin{aligned}\ln L &= \sum_{\nu} f_\nu \ln p_{\nu|\theta_i} \qquad (4.4)\\ &= \sum_{\nu} f_\nu \ln \left[ \prod_j \prod_g p_{jg|\theta_\nu}^{x_{\nu jg}} \right],\end{aligned}$$

in which $p_{jg|\theta_\nu}$ indicates that the response probabilities are the same for individuals with the same response pattern $\nu$. The $x_{\nu jg}$ indicator variable indicates, as it did in Chapter 2, whether the response to item $j$ in response pattern $\nu$ falls in category $g$ ($x_{\nu jg} = 1$) or in another category ($x_{\nu jg} = 0$).

The development of the joint maximum likelihood estimation procedure was mainly the work of Lord (1967) and Birnbaum (1968). Although the log-likelihood function and the term "joint maximum likelihood" indicates a simultaneous estimation of item parameters and ability parameters, the procedure uses a two-stage iterative method. At one stage, the item parameters are held constant and new ability parameters are estimated. To solve identification problems, the mean of the $\theta$'s is fixed to a given value; for the two- and three-parameter model, the latent scale must also be fixed by restricting the variance of the $\theta$'s to be equal to some arbitrary value. Subsequently, the adjusted ability parameters are held constant and new estimates are obtained for the item parameters using a multivariate Newton-Raphson procedure. Details concerning these numerical methods as well as expressions for the likelihood equations that have to be solved in the case of dichotomous variables can be found in Baker (1987) and Hambleton and Swaminathan (1985).

The JML procedure is hampered by a number of problems, the most serious being the presence of "incidental" parameters. As the sample size increases, the number of $\theta_i$ parameters that must be estimated also increases. For this reason, the $\theta_i$ parameters are called incidental parameters. The item parameters are called "structural" parameters because their number does not increase if the sample size grows. The problem is that these structural parameters cannot be estimated consistently in the presence of incidental parameters. Within the context of JML it is impossible to get consistent estimators for the item parameters of the Rasch model. This problem was first studied by Neyman and Scott (1948) in a different context. Andersen (1973a) showed that the item difficulty parameters in the one-parameter model were not consistently estimated by using JML if the number of subjects approached infinity and the number of items was fixed to a given value. If, however, the number of subjects ($N$) and the number of items ($n$) approaches infinity, the JML parameter estimates for item and ability parameters are unbiased, if they exist at all (Haberman, 1977b).[4] In practice, these conditions are never fulfilled.

A second serious drawback to JML are the numerical problems that are often encountered with the estimation of parameters in the two- and three-parameter model. In a number of cases, estimates for the

[4] Haberman also specified a condition with respect to the relationship between $N$ and $n$. Details can be found in Haberman (1977b).

discrimination parameters can degenerate, i.e., become very large, and estimates for the guessing parameter in the three-parameter model can fall outside the region of acceptable values (i.e., the region ranging from zero to one). Finally, the estimation of $\theta$ parameters may become very cumbersome in the case of two- and three-parameter models.

These numerical problems do not appear in the one-parameter model as in this model score groups can be formed on the basis of sufficient statistics for the ability parameters (i.e., the total-scores for the subjects). In most situations the number of score groups is much smaller than the number of subjects. If just one ability parameter has to be estimated for each score group, the estimation problem can be simplified considerably. However, there are also a number of situations known in which the estimation of the item parameters for the one-parameter model is not possible by using JML. Sufficient and necessary conditions for the existence of a JML solution in the case of the one-parameter model are given in Fischer (1981).

In the past, JML was sometimes denoted as "unconditional maximum likelihood" to contrast the procedure with conditional "maximum likelihood" (see Wright and Douglas, 1977a, 1977b). However, the term "unconditional maximum likelihood has also been used by Bock and Lieberman (1970) and Bock (1972) to explain what later was called "marginal maximum likelihood". It is for this reason that the term unconditional maximum likelihood is not used in this study. De Leeuw and Verhelst (1986) denoted JML as "unconstrained maximum likelihood". This terminology corresponds to theoretical results obtained by Holland (1990b). This subject will be discussed again at a later point.

### 4.2.2 Conditional maximum likelihood

Conditional maximum likelihood (CML) depends on the availability of known sufficient statistics for the person parameters. The principle behind this method is very simple. By defining a log-likelihood function conditioning on the sufficient statistics for the ability parameters, these ability parameters vanish from the log-likelihood function and what is left is a function in which only item parameters appear. In the absence of the incidental parameters, the structural item parameters can be estimated consistently.

Whether CML can be used to estimate the item parameters in a specific latent trait model depends on the availability of known sufficient statistics for the person parameters. In the Rasch model (when the manifest indicators are dichotomous) and in the Partial Credit model and its derivative models (the manifest variables are polytomous) these sufficient statistics are equal to the total-score for an individual. [5] No additional assumptions are needed.

If, however, latent trait models with discrimination parameters are used (such as the Lord-Birnbaum model or Bock's Nominal response model), it must be assumed that these discrimination parameters are known a priori in order to estimate the parameters with CML. In these models, the sufficient statistics for the ability parameters $\theta_i$ are equal to the weighted total-scores, with the weights equal to the discrimination parameters (see, for example, Verhelst, 1992).

Despite this restriction, CML is widely used due to the popularity of the Rasch model and the Partial Credit model. The reason for this popularity is that these models have certain properties that relate to the presence of known sufficient statistics which, in a sense, make them unique. In order to properly understand these assertions, it is helpful to distinguish between two strategies that can be discerned in the context of item response theory. One objective is the construction of models with which the covariation that is present in a number of observed variables can be described. This perspective is rather pragmatic and it allows the inclusion of models that do not have known sufficient statistics for the ability parameters, such as the two- and three-parameter models suggested by Birnbaum and Bock's Nominal Response model. The second perspective is based on a number of measurement-theoretical considerations which are derived from the item response model. This tradition can be traced back to Rasch (1960) who introduced the notion of *specific objectivity*.

Specific objectivity has to do with the possibility of comparing any two items independently of all other item parameters and independently of the ability parameters. In the comparison between any two items, only the parameters characterizing those two items should play a role. This can be illustrated as follows. If individual $i$ responds to

[5] It is assumed for the models for polytomous data that the categories of the manifest indicators are scored with successive integers.

two dichotomous items $j$ and $j'$, two indicator variables, $x_{ij}$ and $x_{ij'}$, are defined. These variables take the value 1 if $i$ responds positively to an item; otherwise, the value of the indicator variable is 0. According to the Rasch model, the probability that $x_{ij}$ or $x_{ij'}$ equals 1 can be expressed as

$$\begin{aligned} p(x_{ij} = 1) &= \frac{\exp(\theta_i - b_j)}{1 + \exp(\theta_i - b_j)} \\ p(x_{ij'} = 1) &= \frac{\exp(\theta_i - b_{j'})}{1 + \exp(\theta_i - b_{j'})}. \end{aligned}$$

The total-score computed from these two items is denoted by $x_{i+}$, so $x_{i+} = x_{ij} + x_{ij'}$. This total-score can, of course, take the values 0, 1, or 2. Only the situation in which $x_{i+} = 1$ provides information with respect to the relative position of the item-difficulties $b_j$ and $b_{j'}$. It can easily be derived that the log odds for responding positively rather than negatively to item $j$, provided that $x_{i+} = 1$, depends only on the item parameters $b_j$ and $b_{j'}$:

$$\ln \left[ \frac{p(x_{ij} = 1 \mid x_{i+} = 1)}{p(x_{ij} = 0 \mid x_{i+} = 1)} \right] = b_j - b_{j'}$$

Because these log odds do not depend on the person parameter $\theta_i$, the relative position of the item difficulties can be estimated independently of the specific individuals that have responded to the items.

Similarly, way it can be shown that the comparison between two individuals who respond to a specific item only involves the ability parameters of those two subjects, not any other ability parameters nor any item parameters. The objectivity of the measurement process implied by these claims is called *specific* because the comparison takes always place within a certain frame of reference, i.e., a specific class of items and a specific class of subjects.

If a measurement process allows for the comparison of any two items independently of all other item parameters and all person parameters, and if that measurement process also allows for a comparison of any two individuals that involves only the ability parameters of those two individuals, specific objectivity is said to be present in the measurement process. This definition of specific objectivity is, however, rather vague. There are two methods of further defining specific objectivity.

One method is to translate the claims of independent comparisons into a measurement theoretical framework. It goes beyond the scope of this study to analyze the measurement theoretical analysis of specific objectivity in depth. The foundation for this approach was laid by Rasch (1977) and further elaborated upon by Irtel (1987). However, within the measurement theoretical framework, the claim of specific objectivity does not necessarily lead to the Rasch model. Analyses by Glas (1989) made clear that the two-parameter model proposed by Birnbaum, for example, also has the property of specific objectivity. Only if the demand is made that the parameterization be symmetrical, i.e., that the number of parameters used to characterize an item be equal to the number of parameters used to characterize the ability of a subject, does the Rasch model necessarily follow from the measurement theoretical definition of specific objectivity. If this claim of symmetric parameterization is not made, models other than the Rasch model can also possess the property of specific objectivity.

The second method of regarding specific objectivity is from a statistical point of view. In this often-used approach, specific objectivity is equated to the presence of sufficient statistics for the ability parameters independent of the item parameters. Within this interpretation, it can be shown that for dichotomous data the Rasch model is the only possibility if certain additional assumptions are made (Andersen, 1973a). This unique characteristic of the Rasch model, i.e., the fact that it can be derived from a number of assumptions among which the most important is the availability of sufficient statistics, has also been demonstrated by Fischer (1974).[6] The Rasch model can also be derived as a probabilistic variant of the Guttman scale. This derivation was given by Roskam and Jansen (1984). The relationship between the Guttman scale and the Rasch model, a topic which was already brought to attention in Chapter 2, was also explored by Andrich (1985).

If the measurement process is required to satisfy specific objectivity, i.e., the item response model has to allow for the existence of known sufficient statistics for both the item and the ability parameters, it can be shown that, in the case of dichotomous data, the Rasch model necessarily follows. Within this perspective, the Rasch model is more than

[6]The other assumptions made are rather straightforward, for example, the assumption of local independence, $\lim_{\theta \to -\infty} p_{jg|\theta_i} = 0$, and $\lim_{\theta \to \infty} p_{jg|\theta_i} = 1$.

a two-parameter model with the discrimination parameters restricted to be equal.

Similar results can be obtained for polytomous manifest variables. If the item response function has to be unidimensional, i.e., the ability can be characterized by a scalar, and if sufficient statistics for the ability parameters must exist, the model that results is equal to the Partial Credit model. This result was first obtained by Andersen (1977).[7] The Partial Credit model and all models that can be seen as restricted Partial Credit models are the only item response models that provide known sufficient statistics for the ability parameters when the manifest indicators are polytomous.

That the availability of sufficient statistics for the ability parameters makes it possible to eliminate these ability parameters from the likelihood function by conditioning on them is easily illustrated within the Rasch model.[8] The probability of responding in category 1 of item $j$ can be written as

$$p_{j1|\theta_i} = \frac{\exp(\theta_i - b_j)}{1 + \exp(\theta_i - b_j)}. \tag{4.5}$$

The probability that subject $i$ will respond with an arbitrary response pattern $\nu$ can now be expressed as

$$p_{\nu|\theta_i} = \frac{\exp(x_{i+} \cdot \theta_i - \sum_{j=1}^{n} x_{ij} \cdot b_j)}{\prod_{j=1}^{n} [1 + \exp(\theta_i - b_j)]}, \tag{4.6}$$

in which $x_{i+} = \sum_j x_{ij}$ is the total-score for subject $i$. Again, an indicator variable $x_{ij}$ is used which is 1 if the response of subject $i$ to item $j$ falls in category 1 and is otherwise equal to zero. The factorization theorem (see, for example, Mood, Graybill and Boes, 1974, pp. 307-308) makes it clear that the total-score is a sufficient statistic for the ability parameter $\theta_i$. Similarly, the number of subjects that answered item $j$ correctly, i.e., $x_{+j} = \sum_i x_{ij}$ is a sufficient statistic for the item parameter $b_j$.

[7] Although Andersen derived the Partial Credit model as the model that necessarily follows from the claim that sufficient statistics exist for the scalar-valued ability parameters, the naming "Partial Credit model" was introduced by Masters, who developed this model using the threshold approach. See Chapter 3 for details.

[8] This can also easily be shown for the Partial Credit model, the notation is then a bit more cumbersome. To facilitate the exposition only the Rasch model is dealt with. A through treatment of CML-estimation for the Partial Credit model is given in Masters (1982).

If $\varsigma$ denotes a total-score $(0 \leq \varsigma \leq n)$ and $\Omega_\varsigma$ is the subset of response patterns $\nu$ with $\varsigma_\nu = \varsigma$, then the probability that the total-score for subject $i$ will be equal to $\varsigma$ can be written as

$$\begin{aligned} p_{\varsigma|\theta_i} &= \sum_{\nu \in \Omega_\varsigma} p_{\nu|\theta_i} \\ &= \frac{\exp(\varsigma \cdot \theta_i) \times \sum_{\nu \in \Omega_\varsigma} \exp\left(-\sum_j x_{ij} \cdot b_j\right)}{\prod_{j=1}^n \left[1 + \exp(\theta_i - b_j)\right]} \\ &= \frac{\exp(\varsigma \cdot \theta_i) \times \gamma_\varsigma}{\prod_{j=1}^n \left[1 + \exp(\theta_i - b_j)\right]}, \end{aligned} \tag{4.7}$$

in which $\gamma_\varsigma = \sum_{\nu \in \Omega_\varsigma} \exp\left(-\sum_j x_{ij} \cdot b_j\right)$ are the elementary symmetric functions whose values depend solely on the total-score and the item parameters. The notation $\sum_{\nu \in \Omega_\varsigma}$ is used to indicate that the summation takes places only for those response patterns $\nu$ which result in a total-score equal to $\varsigma$. Finally, the probability that response pattern $\nu$ will be observed given that subject $i$ has a total-score equal to $\varsigma$ is

$$\begin{aligned} p_{\nu|\varsigma} &= \frac{p_{\nu|\theta_i}}{p_{\varsigma|\theta_i}} \\ &= \frac{\exp\left(-\sum_j x_{ij} \cdot b_j\right)}{\gamma_\varsigma}. \end{aligned} \tag{4.8}$$

The likelihood function that must be evaluated for a sample of $N$ subjects is obtained by

$$L = \prod_{i=1}^N p_{\nu|\varsigma} = \prod_\nu p_{\nu|\varsigma}^{f_\nu} \tag{4.9}$$

so that the expression for the log-likelihood becomes

$$\ln L = \sum_\nu f_\nu \ln p_{\nu|\varsigma}, \tag{4.10}$$

in which $p_{\nu|\varsigma}$ is defined as in Equation 4.8.

The CML log-likelihood function is expressed here by a summation of the response patterns $\nu$. It is also possible to let the summation take place over the total-scores $\varsigma$ which leads to an expression that is simpler to evaluate, but that cannot be directly compared to the log-likelihood functions obtained within the framework of JML or the third procedure, marginal maximum likelihood, which is described in the next section. It is for this reason that the CML convention of taking the sum over $\varsigma$ is not followed here.

The only identification problem that must be solved is fixing the location of the latent scale. This can be done by constraining one of the $b_j$ parameters to zero. Andersen (1970) showed that the CML estimators are then consistent, efficient, and asymptotically normally distributed. There are, however, a number of numerical drawbacks to the CML procedure. The evaluation of the elementary symmetric functions can be laborious, particularly if a large number of items are involved. Originally, it was believed that in practice CML could only be carried out without serious problems for analyses involving up to 50 items. Suggestions for solving the numerical problems involved in computing the elementary symmetric functions can be found in Verhelst, Glas, and van der Sluis (1984). The numerical approach suggested by these authors made possible the analysis of large data sets.

#### 4.2.2.1 CML estimation by log-linear models

CML estimates for the item parameters in the Rasch model can also be obtained by fitting somewhat non-standard log-linear models. The close connection that exists between the two types of models has been pointed out by various authors (Mellenbergh and Vijn, 1981; Tjur, 1982; Cressie and Holland, 1983; Kelderman, 1984, 1987; Duncan, 1984b; Hout, Duncan and Sobel, 1987). As a starting point, the expression for the probability of a certain response pattern $\nu$ given a total-score equal to $\varsigma$ is taken to be (see Equation 4.8)

$$p_{\nu|\varsigma} = \frac{\exp\left(-\sum_j x_{ij} \cdot b_j\right)}{\gamma_\varsigma}.$$

The Rasch model based upon this probability was called the "conditional Rasch model" in Kelderman (1984). This model can be written as a log- linear model:

$$\ln p_{\nu|\varsigma} = u_\varsigma + \sum_j x_{\nu j} \cdot u_j, \tag{4.11}$$

in which $u_\varsigma = \ln(1/\gamma_\varsigma)$, $u_j = -b_j$ and $x_{\nu j}$ is the element of the design matrix for pattern $\nu$ and item $j$. Because the vector of $x_{ij}$-values is the same for all individuals responding with the same response pattern $\nu$, the expression $x_{ij}$ used in Equation 4.8 is replaced by $x_{\nu j}$.

As Kelderman has shown, the estimates for the item parameters in this log-linear model are identical to the CML estimates for the Rasch model (Kelderman, 1984, appendix II). The conditional model assumes, however, that the total-score marginals $f_\varsigma$ are fixed by the sampling design. As this assumption is not fulfilled in most practical applications, a model in which the assumption of a fixed total-score distribution is not made would appear more attractive:

$$\begin{aligned} p_\nu &= p_{\nu|\varsigma} \times p_\varsigma \\ &= \frac{\exp\left(-\sum_j x_{ij} \cdot b_j\right)}{\gamma_\varsigma} \times p_\varsigma. \end{aligned} \tag{4.12}$$

The log-linear formulation now becomes

$$\ln p_\nu = u_\varsigma^* + \sum_j x_{\nu j} \cdot u_j, \tag{4.13}$$

in which $u_\varsigma^* = \ln(p_\varsigma/\gamma_\varsigma)$ and $u_j = -b_j$. Although this model does not assume that the totals $f_\varsigma$ are fixed by design, another problem arises because the $u_\varsigma^*$ parameter is now dependent on $p_{\varsigma|\theta_i}$, which, in turn, is dependent on the latent distribution of $\theta$. As was already shown, the latter probability can be expressed as

$$p_{\varsigma|\theta_i} = \frac{\exp(x_{i+} \cdot \theta_i) \times \gamma_\varsigma}{\prod_{j=1}^n [1 + \exp(\theta_i - b_j)]}.$$

When the latent variable $\theta$ is assumed to be continuous, the probability of obtaining a particular total-score can be written as

$$p_\varsigma = \gamma_\varsigma \int_{-\infty}^{+\infty} \frac{\exp(\varsigma \cdot \theta_i)}{\prod_{j=1}^n [1 + \exp(\theta_i - b_j)]}\, d\Phi(\theta), \tag{4.14}$$

in which $\Phi(\theta)$ is the cumulative distribution function of the latent variable.

Because $u_\varsigma^*$ is equal to $\ln(p_\varsigma/\gamma_\varsigma)$, the exponential of $u_\varsigma^*$ equals

$$\exp(u_\varsigma^*) = \int_{-\infty}^{+\infty} \frac{\exp(\varsigma \cdot \theta_i)}{\prod_{j=1}^n [1 + \exp(\theta_i - b_j)]}\, d\Phi(\theta).$$

Cressie and Holland (1973) showed that the elements to the right of the equal sign constitute a moment sequence. Because such a moment sequence must satisfy certain inequality restrictions,[9] these restrictions must be taken into account when estimating the parameters of the Rasch model using the log- linear model in 4.13. Expressions for these inequality restrictions can be found in Cressie and Holland (1973) and Kelderman (1984). Holland (1990a) showed how these restrictions influenced the values of the $u^*_\varsigma$ parameters. As the usual methods for estimating parameters in log-linear models cannot deal with this type of inequality restriction, the question arises whether this log-linear approach to the Rasch model holds any promise. Cressie and Holland (1973) criticized the uncritical use of the log-linear model for estimating parameters in the Rasch model. The estimated set of $\hat{p}_\nu$ probabilities do not have to be true manifest probabilities for the Rasch model. If the inequality restrictions are not met, it is possible to obtain estimated $\hat{p}_\nu$ parameters that could not be generated by any latent trait model.

The log-linear Rasch model in which the inequality constraints are ignored was termed the "extended Rasch model" by Tjur (1982). Tjur showed that even if the inequality constraints are not fulfilled, the estimates for the item parameters are equal to the conditional maximum likelihood estimates. This follows from the fact that the likelihood for this extended Rasch model can be factored in two parts (see Andersen and Madsen, 1977 and de Leeuw and Verhelst, 1986). One part corresponds to the usual conditional likelihood, the other to the $u^*_\varsigma$ parameters. This can easily be shown as follows:

$$\begin{aligned}
\ln L &= \sum_\nu f_\nu \cdot \ln p_\nu \\
&= \sum_\nu f_\nu \cdot \ln(p_{\nu|\varsigma} \times p_\varsigma) \\
&= \sum_\nu f_\nu \cdot \ln p_{\nu|\varsigma} + \sum_\nu f_\nu \cdot \ln p_\varsigma \\
&= \sum_\nu f_\nu \cdot \ln p_{\nu|\varsigma} + \sum_\varsigma f_\varsigma \cdot \ln p_\varsigma. \qquad (4.15)
\end{aligned}$$

[9]For example, the second moment must be greater than or equal to the square of the first moment, otherwise the variance will be negative.

The first part, on the right hand of this expression, is equal to the log-likelihood for CML (see Equation 4.10). The second part pertains to the total-score distribution. In the log-linear parameterization the $u^*_\varsigma$ parameters guarantee that the total-score distribution is precisely reproduced. As is clear from Equation 4.14, however, both the distribution of the latent variable and the item parameters play a role in this second part of the likelihood. The fact that the latent proportions $p_\varsigma$ are also a function of the item parameters has some consequences with respect to the relationship between CML estimators and semi-parametric MML estimators. This is the subject of discussion in Section 4.2.3.

Thus, in the extended Rasch model, the estimated item parameters are equal to the CML estimates for those parameters. However, when the inequality restrictions are not met, the estimated probabilities $\hat{p}_\nu$ do not necessarily have to conform to the Rasch model. Cressie and Holland stress the fact that, despite these complications, the extended Rasch model may have some practical use. If the extended Rasch model does not provide an acceptable fit for a given set of data, the Rasch model will, likewise, not fit the data and, therefore, more complex models will be needed to explain the association between the variables. On the other hand, if the data conform to the Rasch model, they will also fit the extended Rasch model. As stated before, the reverse is not necessarily true.

To illustrate the use of log-linear models in estimating the item parameters of the Rasch model, the data analyzed in the previous chapters were again used. Before turning to these analyses, however, some remarks have to be made about the identifying restrictions that are needed in the context of these log-linear models. To clarify this, the log-linear model in 4.13 has been rewritten below using the Kronecker-delta notation that was also used by Cressie and Holland:

$$\ln p_\nu = \sum_{\varsigma=0}^{n} u^*_\varsigma \cdot \delta(\varsigma, x_{\nu+}) - \sum_{j=1}^{n} x_{\nu j} \cdot u_j,$$

in which $x_{\nu+} = \sum_j x_{\nu j}$ and $\delta(\varsigma, x_{\nu+})$ is defined by:

$$\delta(\varsigma, x_{\nu+}) = \begin{cases} 1 & \text{if } x_{\nu+} = \varsigma \\ 0 & \text{if } x_{\nu+} \neq \varsigma. \end{cases}$$

At first sight, there appear to be $2n+1$ parameters in this log-linear model. However, the actual number of parameters is smaller because of the presence of some linear restrictions. Firstly, it must be noted that

$$\sum_{\varsigma=0}^{n} \delta(\varsigma, x_{\nu+}) = 1,$$

so the number of independent $\delta$-functions is $n$ instead of $n+1$. Furthermore

$$\sum_{j=1}^{n} x_{\nu j} = \sum_{\varsigma=1}^{n} \varsigma \cdot \delta(\varsigma, x_{\nu+}),$$

which indicates the need for another identifying restriction, for example, $x_{\nu 1} = 0$ or $x_{\nu n} = 0$ for all $\nu$, which will fix the item parameter for the first or the last item at a value of zero. This restriction is clearly necessary for fixing the scale of the latent variable.

The number of parameters to be estimated is equal to $2n-1$; therefore, the number of degrees of freedom is equal to

$$df = 2^n - 1 - (2n-1) = 2^n - 2n.$$

The parameters of the log-linear Rasch model were estimated for the data on women's liberation that were also used in the previous chapters. Again, only the data for individuals who satisfied the educational requirements formulated in Chapter 2 were used in the analysis.

The analyses were performed with Haberman's FREQ program. The log-linear Rasch model for the $2^n$-table can be estimated with any log-linear program that enables the user to specify a design matrix. The reason for this is that the $u^*_\varsigma$ interaction parameters that appear in 4.13 are non-standard. Another possibility, pointed out by Kelderman, is to estimate a log-linear model for the $2^n \times (n+1)$-table formed by combining the $2^n$ different response patterns with the $(n+1)$ total-scores. Of course, this table has a large number of cells containing structural zeros. The log-linear Rasch model can be estimated in such a table by fitting the quasi-independence model.[10] Both methods of estimating the parameters in the log-linear Rasch model are, for practical reasons, restricted to situations in which the number of items is not too large.

[10]Information concerning quasi-independence models can be found in standard text books such as Bishop et al. (1975) and Hagenaars (1990).

Kelderman (1987) proposed a "marginalization-by variable" algorithm that can be used when the number of items is large.

Table 4.1 provides the values of the log-likelihood ratios and the Pearson $\chi^2$-statistics that were obtained in the analyses for the total sample, as well as the separate analyses for male and female respondents. Table 4.1 clearly shows that the test results were not satisfactory. The hypothesis that the Rasch model holds in the population clearly had to be rejected for the total sample. The values of the test statistics were smaller in the male and the female subsamples, but that was due to the fact that the number of observations in the subsamples were, of course, smaller than in the total sample. The Rasch model could not be accepted in these subsamples either.

Table 4.1: Testing results for the log-linear Rasch model

| | Pearson $\chi^2$ | $p$ | $G^2$ | $p$ | $df$ |
|---|---|---|---|---|---|
| Male | 42.51 | .000 | 42.98 | .000 | 22 |
| Female | 43.34 | .000 | 41.92 | .000 | 22 |
| Total | 53.30 | .000 | 51.77 | .000 | 22 |

In Chapter 2, as was already noted, it can be seen that the ordering of the items according to their difficulty was not the same for the two subsamples by inspecting the marginal distributions of the items. The same phenomenon was found in the estimated item difficulties for the two subsamples. These estimates are reported in Table 4.2. The results regarding the ordering of the items for male and female respondents were identical to those obtained in Chapter 2 using marginal

Table 4.2: Estimated item difficulties for the Rasch model; log-linear estimates

| | Male | Female | Total |
|---|---|---|---|
| Item 1 | −0.95 | −0.17 | −0.56 |
| Item 2 | −2.21 | −1.53 | −1.86 |
| Item 3 | −0.74 | +0.34 | −0.19 |
| Item 4 | −3.89 | −3.78 | −3.77 |
| Item 5 | 0.00[a] | 0.00[a] | 0.00[a] |

[a]Fixed according to identifying restrictions

distributions. For the male respondents, item 5 was the most difficult item, followed by item 3, whereas the reverse was true for the female respondents.

The item difficulties reported in Table 4.2 are identical to the CML estimates. However, as was mentioned earlier, the log-linear Rasch model does not guarantee that the $p_\nu$ response pattern probabilities will conform to the Rasch model. Cressie and Holland (1973) formulated both the necessary and the sufficient conditions in terms of the $u^*_\varsigma$ parameters. When the sufficient conditions are met, the $p_\nu$-probabilities will conform to the Rasch model. Table 4.3 provides the estimated $u^*_\varsigma$ parameters.

Table 4.3: Estimated $u^*_\varsigma$-parameters in the log-linear Rasch model

| $\varsigma$ | Male | Female | Total |
|---|---|---|---|
| 0 | 0.00[a] | 0.00[a] | 0.00[a] |
| 1 | −3.43 | −3.05 | −3.18 |
| 2 | −5.27 | −4.59 | −4.83 |
| 3 | −6.65 | −5.01 | −5.73 |
| 4 | −7.07 | −4.56 | −5.71 |
| 5 | −6.76 | −3.11 | −4.78 |

[a]Fixed according to identifying restrictions

It can be verified from the obtained values that the sufficient conditions were met. This was true for both the total sample and for the male and female subsamples. The estimated parameters in Table 4.2 can therefore be interpreted as the CML estimates of the item difficulties in a Rasch model. The exact definition of the inequality constraints that must be satisfied as formulated by Cressie and Holland (1973) will not be given, as they have already been reformulated by Lindsay et al. (1991) in a way that is both more informative and easier to interpret. These results will be discussed at a later point.

### 4.2.3 Marginal maximum likelihood

The two maximum likelihood procedures discussed so far each dealt with the simultaneous presence of item and person parameters in their own fashion. JML does not provide a solution to this problem; the presence of person parameters is taken for granted. Ability and item parameters are estimated jointly, although it is known that the estimators are inconsistent. Estimation of ability is facilitated by forming homogeneous ability groups on the basis of identical total-scores or identical response patterns. It is for this reason that JML is labeled a fixed-effects model, i.e., both persons and items are treated as fixed.

Holland (1990a) pointed out that the JML approach starts from what he called the *stochastic subject rationale*. The assumption formalized in the expression for $p_{jg|\theta_i}$, that the response of a person $i$ with ability $\theta_i$ to item $j$ is probabilistic in nature has to do with the fact that human behavior is inherently unpredictable. The uncertainty which is characteristic of human behavior is modeled by $p_{jg|\theta_i}$, which denotes the probability that an individual with latent score $\theta_i$ will answer in category $g$ when confronted with item $j$. As Holland (1990a) noted, it was this interpretation of $p_{jg|\theta_i}$ which led to the joint maximum likelihood function 4.3. The outcome of the measurement process when stochastic subject $i$ meets item $j$ is ruled by the probability function $p_{jg|\theta_i}$, and the likelihood function 4.3 states that these probability functions are all statistically independent. In item response theory, this approach is called the *fixed-score* model or the *functional* model (de Leeuw and Verhelst, 1986). In the context of this model, it does not make sense to talk about the distribution function of the latent variable because the distribution of the latent variable is a point distribution. The persons are regarded as fixed and the person parameters are estimated jointly with the item parameters.

In CML, the problem of the simultaneous presence of item and person parameters is solved by using sufficient statistics for the person parameters. By conditioning on the sufficient statistic for the person parameter, i.e., the total-score, the person parameters can be eliminated from the likelihood function and consistent estimates can be obtained for the item difficulties. The major drawback of this method is that it can only be applied to item response models in which known sufficient statistics for the person parameters are available. This puts limits on the range of latent trait models that can be used.

After the item parameters have been estimated, estimation of the latent ability can still take place using standard maximum likelihood procedures. In practice, this means that the JML likelihood is used in which the item difficulty parameters have been replaced by their CML estimates. Again, this procedure for estimating the $\theta_i$ parameters does not depend on assumptions concerning the latent distribution. This method yields maximum likelihood point estimates of the latent abilities for each response pattern except the two most extreme patterns, i.e., the response pattern in which the answers to all of the items are in the lowest category and the pattern in which all of the answers are in the highest category. It is, of course, assumed that the response categories for the manifest indicators can be ordered. The fact that point estimates are obtained for the latent ability parameters clearly illustrates that it is not necessary to speak of a latent distribution in the context of CML estimation. In this sense, CML can also be used in the context of the fixed-effects or functional model.

When CML is used, it is also possible to include certain hypotheses about the latent population distribution in the estimation procedure. CML can, therefore, be used to estimate the parameters in what de Leeuw and Verhelst (1986) called the *random-score* model or the *structural* model. In this model, subjects are regarded as being sampled at random from a population which is characterized by a certain distribution of the latent variable $\theta$.

Holland (1990a) stated that this model is based on another substantive interpretation of $p_{jg|\theta_i}$. That quantity is now regarded as the proportion of subjects in the population with latent score $\theta_i$ that will respond in category $g$ of item $j$. The stochastic nature of the relationship between the latent variable and the manifest indicators is not the result of inherently unpredictable human behavior, but stems from the fact that subjects are sampled at random from a population with a specified latent distribution. Holland called this interpretation of $p_{jg|\theta_i}$ the *random sampling rationale.* The log-linear Rasch model which does not condition on the total-score is an example of this structural model. As was already shown, the likelihood for this log-linear Rasch model can be written as

$$\ln L = \sum_{\nu} f_\nu \cdot p_\nu$$

and this likelihood can be factored in two parts:

$$\ln L = \sum_i \ln p_{\nu|\varsigma_i} + \sum_i \ln p_{\varsigma_i}.$$

The first part, on the right-hand side, is the likelihood that is used in CML. Andersen and Madsen (1977) proposed estimating the parameters of the latent distribution using the second part of the likelihood. Equation 4.14 makes clear that the probabilities that appear in this part of the log-likelihood, $p_\varsigma$, depend on the latent distribution in the population. The density function $\phi(\theta)$ of this latent variable is equal to the first derivative of $\Phi(\theta)$ in 4.14, assuming that this first derivative exists. Andersen and Madsen (1977) proposed taking a given known density function for $\phi(\theta)$, for example, the normal density or the log-beta density. Using the part of the log-likelihood that is based on $p_\varsigma$ and substituting the CML estimates for the item difficulties in this likelihood function, standard ML methods can be employed to estimate the parameters that describe the assumed density function for the latent variable.

Marginal maximum likelihood (MML) makes explicit use of the density function of the latent variable $\phi(\theta)$ in order to end up with a likelihood function which does not contain any $\theta_i$ person parameters. This is achieved by maximizing the likelihood function:

$$L = \prod_{i=1}^{N} p_{\nu i} = \prod_{\nu} p_{\nu}^{f_\nu}, \tag{4.16}$$

in which the probability for observing response pattern $\nu$ can be expressed as

$$p_\nu = \int p_{\nu|\theta_i} \, d\Phi(\theta).$$

In this fashion, the person parameters $\theta_i$ are integrated out.[11] The resulting likelihood function is based on the probabilities $p_\nu$ which reflect the marginal distribution of the response patterns $\nu$, hence the name *marginal maximum likelihood.*

MML was first proposed by Bock and Lieberman (1970), though the numerical procedures developed in that article were not very workable, owing to their complexity. A major breakthrough was made in an article by Bock and Aitkin (1981), in which it was pointed out that parameter

[11] The Stieltjes form of the integral was used in the notation, a common convention in the literature on this topic. If the cumulative distribution function $\Phi(\theta)$ is differentiable, $d\Phi(\theta)$ can be replaced by the more known expression $\phi(\theta)\, d\theta$, in which $\phi(\theta)$ is the first derivative of $\Phi(\theta)$ and the density function of $\theta$.

estimation using the MML function can be dealt with numerically by applying the EM algorithm (Dempster, Laird and Rubin, 1977). Before going into further detail, some comments have to be made regarding the choice of the density function $\phi(\theta)$. This topic is discussed in the following sections.

#### 4.2.3.1 Parametric estimation

The major problem with MML is that something must be known about the density function $\phi(\theta)$ in order to be able to evaluate the integral in the expression for $p_\nu$. The most rigorous solution to this problem is to assume that the distribution of the latent variable $\theta$ is completely known. For example, one may assume that $\phi(\theta)$ is the density of a standard normal variate. In that case, the integral can be evaluated numerically using Gauss-Hermite quadrature. The continuous normal variable is then discretized on a certain interval and the integration is substituted by a summation:

$$p_\nu = \sum_{k=1}^{q} p_{\nu|X_k} \cdot A(X_k).$$

The optimal set of the $q$ different quadrature points $X_k$ and their corresponding weights $A(X_k)$ can be found in Stroud and Secrest (1966). The set is optimal in the sense that the approximation to the integral is optimal when the quadrature points and their weights are used in this summation. It should be noted that when this procedure is used, identifiability restrictions which fix the scale of the latent variable are no longer needed. In the Rasch model, the value of the common discrimination parameter $a$ has to be estimated (see Thissen, 1982, for details on MML estimation in the Rasch model).

As an example, the data on women's liberation were again analyzed using a Rasch model. In order to obtain results that were comparable with the results in Tables 4.1 and 4.2, the analyses were performed separately for the male and female subsamples, as well as for the total sample.[12] The Rasch model was subsequently fitted with a MML procedure under the assumption that the latent variable followed a normal

[12] Again, only respondents satisfying the educational constraints were retained in the analysis.

distribution in the population. The calculations were performed using Thissen's Multilog program (Thissen, 1988). A normal distribution was assumed for the latent variable. In the estimation process, ten equidistant quadrature points were used ranging from −4.5 to +4.5 at distances of 0.5 point from each other. Because the assumption of a normal distribution for the latent variable set a metric for the latent scale that could not be compared directly to the results of an analysis using CML estimation, the estimated MML item difficulties were standardized so that they had the same mean and standard deviation as the CML estimates. The results of both estimation procedures are given in Table 4.4.

Table 4.4: Estimated item difficulties for the Rasch model; a comparison of CML estimates with parametric MML estimates

| | Male | | Female | | Total | |
|---|---|---|---|---|---|---|
| | CML | MML | CML | MML | CML | MML |
| Item 1 | −0.95 | −0.96 | −0.17 | −0.18 | −0.56 | −0.70 |
| Item 2 | −2.21 | −2.27 | −1.53 | −1.59 | −1.86 | −1.97 |
| Item 3 | −0.74 | −0.74 | +0.34 | +0.38 | −0.19 | +0.12 |
| Item 4 | −3.89 | −3.86 | −3.78 | −3.75 | −3.77 | −3.69 |
| Item 5 | 0.00[a] | +0.03 | 0.00[a] | +0.01 | 0.00[a] | −0.14 |
| $G^2$ | 42.98 | 47.7 | 41.92 | 46.3 | 51.77 | 57.2 |
| *df* | 22 | 25 | 22 | 25 | 22 | 25 |

[a]Fixed according to identifying restrictions

The results made clear that the CML estimates resulted in a slightly better fit than the parametric MML estimates, though this was at the cost of a loss of three degrees of freedom. The parametric MML estimation procedure imposed some extra constraints on the distribution of the latent variable, whereas with CML the latent distribution was not even a part of the likelihood that was maximized. This difference between the two estimation procedures was also reflected by the difference in degrees of freedom. Furthermore, the differences between the estimated item difficulties in the two estimation schemes became negligible when attention was given to the male and female subsamples separately. The differences were somewhat larger, however, for the total sample. The overall picture gained by this comparative analysis is, however, apparent: the two sets of estimated item difficulty parameters have a distinct resemblance. The two estimation procedures provide highly comparable, though not identical, results.

Another less restrictive solution is to assume that the density function $\phi(\theta)$ belongs to a given parametric family of distributions, and to estimate the parameters describing $\phi(\theta)$. It could be assumed, for instance, that the latent variable $\theta$ is normally distributed with unknown mean $\mu$ and unknown variance $\sigma^2$. Subsequently, the distribution parameters $\mu$ and $\sigma$ can be estimated along with the item parameters.[13]

Within this tactic of estimation, two different approaches can be distinguished. Firstly, it is possible to estimate the parameters describing the population distribution by using the MML function in 4.16, provided that the item parameters are known. It is immaterial whether these parameters were estimated using CML or obtained from a previous calibration sample. This approach is similar to the one suggested by Andersen and Madsen (1977), but it should be noted that the likelihood function which is used to estimate the parameters defining $\phi(\theta)$ is different. If the item parameters are known, this approach of estimating the parameters defining the latent distribution can also be used for the two- and three-parameter models (this is not true of the Andersen/Madsen solution). Examples of this estimation procedure can be found in Mislevy (1984). In this article, several parametric possibilities are presented for $\phi(\theta)$, such as the normal distribution, a mixture of normal components, and the beta-binomial.

An alternative approach is to estimate the item parameters and the parameters describing $\phi(\theta)$ simultaneously. Examples can be found in Rigdon and Tsutakawa (1983) and Glas (1989).

Regardless of whether the density function $\phi(\theta)$ is considered completely known or $\phi(\theta)$ is assumed to belong to some parametric family with unknown parameters that have to be estimated, the MML estimation is said to be parametric. Within the framework of MML it is,

[13] Whether both $\mu$ and $\sigma$ can be estimated depends on whether sufficient restrictions are imposed on the item parameters to fix the location and/or the metric of the latent scale.

however, also possible to employ semi-parametric estimation.[14] Semi-parametric estimation is discussed in the following sections.

#### 4.2.3.2 Semi-parametric estimation

There are two possible strategies that can be used in order to execute semi-parametric estimation. Both methods are characterized by the discretization of the latent scale at a number of node-points. One possibility is to fix a number of node-points on the latent axis and to assign particular values to these node-points. If this tactic is chosen, estimates must be obtained for the item parameters and the latent proportions which belong to the fixed latent node points. The weights $A(X_k)$ are no longer fixed, but have to be estimated from the data. This procedure will be called semi-parametric estimation. As was illustrated in the examples in Chapter 3, the discretization of the latent trait and the assumption that the values of the latent nodes are known results in latent trait models that are identical to certain restricted latent class models. Therefore, within the framework of the semi-parametric MML approach it is possible to bridge the gap between latent trait models and latent class analysis.

Another possibility is to estimate not only the latent proportions but also the values of the latent nodes. In this case the entire latent distribution is free and has to be estimated. Only the number of latent node points are fixed in advance. This will be called fully semi-parametric estimation. In this section, semi-parametric estimation with fixed node-points is discussed.

As a starting point is taken the expression for the $p_\nu$ probabilities.

$$p_\nu = \sum_{t=1}^{T} p_{\nu|\theta_t} \cdot p_{\theta_t}.$$

The structure of this expression is the same as the one that is used

[14]The estimation methods discussed here under the heading semi-parametric estimation have, in a number of articles, been defined as non-parametric. However, it should be noted that the relationship between the latent variable and the manifest indicators is still assumed to be parametric, i.e., this relationship is assumed to follow the logistic distribution. Here, the term non-parametric will be reserved for those procedures in which $p_{jg|\theta_t}$ is also permitted to be non-parametric. Examples of this approach can be found in Holland and Rosenbaum (1986) and Croon (1990, 1991a, 1991b).

for the Gauss-Hermite quadrature. The difference is that the values $\theta_t$ and the number $T$ of these values can be chosen freely (within certain constraints) by the researcher. The second and most important difference is that the $A(X_k)$ weights are now replaced by the $p_{\theta_t}$ probabilities which are estimated using the empirical data. The form of the latent distribution is only fixed in advance in the sense that the latent distribution can be approximated by a discrete distribution with a specified number of node-points. This procedure is extremely flexible and it allows for the application of MML to a great variety of item response models. In Chapter 3, this semi-parametric MML approach was used extensively.

The parameters in this model (i.e., the item parameters and the latent proportions $p_{\theta_t}$) can be estimated using several algorithms. It is possible to use Haberman's *NEWTON* program which incorporates the stabilized Newton-Raphson procedure. The application of this program to models containing linear restrictions on the relations between the latent and the manifest variables did not, however, prove satisfying. The algorithm is sensitive to the choice of initial estimates and for some data sets it proved very difficult to obtain convergence. The simultaneous estimation of item parameters and the latent probabilities $p_{\theta_t}$ is, on the other hand, rather straightforward if the EM algorithm is used. The estimation problem at hand is similar to the estimation problems that were dealt with in Chapter 2, i.e., the estimation of parameters in the latent class model. In the following, the same notation is used as was in Chapter 2 in order to maintain consistency in the expressions. As the essentials of the EM algorithm were already explained in Chapter 2, only the main steps will be indicated here.

In the E-step, the frequencies for the "complete" data-matrix $(e_{\nu\theta_t})$ are estimated (see Equation 2.30). Estimation of the item parameters can then take place in the M-step by maximizing a likelihood function that is based on this complete data-matrix. The expression for this likelihood function is the same as the one used for latent class analysis:

$$L = \prod_{\nu} \prod_{t=1}^{T} [p_{\nu\theta_t}]^{e_{\nu\theta_t}}$$

$$= \left(\prod_t \prod_j \prod_g p_{jg|\theta_t}{}^{q_{jg\theta_t}}\right) \times \prod_t p_{\theta_t}^{e_{+\theta_t}}.$$

In this expression, the symbol $e_{+\theta_t}$ denotes the frequency with which individuals belong to latent "class" $\theta_t$. Thus,

$$e_{+\theta_t} = \sum_\nu e_{\nu\theta_t}.$$

The number of individuals that belong to latent "class" $\theta_t$ and respond in category $g$ of item $j$ is indicated by $q_{jg\theta_t}$, which is defined as

$$q_{jg\theta_t} = \sum_\nu e_{\nu\theta_t} \cdot x_{\nu jg}.$$

The variable $x_{\nu jg}$ is, as before, an indicator variable that takes the value 1 if the response to item $j$ in response pattern $\nu$ falls in category $g$. Otherwise, this indicator variable is 0.

Again, it is clear that the latent proportions can be estimated independently of the item parameters that define the response probabilities $p_{jg|\theta_t}$. This means that the item parameters can be estimated using just the first part of this likelihood function. Therefore, the ML estimates for the item parameters are those values that maximize

$$\ln L = \sum_t \sum_j \sum_g q_{jg\theta_t} \cdot \ln p_{jg|\theta_t}.$$

Because the item response models dealt with are all defined by the logistic distribution, the M-step is equivalent to estimating parameters in an ordinary logistic regression problem. Below, the Nominal Response model is taken as an example because, as was shown in Chapter 3, this model is the most general item response model used in this study. The response probabilities for this model are defined as

$$p_{jg|\theta_t} = \frac{\exp[a_{jg} \cdot \theta_t + c_{jg}]}{\sum_{h=0}^{m_j} \exp[a_{jh} \cdot \theta_t + c_{jh}]}.$$

Substituting this expression in the likelihood function results in

$$\ln L = \sum_t \sum_j \sum_g \left[q_{jg\theta_t} \cdot a_{jg} \cdot \theta_t + q_{jg\theta_t} \cdot c_{jg} - q_{jg\theta_t} \cdot \ln S_{j\theta_t}\right],$$

in which $S_{j\theta_t}$ is defined as

$$S_{j\theta_t} = \sum_{h=0}^{m_j} \exp(a_{jh} \cdot \theta_t + c_{jh}).$$

The parameters $a_{jg}$ and $c_{jg}$ can be estimated using Newton-Raphson. Hence, the first and second derivatives of the log-likelihood with respect to these parameters are needed. The expressions for the first derivatives can readily be obtained:

$$\begin{aligned}
\frac{\partial \ln L}{\partial a_{jg}} &= \sum_t [q_{jg\theta_t} - e_{+\theta_t} \cdot p_{jg|\theta_t}] \cdot \theta_t \\
\frac{\partial \ln L}{\partial c_{jg}} &= \sum_t [q_{jg\theta_t} - e_{+\theta_t} \cdot p_{jg|\theta_t}].
\end{aligned}$$

Using these expressions, it is also easily verified that the second derivatives are equal to the following expressions:

$$\begin{aligned}
\frac{\partial^2 \ln L}{\partial a_{jg}^2} &= -\sum_t e_{+\theta_t} \cdot p_{jg|\theta_t} \cdot (1 - p_{jg|\theta_t}) \cdot \theta_t^2 \\
\frac{\partial^2 \ln L}{\partial c_{jg}^2} &= -\sum_t e_{+\theta_t} \cdot p_{jg|\theta_t} \cdot (1 - p_{jg|\theta_t}) \\
\frac{\partial^2 \ln L}{\partial a_{jg} \partial a_{jh}} &= \sum_t e_{+\theta_t} \cdot p_{jg|\theta_t} \cdot p_{jh|\theta_t} \cdot \theta_t^2 \\
\frac{\partial^2 \ln L}{\partial c_{jg} \partial c_{jh}} &= \sum_t e_{+\theta_t} \cdot p_{jg|\theta_t} \cdot p_{jh|\theta_t} \\
\frac{\partial^2 \ln L}{\partial a_{jg} \partial c_{jg}} &= -\sum_t e_{+\theta_t} \cdot p_{jg|\theta_t} \cdot (1 - p_{jg|\theta_t}) \cdot \theta_t \\
\frac{\partial^2 \ln L}{\partial a_{jg} \partial c_{jh}} &= \sum_t e_{+\theta_t} \cdot p_{jg|\theta_t} \cdot p_{jh|\theta_t} \cdot \theta_t.
\end{aligned}$$

It should be noted that the second derivatives with respect to parameters pertaining to different items (such as $\frac{\partial^2 \ln L}{\partial a_{jg} \partial a_{j'g}}$) are equal to zero for the Nominal response model. Hence, the matrix of second derivatives has a block-diagonal structure. Each block pertains to the second derivatives with respect to the parameters of one specific item.

With these first and second derivatives the provisional estimates for $a_{jg}$ and $c_{jg}$ can be improved at each M-step. Because of the block-diagonal structure of the matrix of second derivatives, the estimation of the item parameters can be carried out for each item separately. However, for a number of other item response models this is not the case. When equality constraints are imposed on parameters pertaining to different items, the parameters for these items have to be estimated simultaneously.

For the other item response models discussed in Chapter 3, the necessary first and second derivatives can be derived in a similar way. The resulting expressions can be found in Appendix D.

Using the results from Appendix A, it is possible to obtain estimates for the variance-covariance matrix of the estimated parameters. When one is interested in estimating this matrix via the inverse of the expected information matrix, then expressions are needed for the first derivatives of $p_\nu$ with respect to the item parameters $a_{jg}$ and $c_{jg}$ and also the first derivatives of $p_\nu$ with respect to the latent proportions $p_{\theta_t}$. The latter are equal to the expressions that were already given in Appendix A for the general latent class model. The first derivatives for the item parameters are, of course, dependent of the specific type of item response model. For Bock's nominal response model the following results are readily obtained:

$$\begin{aligned}
\frac{\partial p_\nu}{\partial a_{jg}} &= \sum_t p_{\theta_t} \cdot p_{\nu|\theta_t} \cdot \theta_t (x_{\nu jg} - p_{jg|\theta_t}) \\
\frac{\partial p_\nu}{\partial c_{jg}} &= \sum_t p_{\theta_t} \cdot p_{\nu|\theta_t} (x_{\nu jg} - p_{jg|\theta_t}).
\end{aligned}$$

Although one of the main advantages of this semi-parametric approach is that it is possible to estimate parameters for models without known sufficient statistics, the procedure will be illustrated in this section by fitting again the Rasch model to the data on women's lib. In this way a comparison can be made with the CML procedure as well as with other MML procedures. First, the expression for the conditional probability that item $j$ will be responded to in category 1 will be given once more. In order to take account of the fact that the latent continuum is discretized, this probability will be denoted by $p_{jg|\theta_t}$ instead of by $p_{jg|\theta_i}$. The expression for the response probability in the discretized

Rasch model is:

$$p_{j1|\theta_t} = \frac{\exp(\alpha \cdot \theta_t + c_j)}{1 + \exp(\alpha \cdot \theta_t + c_j)}.$$

Note that the common slope parameter $\alpha$ is used here as a scaling constant that catches up with different choices for the set of $\theta_t$ values. So the value of $\alpha$ must be estimated. The $c_j$ parameters are equal to the negative of the item difficulties. Because the location of the latent scale is fixed by the choice of the $\theta_t$ values, all five item parameters $c_j$ can be estimated. When the number of node points that is used equal $T$, there are $T-1$ independent latent proportions to be estimated. The number of degrees of freedom for this model is then equal to:

$$df = 2^n - n - T - 1.$$

We are left of course with the problem of choosing a specific number of node points as well as choosing the values that must be assigned to these node points. The problem of choosing a specific number of node points is of course equivalent to choosing a specific number of latent classes in latent class analysis. Noting this equivalence does not solve the problem, however. It is known that in LCA the number of latent classes can not be extended infinitely, because this leads at a certain point to identification problems. This observation is also made in the context of semi-parametric estimation of latent trait models (see for example Mislevy, 1984).

When there are no theoretical grounds for taking a specific number of latent node points, as will very often be the case, one strategy is to vary the number of node points and to weight the benefit of a reduction in the likelihood ratio against the drawback of a less parsimonious and more complex model. As was made clear in the Chapter 2, choosing between models with a different number of latent node points (or classes) cannot be guided by resorting to statistical tests, as the difference in likelihood ratios is not longer asymptotically $\chi^2$-distributed. Also, the Akaike information criterium and other criteria that are based on AIC cannot solve this problem. Hence, again it is not possible to base the choice for a specific number of latent nodes on statistical or theoretical considerations. In the context of estimating Rasch models there is however one handhold. Before turning to this topic, some results will be presented for the semi-parametric estimation of the Rasch model to the data on women's lib. Table 4.5 provides the values of the likelihood ratio and Pearson $\chi^2$-statistics when the Rasch model is fitted to the data of the male, the female and the total sample.[15] All calculations were performed with the program DILTRAN, a program for the semi-parametric estimation of latent trait models[16] (Heinen and Vermaseren, 1992). The values for the node points were chosen on the interval from

Table 4.5: Semi-parametric estimation of the Rasch model; testing results for models with 2, 3 and 4 latent node points

| | | $T=2$ | $T=3$ | $T=4$ |
|---|---|---|---|---|
| Male | Pearson $\chi^2$ | 60.27 | 45.14 | 42.54 |
| | $G^2$ | 64.38 | 45.89 | 42.98 |
| | *df* | 24 | 23 | 22 |
| Female | Pearson $\chi^2$ | 62.50 | 42.82 | 43.32 |
| | $G^2$ | 68.69 | 44.40 | 41.92 |
| | *df* | 24 | 23 | 22 |
| Total | Pearson $\chi^2$ | 92.62 | 54.16 | 53.31 |
| | $G^2$ | 98.19 | 54.97 | 51.77 |
| | *df* | 24 | 23 | 22 |

$-3$ to $+3$ in order to preserve some comparability with the assumption of a standard-normal distributed latent variable. The node-points are assumed to be equally spaced on this interval. So for the case with two nodes, the values are $-1$ and $+1$; with three nodes, the values are $-1.5$, 0 and 1.5; finally, with four node points, the values $-1.8$, $-0.6$, $+0.6$ and 1.8 are used.

It is clear from the results in Table 4.5 that a model with two latent node points does not provide an acceptable fit. Note that this model differs from an unrestricted latent class model with two latent classes in that the association between the latent variable and a manifest indicator is the same for all manifest items. The models with three and four latent nodes provide a better, but not yet acceptable fit. These models treat the latent variable as a variable on an interval scale with equidistant scoring. The difference in the likelihood ratio between the model with three nodes and a model with four nodes, is not very large.

[15]Again only respondents fulfilling the educational requirements were retained in the analyses.

[16]This program is currently only available for VAX computers working under the VMS operating system. Copies of this program can be obtained from the author.

When more than four nodes were used, the fit could not be improved. Moreover, when trying to estimate a model with more than four nodes for these data, numerical problems were frequently encountered. It is also questionable whether models with more than four latent nodes are identified. This raises the question what is so special about the solution with four node points. To examine this, the estimated item parameters for the four latent classes solution are presented in Table 4.6. It should be kept in mind that the item parameters $c_j$ are equal to the negative of the item difficulties. The standard errors are denoted between parentheses after the values of the estimated parameters.

Table 4.6: Estimated item parameters for the Rasch model; semi-parametric estimates; $T = 4$

| | | Male | | Female | | Total | |
|---|---|---|---|---|---|---|---|
| Discrimination parameter : | Slope | +1.84 | (0.12) | +1.80 | (0.15) | +1.86 | (0.10) |
| | Item 1 | −1.89 | (0.21) | −1.15 | (0.28) | −1.61 | (0.22) |
| | Item 2 | −0.63 | (0.19) | +0.20 | (0.28) | −0.31 | (0.22) |
| Easiness | Item 3 | −2.10 | (0.21) | −1.67 | (0.28) | −1.97 | (0.22) |
| parameters : | Item 4 | +1.06 | (0.25) | +2.45 | (0.28) | +1.61 | (0.21) |
| | Item 5 | −2.84 | (0.23) | −1.33 | (0.28) | −2.16 | (0.22) |

The estimated value of the slope is dependent on the choice of the $\theta_t$ values. When the estimated item parameters are examined more closely, it becomes clear that they are exactly equal to the negative of the estimated item difficulties in Table 4.2, i.e., the CML estimates. The semi-parametric solution that is obtained here is therefore identical to the CML solution. That it is possible that these two procedures can give the same results was already mentioned by Clogg (1988). It should be remembered that the CML estimates for the item parameters can be obtained by fitting a log-linear model. In addition to the item parameters, this log-linear model includes a set of $u^*_\varsigma$ parameters to insure that the distribution of the total-score variable is reproduced exactly. As was shown in one of the previous sections, there are exactly $n$ different independent $u^*_\varsigma$ parameters to be estimated in this log-linear model. When a latent variable $\theta$ is introduced with values $\theta_t$, where $t = 1, \ldots, T$, the set of $u^*_\varsigma$ parameters can be written as polynomial function of this variable $\theta$ with at most $(n - 1)$ different terms. Hence, when there are 5 dichotomous items, there can be no more than four latent classes when a Rasch model is fit to these data.

This rather simple rule of thumb in establishing the maximum number of latent nodes also makes clear why models with more than 4 latent classes are not identified. Although in this example the CML estimates for the item parameters could be obtained with a discretized latent trait model with $(n-1)$ latent nodes, three additional remarks should be made in this context.

Firstly, choosing fewer than $(n-1)$ latent classes can also lead to a acceptable fit and so there is no a priori reason for always taking $(n-1)$ latent classes, except when one is only interested in obtaining CML estimates. At present, however, there seems to be no rationale for preferring CML estimates a priori.

Secondly, although the semi-parametric estimates for the item parameters in the examples presented were exactly the same as the CML estimates, this is not necessarily the case. It depends on the data whether the distribution of the total-score variable can be reproduced exactly by a restricted latent class model. A number of very important results in this context were established by Lindsay et al. (1991). This topic is discussed in more detail in the next section.

Finally, sometimes the fit can be improved considerably by choosing a spacing of the latent nodes that is not equidistant, as the analyses in Chapter 3 have pointed out. So, in principle it is possible to search for an optimal spacing of the latent nodes.

#### 4.2.3.3 Fully semi-parametric estimation

In fully semi-parametric estimation both the latent proportions $p_{\theta_t}$ and the latent node points $\theta_t$ are estimated from the data. This estimation procedure has been proposed for the Rasch model by de Leeuw and Verhelst (1986). The same method was developed independently by Follman (1988). Fully semi-parametric MML can easily be extended to the Partial credit model.

As a starting point, the partioning of the likelihood function of MML as shown in Equation 4.15 is taken up once again:

$$\ln L_{MML} = \sum_{\nu} f_\nu \cdot \ln p_\nu$$

$$\begin{aligned} &= \sum_{\nu} f_{\nu} \cdot \ln p_{\nu|\varsigma} + \sum_{\varsigma} f_{\varsigma} \cdot \ln p_{\varsigma} \\ &= \ln L_{CML} + \ln L_{\varsigma}. \end{aligned}$$

Maximization of the first part on the right-hand side yields estimates for the item-parameters that are identical to CML estimates. This part is a function of the item parameters, but not of the parameters of the distribution of the latent variable $\Phi(\theta)$. The second part on the right-hand side refers to the distribution of the total-scores. It measures how well the observed total-score distribution $(f_{\varsigma} : \varsigma = 0, \ldots, \varsigma_{max})$ can be approximated by the theoretical probability distribution $(p_{\varsigma} : \varsigma = 0, \ldots, \varsigma_{max})$. For the case of dichotomous variables, it has already been shown that the latter probability distribution is a function of both the item parameters and the parameters of the distribution function $\Phi(\theta)$ (see Equation 4.14). These results can easily be generalized to the situation of polytomous items. When the Partial Credit model is specified to model the relations between the latent and the manifest variables, the response probabilities can be expressed as

$$p_{jg|\theta_i} = \frac{\exp(g \cdot \theta_i + c_{jg})}{\sum_h \exp(h \cdot \theta_i + c_{jh})}. \tag{4.17}$$

Equation 4.14 can now be generalized to

$$p_{\varsigma} = \gamma_{\varsigma} \int_{-\infty}^{+\infty} \frac{\exp(\varsigma \cdot \theta_i)}{\prod_{j=1}^{n} \left[\sum_h \exp(h \cdot \theta_i + c_{jh})\right]} \, d\Phi(\theta), \tag{4.18}$$

in which $\Phi(\theta)$ is the cumulative distribution function of the latent variable and with

$$\gamma_{\varsigma} = \sum_{\nu \in \Omega_{\varsigma}} \exp \sum_{j} \sum_{g} (x_{\nu jg} \cdot c_{jg})$$

The notation used is analogous to the one introduced earlier. The symbol $\varsigma$ denotes the total-score over $n$ items; each item is scored with successive integers $0, \ldots, m$, so $0 \leq \varsigma \leq (n \times m)$. The subset of response patterns $\nu$ with $\varsigma_{\nu} = \varsigma$ is indicated by $\Omega_{\varsigma}$, and the summation is taken over all response patterns $\nu$ that result in a total-score equal to $\varsigma$. The variables $x_{\nu jg}$ are indicator variables that take the value 1 if in response pattern $\nu$ item $j$ is responded to in category $g$; otherwise the variable is equal to 0.

Thus, in the case of the Partial Credit model it is also true that the theoretical probability distribution of $p_\varsigma$ depends on both the item parameters $c_{jg}$ and the parameters of the distribution function $\Phi(\theta)$.

It is important to investigate the conditions under which the CML estimates are identical to the MML estimates. Let the CML estimates for the item parameters be denoted by $\hat{c}_{jg}$; hence, the set of $\hat{c}_{jg}$ maximize $L_{CML}$. It can now be shown that $L_{MML}$ and $L_\varsigma$ are maximized by these CML estimates if and only if a distribution function $\Phi(\theta)$ can be found for which $L_\varsigma$ with the item parameters equal to $\hat{c}_{jg}$ attains its maximal realizable value. The latter value is obtained by the unconstrained estimates under a multinomial model. Maximizing $\ln L_\varsigma = \sum_\varsigma f_\varsigma \cdot \ln p_\varsigma$ yields the following estimates for the probabilities $p_\varsigma$:

$$\hat{p}_\varsigma = \frac{f_\varsigma}{N} \text{ with } N = \sum_\varsigma f_\varsigma.$$

Hence, the problem is reduced to the question whether, given the CML estimates of the parameters $c_{jg}$, a distribution function $\Phi(\theta)$ can be found such that

$$\gamma_\varsigma \int_{-\infty}^{+\infty} \frac{\exp(\varsigma \cdot \theta_i)}{\prod_{j=1}^{n} \left[\sum_h \exp(h \cdot \theta_i + c_{jh})\right]} d\Phi(\theta) = \frac{f_\varsigma}{N} = \hat{p}_\varsigma \tag{4.19}$$

Now define:

$$\begin{aligned} \xi &= \exp(\theta) \text{ and} \\ dG(\xi) &= \prod_{j=1}^{n} \left[\sum_h \exp(h \cdot \theta_i + c_{jh})\right]^{-1} d\Phi(\theta). \end{aligned}$$

Then:

$$\int_0^{+\infty} \xi^\varsigma dG(\xi) = \frac{\hat{p}_\varsigma}{\gamma_\varsigma} = w_\varsigma. \tag{4.20}$$

Similar results for the Rasch model were first achieved by Cressie and Holland (1983) and later by de Leeuw and Verhelst (1986). A more general result was obtained by Holland (1990a).

It can be seen from Equation 4.20 that solving this equation implies solving a power moment problem. The necessary and sufficient conditions on the set of moments $\{w_\varsigma : \varsigma = 0, \ldots, \varsigma_{max}\}$ can be stated in the following way:

Case I: $\varsigma_{max} = 2 \cdot y$ is even.

Let $U$ be a matrix of order $(y+1) \times (y+1)$ with elements

$$u_{i,j} = w_{i+j-2}$$

for $i, j = 1, \ldots, y+1$, and let $V$ be the $v \times v$ matrix with elements

$$v_{i,j} = w_{i+j-1}$$

for $i, j = 1, \ldots, y$. Then the moment problem has a solution if $U$ and $V$ are both positive semi-definite.

Case II: $\varsigma_{max} = 2 \cdot y + 1$ is odd.

The two $(y+1) \times (y+1)$ matrices $U$ and $V$ are defined as follows:

$$\begin{aligned} u_{i,j} &= w_{i+j-2} \\ v_{i,j} &= w_{i+j-1} \end{aligned}$$

for $i, j = 1, \ldots, y+1$. The moment problem has a solution if both matrices $U$ and $V$ are positive semi-definite.

If both matrices are positive semi-definite, there is a bounded non-negative measure $dG(\xi)$ such that

$$w_\varsigma = \int_0^{+\infty} \xi^\varsigma dG(\xi)$$

for $\varsigma = 0, \ldots, \varsigma_{max} - 1$, and for which

$$w_y = \int_0^{+\infty} \xi^\varsigma dG(\xi) + \lambda$$

in which $\lambda \geq 0$ is the mass at infinity. If both matrices are positive definite, a representation with $\lambda = 0$ is possible.

It is thus possible to specify the conditions under which it is possible to estimate the distribution function $G(\xi)$. The inequality conditions that follow from the claim that the matrices $U$ and $V$ must be positive semi-definite are the same as those discussed by Cressie and Holland (1983) and Kelderman (1984).

But the question of how this latent distribution $G(\xi)$ can be estimated has not been answered yet. The latent distribution cannot be determined uniquely by maximum likelihood unless one restricts the class of possible solutions. This can be done by estimating a discrete distribution function $G^*$ with a specified number of support points (=nodes) that has the same first $n$ moments as the continuous distribution function $G$. Karlin and Studden (1966) have proven that such a discrete distribution function exists. This makes it possible to estimate a discrete latent distribution in a semi-parametrical way. Results with respect to the consistency of such estimates can be found in Kiefer and Wolfowitz (1956). A rather technical elaboration of this topic can be found in Engelen (1989).

Suggestions for using Karlin and Studden's (1966) results to obtain fully semi-parametric estimation of the parameters in the Rasch model were first formulated by de Leeuw and Verhelst (1986). The same ideas were developed independently by Follman (1988). Clogg (1988) arrived at a formally equivalent model, but he started from a restricted latent class point of view. Following Karlin and Studden (1966), de Leeuw and Verhelst (1986) restricted themselves to a description of *canonical solutions* for the estimation of the latent distribution. These canonical solutions are step functions.

1. In the case that $\varsigma_{max}$ is even, this step function has $(\varsigma_{max} + 2)/2$ points of increase, with the first point being equal to $-\infty$. (Note that this first point of increase corresponds to a latent class in which the response pattern $(0, 0, \ldots, 0)$ is given with probability 1.)

2. In the case that $\varsigma_{max}$ is odd, the canonical solution has $(\varsigma_{max}+1)/2$ points of increase, none of which is specified in advance. Note also that these observations on canonical solutions only apply if the necessary and sufficient conditions on the "moments" are satisfied, i.e., if the marginal distribution of the total-scores can be fitted perfectly

An obvious question pertains to the relationship between the MML estimates obtained with fully semi-parametric estimation on the one hand and CML estimates on the other hand. Some intuitive notions can already be made in this context. When the inequality conditions formulated by Cressie and Holland (1983) are met, the estimated $u^*_\varsigma$ parameters in the log-linear model in Equation 4.13 are compatible with the existence of a latent distribution. On the other hand, when the discrete latent structure estimated by fully semi-parametric MML leads to estimated probabilities $\hat{p}_\varsigma$ that are equal to the observed proportions $(f_\varsigma/N)$ (i.e., when the total-score distribution is being reproduced by the MML estimates), the partioning of the complete likelihood is the same as in the case of log-linear analysis. It seems obvious that in this case both procedures yield the same estimated item parameters. De Leeuw and Verhelst (1986) were the first to formalize the relationship between fully semi-parametric MML estimates and CML estimates. They showed that the MML estimates for the item parameters are asymptotically equal to the CML estimates as $N \longrightarrow \infty$ for a discrete distribution function $G^*$ with an infinite number of support points. More general results were obtained by Lindsay et al. (1991).

Lindsay et al. (1991) make a distinction between concordant and discordant cases. A case is concordant if one succeeds in fitting the observed total-score distribution exactly; it is discordant in all other cases. With respect to the concordant case, a further distinction is drawn between *borderline concordance* and *PD (positive-definite) concordance.* PD concordance is achieved if all determinantal inequalities that arise in checking for concordance hold strictly. Borderline concordance is achieved if at least one of these inequalities holds as an equality. Lindsay's (1991) main results may now be stated in the following way:

1. For concordant cases, the MML and CML estimates of the item parameters are identical.

2. In the case of borderline concordance, the latent distribution can be estimated uniquely. Moving beyond the critical number of node points does not yield a better fit; moreover, the corresponding latent distribution will degenerate in the following sense: the extra number of node points added beyond the critical number will collapse so that, in the end, one always obtains the latent distribution which assigns non-zero probability only to the critical number of node points. This latent distribution is estimated uniquely.

3. In the case of PD concordance, the latent distribution cannot be estimated uniquely: increasing the number of node points beyond the critical number does not improve the fit, but will result

in definitely distinct latent distributions. In this case, the latent distribution is not identified: one obtains a different estimate for each different number of node points. However, the latent distribution may be uniquely determined for a fixed number of node points.

4. In the discordant cases, MML and CML estimates are different.

5. For discordant cases, the estimates of the latent distributions are also unique and maximally $\varsigma_{max}/2$ distinct node points are involved.

Both de Leeuw and Verhelst (1986) and Lindsay et al. (1991) indicated how the EM algorithm can be used to obtain fully semi-parametric estimates with a discretized latent distribution. To illustrate this, the results of an analysis on the data on women's liberation for the male sample will be presented. Again, only respondents who satisfied the educational requirements were used in the analysis.

The parameters of the Partial Credit model were estimated with three, four and five latent nodes, respectively. In each case the same value for the log-likelihood ratio ($G^2 = 218.06$) was found; the MML parameter estimates $\hat{c}_{jg}$ were very much alike. None of these three models could reproduce exactly the distribution of the total-score. All these results indicate that this is a discordant case. In order to verify this, the CML estimates for the parameters were also calculated with log-linear models. If the data really is discordant, the CML estimates should differ from the MML estimates.

It should be noted that the value for the log-likelihood ratio obtained with fully semi-parametric MML is almost equal to the values for this ratio obtained with MML with four fixed latent nodes (see Table 3.9). In order to investigate this further, the estimated $c_{jg}$ parameters and the estimated latent structure should be inspected. Table 4.7 presents the estimated latent structure both for semi-parametric with fixed and free values for the latent nodes. With fixed latent nodes, the first latent class has a probability near zero. The other three latent classes have probabilities which are almost equal to the corresponding latent proportions for MML with free latent nodes. The latter procedure estimates the two distances between the values for the latent nodes as

Table 4.7: Estimated latent structure; semi-parametric MML with free and fixed nodes; Partial Credit model for the male subsample

| Free latent nodes | | Fixed latent nodes | |
|---|---|---|---|
| $\theta_t$ | $p_{\theta_t}$ | $\theta_t$ | $p_{\theta_t}$ |
| | | −2.1 | 0.02 |
| −1.7 | 0.11 | −0.7 | 0.11 |
| +0.1 | 0.49 | +0.7 | 0.49 |
| +1.6 | 0.40 | +2.1 | 0.38 |

almost equal. As this was also the assumption for the analysis with fixed latent nodes, this probably explains why both procedures gave approximately the same fit.

The comparison of the estimated $c_{jg}$ parameters obtained with different estimation procedures can sometimes be hampered by the fact that the location of the latent scale is set arbitrarily. This was already discussed in Section 3.3.3.2. Therefore, it is more convenient to calculate the estimated thresholds and to fix the location on the latent scale by fixing one of the thresholds to, for instance, zero.

Table 4.8 gives the values for the estimated $c_{jg}$ parameters as well as the estimated threshold values, when the thresholds are relocated such that the first threshold (i.e., $\delta_{j1}$) equals 0. These estimated values are given for the Partial Credit model estimated for the male subsample, where the estimation was done (i) by fully semi-parametric MML with three latent nodes (ii) by semi-parametric MML with four fixed latent nodes and (iii) by CML using log-linear models.[17] The estimates for the threshold parameters obtained with the two semi-parametric MML methods are nearly identical. Given that the value of the log-likelihood was almost equal for the two methods, this is not very surprising.

When CML is used, the estimates are different: the distance between the two estimated thresholds is much greater. CML also resulted in a value for the log-likelihood ratio that was different from the values that were found with MML. The log-likelihood ratio for CML was 192.94. All these results conform to the theoretical conclusions that were drawn earlier in this section. This is a discordant case. The parameters es-

[17]The fully semi-parametric MML procedure was carried out by an ad-hoc program written by Croon. The analyses with semi-parametric MML with fixed latent nodes were performed with DILTRAN, while the CML estimates were obtained with FREQ.

Table 4.8: Estimated category-parameters and thresholds; three different estimation procedures

| | | Fully semi-parametric MML with $T = 3$ | | Semi-parametric MML with $T = 4$ fixed nodes | | Conditional ML | |
|---|---|---|---|---|---|---|---|
| | | $c_{jg}$ | $\delta_{jg}$ | $c_{jg}$ | $\delta_{jg}$ | $c_{jg}$ | $\delta_{jg}$ |
| Item 1 | Cat. 1 | 0.66 | 0.00 | 0.07 | 0.00 | −0.14 | 0.00 |
| | Cat. 2 | 0.44 | 0.88 | −0.75 | 0.89 | 0.46 | −0.74 |
| Item 2 | Cat. 1 | 0.64 | 0.00 | 0.04 | 0.00 | −0.18 | 0.00 |
| | Cat. 2 | 1.66 | −0.38 | 0.47 | −0.39 | 1.67 | −2.03 |
| Item 3 | Cat. 1 | −0.66 | 0.00 | −1.24 | 0.00 | −1.49 | 0.00 |
| | Cat. 2 | −0.60 | −0.72 | −1.79 | −0.69 | −0.59 | −2.39 |
| Item 4 | Cat. 1 | 1.41 | 0.00 | 0.83 | 0.00 | 0.59 | 0.00 |
| | Cat. 2 | 3.81 | −0.99 | 2.61 | −0.95 | 3.81 | −2.63 |
| Item 5 | Cat. 1 | −1.39 | 0.00 | −1.98 | 0.00 | −2.23 | 0.00 |
| | Cat. 2 | −1.71 | −1.07 | −2.91 | −1.05 | −1.71 | −2.75 |

timated with CML differ from the ones estimated via semi-parametric MML. The log-likelihood ratio obtained with CML is smaller than the one found with MML because in CML a part of the total log-likelihood is discarded. The restrictions that result from the assumption of a proper latent distribution function play a significant role in MML, but are neglected in CML. The difference in the estimated parameters is clearly shown in Figure 4.1.

The category-characteristic curves for MML are represented by the solid lines, whereas the CML curves are depicted as dashed lines. It is very remarkable that the middle categories are much less important according to the CML estimates, as compared with the MML estimates. It is also obvious that the distance between the two thresholds is greater for CML than for MML.

### 4.2.4 Evaluation of the maximum likelihood estimation methods

The differences between the three maximum likelihood methods presented in this chapter (i.e., JML, CML and MML) are not merely numerical. Each maximizes its own likelihood function. Therefore, one could raise the question whether one of the estimation schemes is to be preferred to the others.

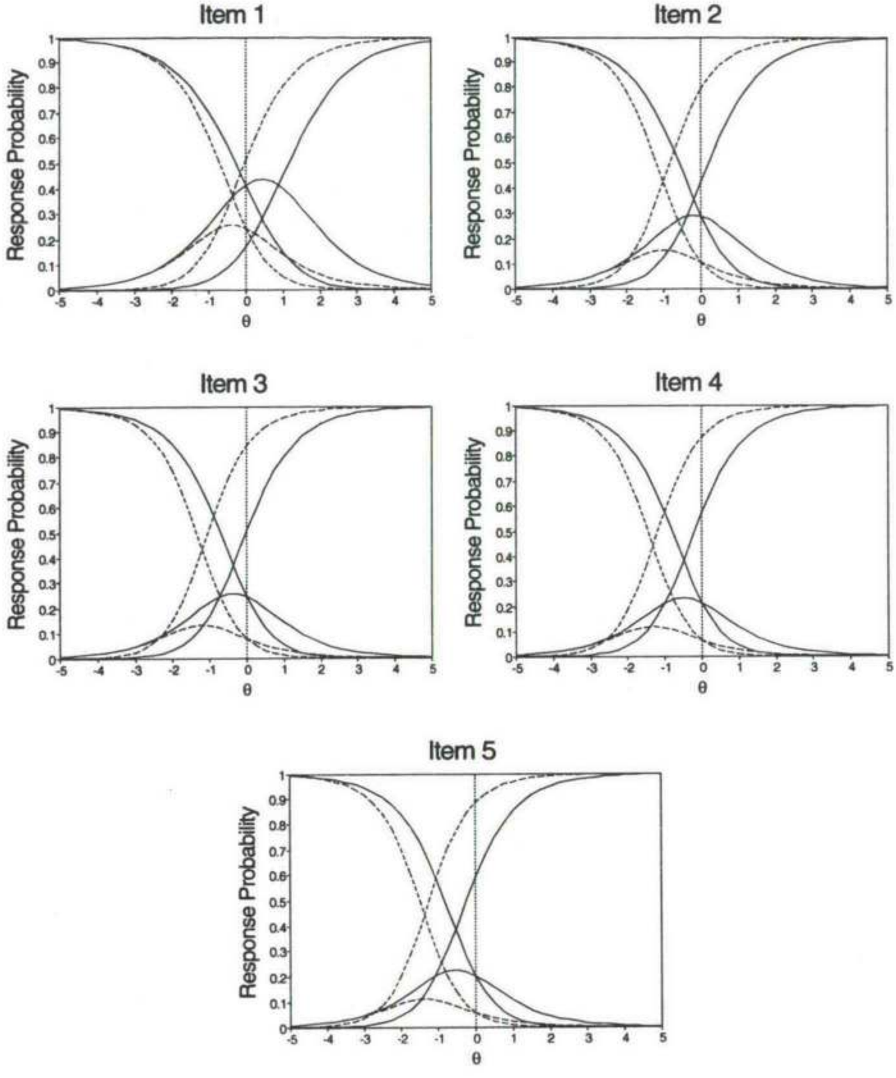


Figure 4.1: Partial Credit model: category characteristic curves for MML (solid lines) and CML (dashed lines)

There are a number of objections that can be raised against the use of JML. This method is numerically very troublesome for the two- and three-parameter models and the item parameter estimates are not consistent. With respect to the relationship with the other two methods, the following observations can be made. First, as Holland (1990b) has noted, it is possible to rewrite the MML likelihood function in a way that is comparable with the JML likelihood function. Holland shows that in JML the person parameters $\theta_i$ are free to vary over all possible functions of the response patterns $\nu$, whereas in MML the parameters $\theta_i$ are in some complicated way constrained by the distribution function $\Phi(\theta)$. As Holland notes, this is in accordance with de Leeuw and Verhelst (1986) who refer to JML as *unconstrained maximum likelihood.* The fact that these complicated constraints are ignored in JML also explains why JML will lead to $G^2$-ratios and to estimated standard errors for the parameters that are both too small. As such, JML presents a view that is too optimistic.

It is also possible to link JML and CML. This link originates from the problem of estimating person parameters $\theta_i$ in CML. Because in CML the $\theta_i$ parameters are conditioned out of the likelihood, these parameters have to be estimated in a separate step, after the item parameters have been estimated. There are a number of possibilities for estimating the person parameters.

Firstly, one could use a likelihood function in which the item parameters are conditioned out by using the number-correct per item as a sufficient statistic for the item parameters. This approach has been advocated by Fischer (1974). As Zwinderman (1991a) has noted, this procedure is of little practical value when the number of subjects is large because the symmetric functions cannot be evaluated.

A second alternative (and herein lies the link to JML) is to use the JML function with the item parameters fixed at their estimated CML values. This procedure is also subject to criticism because with a small number of items the estimates for $\theta_i$ will be biased and the standard errors of these estimates will be relatively large. Both procedures have the disadvantage that no $\theta_i$ estimates are available for the two most extreme response patterns (i.e., all items answered incorrectly or all items answered correctly). Moreover, both procedures also seem to arise from what Holland (1990b) has called the *stochastic subject rationale.*

So the $\theta_i$ parameters are seen as a set of parameters that belong to subjects, while the subjects are considered a "fixed" factor. Hence, both procedures pertain to what de Leeuw and Verhelst (1986) have called the functional Rasch model.

A third procedure is the one suggested by Andersen and Madsen (1977). Their approach differs from the former two because a latent distribution function $\Phi(\theta)$ is assumed to exist and the latent quantities $\theta_i$ are now no longer considered to be merely a set of parameters belonging to "fixed" individuals. Rather, the $\theta_i$ quantities are seen as latent scores belonging to subjects that were sampled at random. In other words, it is the *random sampling rationale* (Holland, 1990b) which governs the ideas of Andersen and Madsen. Their procedure pertains to the *random-score* model or *structural* model (de Leeuw and Verhelst, 1986). Within the context of the Andersen and Madsen approach, no individual $\theta_i$ parameters are estimated, but the parameters describing the distribution function $\Phi(\theta)$ are to be estimated.

This procedure has been briefly described earlier in this chapter. It is an interesting procedure for two reasons. Firstly, it uses the same starting point as some MML procedures and hence it sheds light on the relationship between CML and MML. Secondly, it gives the opportunity to comment upon the usefulness of the Andersen and Madsen approach in the light of the results obtained by Lindsay et al. (1991) with regard to MML estimation.

Andersen and Madsen take as a starting point the likelihood in terms of the probabilities $p_\nu$ which, as was noted earlier, can be factored in two terms by conditioning on the total-scores. One of these terms is equal to the CML-likelihood and the other pertains to total-score distribution. The first term is used to obtain CML estimates for the item parameters. The second term depends, among other things, on the latent distribution function $\Phi(\theta)$. After the CML estimates for the item parameters have been obtained, this second term is used to estimate the parameters describing $\Phi(\theta)$.

Herein lies a substantial difference with MML: within the context of MML the item parameters and the parameters characterizing the latent distribution are estimated jointly. This has given rise to a serious criticism of MML, for the hypothesis tested within the MML framework pertains to both a measurement model for the items and some assumptions regarding the latent distribution. When the latent trait model fails to fit the data, it is not clear what the reason is for this bad fit. Is the measurement model specified incorrectly, or are the assumptions with respect to the population distribution incorrect? In this sense, the approach suggested by Andersen and Madsen seems a reasonable alternative. First, the measurement model is tested using CML and afterwards structural hypotheses pertaining to the population distribution can be tested separately. However, this approach assumes that the estimated item parameters are compatible with a latent distribution function $\Phi(\theta)$. The results obtained by Cressie and Holland (1983) and Lindsay et al. (1991) have pointed out that CML estimates for the item parameters are not necessarily compatible with the idea of a latent distribution function. As was mentioned before, this is caused by the fact that the second term in the partioned likelihood represents a power moment problem and that the solution of this problem does not necessarily guarantee that the estimated total-score distribution equals the observed distribution. So, conditioning on these total-scores as is done in CML can lead to different estimates as compared to semi-parametric MML where the latent distribution is characterized directly.

If the latent trait model is conceived of as being built upon Holland's random sampling rationale and, hence, a latent distribution of scores is thought to exist, semi-parametric MML is to be preferred. The fact that goodness of fit tests in this MML framework necessarily pertain to complex hypotheses in which both the measurement model and the latent population distribution play a role is a direct consequence of the random sampling rationale. It is the price that has to be paid for in structural latent trait models. The alternative is to stick with the stochastic subject rationale, but as Holland (1990b) notes, it is the random sampling rationale that provides the basis for statistical inference in latent trait models. Thus, the fixed-subjects approach seems less attractive. And if this fixed-subjects view is not adopted, CML estimation for the item parameters is less attractive because the only thing one can do is to hope that the estimates for the item difficulties are compatible with the existence of a proper latent distribution function. As Lindsay et al. (1991) have shown, this is only so in the PD-concordant case and whether this case appears in a specific analysis depends on the structure in the observed data.

Holland (1990b) also made some relevant comments with respect to these problems. He showed that CML and semi-parametric MML would always give the same results for the estimated item parameters *if* the second term in the partioned likelihood (i.e., the part that pertains to the total-score distribution) only depended on the latent distribution $\Phi(\theta)$ and not on the item parameters as well. As this is only approximately true, CML and MML will not always give the same results. These notions lead to some other interesting conclusions. When the Guttman scale model is considered as a deterministic Rasch model, it is easy to see that an individual's location on the latent continuum fully determines the total-score for that individual. In the usual Rasch models this is only approximately true because of the stochastic nature of the item response functions. However, in most data sets that were analyzed in this study, the data satisfied the PD concordant case and so CML and MML item parameter estimates are the same. In most practical situations CML works just fine. But this observation is only true for dichotomous indicators. With polytomous data, the relationship between the position that an individual takes on the latent scale on one hand and the observed total-score for that individual on the other hand is far less tight. The consequence is that for polytomous data, CML and MML will only rarely lead to the same estimated item parameters. In such cases the use of MML should definitely be preferred because the latent population distribution is characterized directly by this method.

The remaining question is whether semi-parametric or fully semi-parametric MML should be used. At this time, fully semi-parametric MML is only applicable in latent trait models for which known sufficient statistics for the $\theta_i$-parameters exist. This method can therefore be used for the Rasch model (in the case of dichotomous data) and for the Partial Credit model and the models that can be derived from this model (in the case of polytomous data). The semi-parametric method in which the values of the latent nodes are fixed in advance is more widely applicable and consequently, more advantageous. A second benefit of this procedure is that it is computationally much simpler than fully semi-parametric MML. However, the fact that a set of $\theta_t$ values has to be chosen can be regarded as a drawback of this procedure. In those cases in which both semi-parametrical procedures could be employed, the disadvantage of selecting the values of the latent node points in advance is, however, not always serious because in the case of PD concordance, the latent structure is not identified. With the correct choice of the number of latent nodes, both procedures will lead to the same estimated item parameters. Only in discordant cases (and as stated earlier, this will most notably be the case with polytomous data) is the latent structure unique and identifiable. In this case the two semi-parametric methods can lead to different solutions. The extent to which the estimated latent structure and the estimated item parameters will be different with both estimation schemes is not clear at this moment. Further research is needed to clarify these topics.

## 4.3 Testing latent trait models

The problem of testing log-linear models and latent class models was taken up in Chapter 2. To summarize this discussion briefly, there are basically two methods for evaluating the goodness of fit of such models.

One way is to perform statistical tests. The null-hypothesis for the overall test states that the observed data are sampled from a population for which the specified model is true. If this hypothesis has to be rejected, it will not always be clear which particular aspects of the data are responsible for the bad fit. Performing conditional tests on more specific hypotheses can help to identify the sources of bad fit and to indicate alternative models that will be more successful in describing the structure of association in the data.

A second way of analyzing and comparing competing models is to use descriptive indices. A benefit of these indices is that the problems that arise because of the asymptotic character of the statistical tests are circumvented. A drawback, however, is that the evaluation of the values of these measures is rather subjective.

The same methods for evaluating the goodness of fit of competitive models can be discerned in the context of latent trait models. First of all, there is, at least theoretically, the possibility of performing an overall test for goodness of fit. This can be done by using either the Pearson $\chi^2$-statistic or the log-likelihood ratio. The formulas for both test statistics were given in Chapter 2. Both statistics are asymptotically distributed as $\chi^2$-variables when certain conditions are satisfied. Read and Cressie (1988) showed that the Pearson statistic and the log-likelihood ratio are special cases of a more general statistic, from which other test statistics can be derived that share some of the properties of the Pearson statistic and the log-likelihood ratio. In some situations with sparse data, these new statistics may be better than Pearson's $\chi^2$ or $G^2$.

It is important to note that the conditions under which Pearson's $\chi^2$ and $G^2$ are asymptotically $\chi^2$ distributed will never be satisfied if the parameters of the model are estimated with JML because JML does not provide consistent estimates for the parameters. These difficulties do not arise with CML or MML.

As was already noted in Chapter 2, it is known that the expected frequencies $(N \cdot p_\nu)$ should be sufficiently large in order for the test statistics to have a sampling distribution that approximates the theoretical $\chi^2$ distribution to a satisfying level. While this creates problems when these statistics are used in latent class analysis, these problems are surpassed completely by the difficulties that arise within the setting of applied latent trait models. The reason is quite obvious: in psychological testing the number of items, and therefore the number of possible response patterns, is much larger than in the case of latent class analysis. So, in research where latent trait models are applied, the expected frequencies will be, on the whole, smaller.

Of course, this is not to say that there is any fundamental difference between latent class models and latent trait models as to the potential use of both test statistics. Several authors have proposed circumventing these problems by pooling response patterns in order to increase the expected frequencies. This introduces new problems. As Glas (1989) noted, the way response patterns are pooled is a function of the data itself. Therefore, it is hardly possible to derive the sampling distribution of test statistics based on pooled data.

Besides the possibility of testing global hypotheses concerning goodness of fit, one could also test more specific hypotheses in order to gain some understanding with regard to the possible sources of bad fit. One way to do so is to make use of estimated standard errors of the model parameters. Hypotheses concerning individual parameters can be tested by calculating estimated standardized values of the parameters. This procedure is identical to the one discussed in Chapter 2. Hypotheses with respect to single parameters can also be tested by formulating conditional tests, i.e., by comparing different models. Such conditional tests are also interesting when the hypotheses concern a set of parameters. This topic will be discussed later. Within the framework of latent trait models, the question has been raised whether the testing of hypotheses should be aimed at the assumptions that underlie the model (see, for example, Molenaar, 1983b). For the Rasch model, this has led to a number of different test statistics. Usually, these statistics are divided into two categories. These two types of test statistics are discussed in the following section.

**Test statistics based on first-order information** The first category is based on *first-order information* (Zwinderman, 1991a): the subjects are divided into subgroups and the number of positive responses to an item within these subgroups are compared with the expected numbers of positive responses. Because the observed numbers are the sufficient statistics for the item parameters in the subgroups, this procedure in effect studies the between-groups variation of the estimated item parameters (Van den Wollenberg, 1979). If the Rasch model holds, the estimated item parameters for subgroups can only differ within chance limits because the item parameters are sample independent. Thus, test statistics which are based on first order information are sensitive to violation of this assumption of sample independence.

The subgroups of subjects can be arranged according to any criterion. One could use background or demographic variables (for example, gender) or specific test items as in Van den Wollenberg's *splitter item technique* (Van den Wollenberg, 1979, 1982a). Very often, subgroups are obtained by partioning the sample according to observed total-scores. In that case, the testing procedure is especially suitable for testing the assumptions of monotonicity and sufficiency (Zwinderman, 1991a). Because sufficiency of the total-scores implies that the item characteristic curves run parallel, testing differences between item parameters estimated within such subgroups is identical to testing for variation in the slopes of these curves.

A very straightforward test statistic within this first category of test statistics is Andersen's (1973b) *conditional likelihood ratio test.* The general formulation for this test statistic is as follows (see, for instance, Zwinderman, 1991a):

$$Z = -2\ln\left[\frac{\hat{L}_c}{\prod_{a=1}^{m}\hat{L}_c^{(a)}}\right]$$

The number of different groups is equal to $m$. The maximum of the usual conditional likelihood (see Equation 4.9) is denoted by $\hat{L}_c$, while the maximum for the conditional likelihood in subgroup $a$ is represented by $\hat{L}_c^{(a)}$. When the Rasch model holds and the item parameters are the same in different subgroups, $\prod_{a=1}^{m}\hat{L}_c^{(a)}$ will be equal to $\hat{L}_c$. Andersen proved that the test statistic is asymptotically distributed as a $\chi^2$-variable. The number of degrees of freedom for this test is equal to $(n-1)(m-1)$ when the number of items is, as usual, equal to $n$. When the subgroups are formed according to the total-scores, the number of these groups is equal to $(n-1)$ because individuals with a total-score equal to either 0 or $n$ do not provide any information about the item parameters. In that case, the number of degrees of freedom is equal to $(n-1)(n-2)$. A significant result indicates that the item parameters are not the same across subgroups and, as a result, the Rasch model does not hold.

Zwinderman (1991a) noted an interesting feature of this statistic: Andersen's conditional likelihood ratio is *conjugate* with respect to CML estimation. This means that the same estimated item parameters are obtained whether Andersen's statistic is minimized with regard to the "overall" item parameters (i.e., for the total sample) or ordinary CML methods are used. Van den Wollenberg (1988) also prefers the conditional likelihood ratio test from a theoretical point of view. A likelihood statistic is theoretically more suitable when the parameters are estimated using maximum likelihood methods. From that standpoint, Andersen's statistic is preferable to other candidates such as Martin-Löf's statistic (Martin-Löf, 1973) or Van den Wollenberg's $Q_1$-statistic (Van den Wollenberg, 1982b) because these statistics are both of the Pearsonian $\chi^2$-type.

Martin-Löf's statistic is based on a comparison of the observed and expected frequencies of the total-score distribution. It is expressed as

$$T = \sum_{\varsigma=1}^{n-1}(\boldsymbol{d}_\varsigma)'(\boldsymbol{V}_\varsigma)^{-1}(\boldsymbol{d}_\varsigma),$$

in which $\boldsymbol{d}_\varsigma$ is the vector of differences between the observed and expected total-score frequencies. The summation takes place over subgroups formed according to the total-score; individuals with a total-score equal to either 0 or $n$ (hence, individuals passing none or all items) are left out. The matrix $\boldsymbol{V}_\varsigma^{-1}$ is the variance-covariance matrix of the conditional expectations $\mathcal{E}(f_{\varsigma j} \mid f_\varsigma)$, hence, the expected number of positive responses to item $j$ within the subgroup with total-scores equal to $\varsigma$. Van den Wollenberg's $Q_1$-statistic is a special case of Martin-Löf's statistic in the sense that for $Q_1$ it is assumed that the off-diagonal values of the matrix $\boldsymbol{V}_\varsigma^{-1}$ are equal.

From a practical point of view, it is advisable to use the $Q_1$-statistic suggested by Van den Wollenberg because it is easier to calculate and it is also applicable in situations in which Andersen's statistic cannot be used. Moreover, Van den Wollenberg showed in a number of simulation studies that the three statistics are highly correlated. However, it has been proven that $T$ follows asymptotically a $\chi^2$-distribution, while the sampling distribution of $Q_1$ is unknown. Glas (1989) developed the same statistic as Martin-Löf, but in a more general context. Glas showed how this test statistic can be generalized for the unidimensional polytomous Rasch model and for the Partial Credit model. Moreover, Glas developed a similar test statistic that can be used in the context of MML. The interested reader is referred to the literature given above for further details.

A final remark concerns the availability of still other test statistics that were suggested to test for differences in slopes of item characteristic functions (see, for example, Wright and Panchapakesan, 1969 and Fischer and Scheiblechner, 1970). These statistics have been criticized on theoretical grounds by Van den Wollenberg (1988). On the basis of this criticism, it can be concluded that these statistics should no longer be applied because the alternatives presented above are more tractable both from a theoretical and a practical point of view.

**Test statistics based on second-order information** The second category of test statistics relies on *second-order information.* This kind of information can be found in two-way tables for pairs of items. Tests based on this kind of data are sensitive to violation of the assumption of a unidimensional latent trait and the assumption of local independence. When the data are in effect generated by a Rasch model, there should be no association between pairs of manifest indicators when the latent trait is partialed out. Hence, this test compares the observed with the expected two-way tables for item pairs. Van den Wollenberg (1979, 1982b) proposed a test statistic named $Q_2$ for this purpose. Glas (1989) formulated an alternative statistic which is easier to calculate and for which it can be proven that it has a sampling distribution that is asymptotically $\chi^2$-distributed. Glas also generalized this test statistic for MML estimation.

Nearly all test statistics that fall in the two categories mentioned here apply to the Rasch model only. The statistics suggested by Glas, however, can be used in a context that is slightly broader; Glas's statistics can be used with polytomous data, provided that known sufficient statistics exist. This broadens the scope to the Partial Credit Model, but latent trait models such as the Lord-Birnbaum model or Bock's Nominal Response model are not included. Glas has also generalized the use of the test statistics to the context of MML estimation of the parameters in the Rasch model or Partial Credit model. With this generalization, new relevant assumptions that underlie the model come into the picture, i.e., assumptions regarding the distribution of the latent variable. Glas (1989) developed the $R_0$-statistic for this purpose. The sampling distribution of this statistic is approximately $\chi^2$-distributed. This test can only be applied for the Rasch model and the Partial Credit Model, i.e., for models with known sufficient statistics for the ability parameters.

Zwinderman (1991a) noted that it is almost impossible to separate between violations of the assumptions of the Rasch model (i.e., sample independence, equal slopes, and the like) and misspecification of the latent distribution. The statistic proposed by Glas seems sensitive to both kinds of violations. Simulation studies performed by Zwinderman have also pointed out that misspecification of the latent distribution results in MML estimators that are slightly biased and less efficient. Zwinderman, among others, used this result to criticize MML and to plead for a combined use of CML and MML as in the approach suggested by Andersen and Madsen (1977). As was stated in previous sections, the problem of misspecification of the latent distribution can be circumvented by using (fully) semi-parametric MML methods. Moreover, if one is indeed interested in estimating properties of the latent distribution, it is natural to start within a MML framework because CML estimators do not guarantee that the estimated values for the item parameters are reconcilable with a proper distribution of the latent variable.

**Conditional tests** As was stated before, models can be evaluated statistically by carrying out formal tests. Besides the global tests, there are also a number of specific test procedures for the Rasch model. The discussion in Chapter 2 made it clear that tests can be unconditional or conditional. Conditional tests are based on the comparison of hierarchically nested models. So far, the tests discussed in this section were unconditional, but a number of the specific tests presented here can also be translated into conditional terms. Kelderman (1989) has pointed out these conditional tests by using the log-linear formulation of the Rasch model:

$$\ln p_\nu = u^*_\varsigma + \sum_j x_{ij} \cdot u_j$$

This log-linear model pertains to a contingency table formed by *item 1* $\times$ *item 2* $\times \cdots \times$ *item n* $\times$ *total-score*. The model assumes that the two-variable interactions are absent and because the table is incomplete, the specified model is a model for *quasi-independence* (Goodman, 1968). The fit of this model can now be compared with the fit of other log-linear models which are more or less restrictive. An example of such a model is the log-linear model which is more general than the Rasch model because it also includes the two-variable interactions for *item j* $\times$ *total-score* for $j = 1, \ldots n$. These two-variable interactions can only reach significant values when the item difficulties vary over different score-groups (i.e., groups that are composed according to the total-score). Kelderman has shown that a test of the difference in the log-likelihood ratios of this model and the log-linear Rasch model is equivalent to Andersen's conditional likelihood ratio test. The test can, of course, be made more specific by including only certain two-variable interactions. Other alternative models that enable us to test interesting hypotheses can be developed by including background variables. Kelderman (1989) shows how hypotheses regarding item bias can be tested by including appropriate interactions between items and background variables into the log-linear model. Two-variable interactions for pairs of items can be used to test hypotheses regarding local independence. Finally, it is easy to generalize the log-linear Rasch model to a multidimensional latent trait model by calculating total-scores over different subsets of items. Models that are related to these multidimensional log-linear Rasch models are also discussed by Goodman (1990).

The log-linear approach for testing assumptions in the Rasch model as developed by Kelderman is very challenging, but it also rather restrictive. It can only be used within the context of the Rasch model; in more general models such as the Lord-Birnbaum model, the total-score is no longer a sufficient statistic for the ability parameter. Therefore, if the objective is to compare the one-parameter model with a two-parameter model, it seems more fitting to start with a two-parameter model. A check can be made to determine whether the discrimination parameters can be restricted to be equal to each other. Kelderman's strategy is not entirely satisfying in this regard: when his conditional test identifies items that are not equally difficult for different score groups, the Lord-Birnbaum model might be a better fitting alternative, but it does not have to be so. In other words, Kelderman's approach is satisfactory when it comes to merely checking some assumptions made in the Rasch model, but it is not very adequate for comparing competing models, when these rival models have varying slope parameters. In addition, the log-linear approach uses CML estimation. As was argued before, this procedure will not always be appropriate when the data are polytomous. It therefore makes sense to compare rivaling models without restricting ourselves to log-linear models.

**Descriptive methods** There are a number of limitations to the use of statistical tests in item response theory. In small samples statistical tests will probably not be very sensitive to severe violations of the model assumptions. Moreover, given the fact that the sampling distributions of the tests statistics are only asymptotical $\chi^2$-distributed, the test results will not always be very reliable in relatively small samples. With large samples, however, small deviations will give rise to significant results. Finally, tests do not give much information with regard to the degree of fit or bad fit. This means that descriptive methods are needed in addition to the formal statistical tests.[18] To gain some understanding with regard to the goodness of fit of a model in a descriptive way, the methods presented in Chapter 2 can be used. The goodness of fit indices based on the value of the log-likelihood ratio as well as the information criteria can also prove to be of some value in the application of latent trait models.

Van den Wollenberg (1988) and Zwinderman (1991a) review a number of alternative descriptive methods that can be used for the Rasch model. These methods are similar to the test statistics that were developed for the Rasch model because the descriptive models have also been developed to be sensitive for the violation of the typical Rasch model assumptions. Again, first-order information is used to check for sample independence and second-order information helps to control for conditional independence. The reader is referred to Zwinderman's excellent review of (1991a) for details.

[18]The term *descriptive methods* is used here in a rather loose way. Zwinderman (1991a) makes a distinction between heuristic, diagnostic and descriptive methods. In the terminology used here, descriptive methods cover all these different methods.

# Chapter 5

# Some extensions of latent trait models: introducing additional latent and manifest external variables

## 5.1 Introduction

This chapter deals with two major extensions of latent trait models. Firstly, multidimensional latent trait models are studied. The latent trait models that were discussed in the previous chapters, are all based on the assumption that the association between the manifest indicators is explained by just one latent trait. However, in a number of situations this assumption is not realistic. Moreover, in some cases there may be a necessity for testing specific assumptions concerning unidimensionality. Therefore, an extension of the latent trait models discussed in Chapter 3 in the direction of multidimensionality is needed. This generalization can be accomplished in a number of ways. In this chapter the focus is on the multidimensional models that can be written as log-linear models with several latent variables. In this fashion, maximum comparability with the unidimensional models discussed earlier is preserved. Alternative methods of defining multidimensional latent trait models are discussed briefly.

A second major extension of latent trait models is the inclusion of external variables. External variables are variables that are not part of the measurement model, i.e., they are not indicators of the latent trait.[1] The inclusion of external variables in the analysis makes it possible to estimate the strength of the association between latent traits on the one hand and external variables on the other. This generalization of latent trait models also makes it possible to test specific hypotheses concerning the relationship between latent traits and external variables.

The most well-known application of the examination of the relationship between latent variables and external variables in item response theory concerns the study of item bias, i.e., the differences in the measurement model for specific subgroups; are there differences between subgroups with regard to item difficulties, etc. Multi-group analyses are performed to study the relationship between latent traits and external variables. Another interesting development in this field was prompted by the proposal of a number of latent trait models by Fischer (1974, 1983, 1987, 1989) and Fischer and Parzer (1991). They proposed latent trait models in which the item parameters are a function of certain structural characteristics of the items or a function of differences in the treatment that individuals have or have not received. This made possible the inclusion of the treatment variables which define the experimental design in the latent trait model.

An alternative method of studying the relationship between latent variables and external variables is to assign scores on the latent variables to individuals. Once scores on the latent trait are assigned, the relationship with external variables can be explored using routine statistical procedures. Interesting questions in this field concern the manner in which these latent scores can be estimated and the identifiability of latent scores. The problems that show up in estimating latent scores are similar to the problems with respect to estimating factor scores (see, for example, Steiger, 1979). These topics are also reviewed in this chapter.

[1]The term *external variables* should not be confused with the term *exogenous variables*. The latter term refers to variables that operate in a specified structural model only as independent variables, i.e., the variation in exogenous variables (as opposed to endogenous variables) is not explained within the model. External variables can function in such structural models both as exogenous variables and as endogenous variables.

## 5.2 Multidimensional latent trait models

There are several reasons for generalizing from unidimensional to multidimensional models. Firstly, of course, multidimensional models may provide a better fit simply because the items may tap more than one distinct latent trait. Secondly, multidimensional models offer opportunities for studying the relationships between latent variables; the association between these latent variables is not disturbed by measurement error. This can be helpful in analyzing data from panel research or non-equivalent control group designs, as is shown in several examples given by Hagenaars (1992a, 1992b). Finally, studying multidimensional models helps to clarify the concepts of unidimensionality and local independence. Some authors have used these concepts in a way that suggests that unidimensionality implies local independence, thereby suggesting that the assumptions of unidimensionality and local independence are identical (see, for example, Hambleton & Swaminathan, 1985 and Lord & Novick, 1968). Multidimensional models make the dimensionality structure of the model explicit. Within this structure, the assumption of local independence can be retained. Within the context of multidimensionality it is clear that the assumptions concerning dimensionality and local independence are separate ones. In fact, Hagenaars (1990) has used models with multidimensional latent variables to analyze *local dependence* between manifest indicators.

The discussion in this chapter starts with a general multidimensional latent trait model that follows directly from Bock's Nominal Response model. This model can be both extended and restricted in a number of ways. Some of these restricted models are reviewed briefly. Following this discussion, the estimation of the parameters in this model is dealt with. The section is concluded with the presentation of an example.

### 5.2.1 A general multidimensional latent trait model

**A model with two latent variables.** This discussion of multidimensional latent trait models starts with the example depicted in Figure 5.1. This figure shows how four manifest indicators, denoted $A$ through $D$, are influenced by two latent variables, designated $\theta^{(1)}$ and $\theta^{(2)}$. The model represented by this figure is an ordinary factor model with four manifest variables and two common factors.

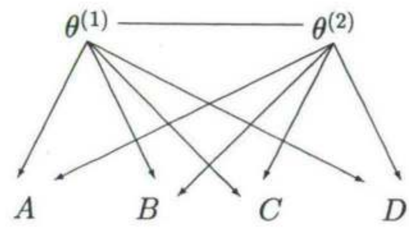


Figure 5.1: A model with two latent variables

Models like the one in Figure 5.1 can be used with both continuous and discrete latent variables. When it is assumed that the latent variables are discrete, the model in Figure 5.1 can be expressed as a log-linear model. It should be noted that the model in Figure 5.1 is not identified. However, the model used here is only for explanatory purposes, so that the problem of identification is not relevant. Using the notation introduced in Chapter 2, the logarithm of the joint probabilities can be written as

$$\begin{aligned} \ln p_{gk\ell jr\,s}^{ABCD\theta^{(1)}\theta^{(2)}} &= u + u_g^A + u_k^B + u_\ell^C + u_j^D + u_r^{\theta^{(1)}} + u_s^{\theta^{(2)}} \\ &\quad + u_{gr}^{A\theta^{(1)}} + u_{kr}^{B\theta^{(1)}} + u_{\ell r}^{C\theta^{(1)}} + u_{jr}^{D\theta^{(1)}} \\ &\quad + u_{gs}^{A\theta^{(2)}} + u_{ks}^{B\theta^{(2)}} + u_{\ell s}^{C\theta^{(2)}} + u_{js}^{D\theta^{(2)}} \\ &\quad + u_{r\,s}^{\theta^{(1)}\theta^{(2)}}. \end{aligned} \tag{5.1}$$

Expressing the common factor model in this way makes it clear that no specific assumptions are made with respect to the measurement level of the latent variables. The two-variable interaction terms are not linearized or restricted in any other fashion.

When the latent variables are variables on an interval level, the two-variable interactions may be linearized in the latent variables. Taking variable $A$ as an example, the following restrictions are made:

$$\begin{aligned} u_{gr}^{A\theta^{(1)}} &= u_g^{A^{(1)}} \cdot \theta_r^{(1)} \\ u_{gs}^{A\theta^{(2)}} &= u_g^{A^{(2)}} \cdot \theta_s^{(2)}. \end{aligned} \tag{5.2}$$

The same restrictions are made for the two-variable interactions pertaining to the other manifest variables. The two-variable interaction that describes the relationship between the two latent variables, i.e., $u_{r\ s}^{\theta^{(1)}\theta^{(2)}}$, is *not* restricted in the model in Equation 5.1. It is possible to impose specific restrictions on the association parameter $u_{r\ s}^{\theta^{(1)}\theta^{(2)}}$, but that requires certain extensions of the model that are discussed at a later point.

With the restrictions on the two-variable interactions describing the relationship between a latent and a manifest variable as defined above, the conditional probability for responding in category $g$ of item $A$, given scores $r$ and $s$ on the latent variables $\theta^{(1)}$ and $\theta^{(2)}$ respectively, is expressed as

$$p_{gr\ s}^{\bar{A}\theta^{(1)}\theta^{(2)}} = \frac{\exp(u_g^{A^{(1)}} \cdot \theta_r^{(1)} + u_g^{A^{(2)}} \cdot \theta_s^{(2)} + u_g^A)}{\sum_h \exp(u_h^{A^{(1)}} \cdot \theta_r^{(1)} + u_h^{A^{(2)}} \cdot \theta_s^{(2)} + u_h^A)}. \tag{5.3}$$

This expression for the conditional probability will now be reformulated by using a notation similar to the one used in Chapter 3 for defining the various latent trait models. The conditional response probability for responding in category $g$ of item $j$, given scores $r$ and $s$ on the latent variables $\theta^{(1)}$ and $\theta^{(2)}$, respectively, is thus

$$p_{jg|\theta_r^{(1)};\theta_s^{(2)}} = \frac{\exp(a_{jg}^{(1)} \cdot \theta_r^{(1)} + a_{jg}^{(2)} \cdot \theta_s^{(2)} + c_{jg})}{\sum_h \exp(a_{jh}^{(1)} \cdot \theta_r^{(1)} + a_{jh}^{(2)} \cdot \theta_s^{(2)} + c_{jh})}. \tag{5.4}$$

The parameters $a_{jg}^{(1)}$ and $a_{jg}^{(2)}$ are, of course, the two-variable interactions describing the relationship between the two latent variables on the one hand and the manifest item $j$ on the other. When Equation 5.4 is seen as defining a discretized latent trait model, these $a_{jg}$-parameters serve as discrimination parameters belonging to the respective latent variables. The category difficulty parameter $c_{jg}$ is equal to the one-variable interaction $u_g^A$ in the log-linear model defined earlier.

This multidimensional discretized latent trait model uses a linear combination of the latent scores as the working part of the model. Therefore, it is sometimes denoted as a *linear* MIRT (multidimensional item response theory) model (see, for example, Spray et al., 1990). An important characteristic of this model is that it is *compensatory* (Ansley and Forsyth, 1985). This term can easily be explained within the context of measuring abilities. It means that an individual who has a low score on one latent dimension can still have a high probability of responding with the correct answer, provided that this individual scores sufficiently high on the other dimension(s). Low ability on one particular dimension can be compensated for by high abilities on the other dimensions.

It is also possible to define non-compensatory multidimensional latent trait models. In that case, an elementary probability for responding in category $g$ of item $j$ is defined for each dimension by using an ordinary one-parameter or two-parameter logistic model. The overall probability of responding in category $g$ of item $j$ in the complete multidimensional space is then defined as the *product* of these elementary probabilities. It is clear that the minimum of the elementary probabilities sets an upper bound for the overall probability. The multiplicative character of this model makes it non-compensatory. Examples of this type of model can be found in Sympson (1978) and Embretson (1984). A comparison between compensatory and non-compensatory multidimensional latent trait models is made in Spray et al. (1990). In this chapter attention is restricted to the linear compensatory model because this model has direct links to the log-linear and latent class models.

**A model with $q$ latent variables.** The model in Equation 5.4 can now be generalized to an arbitrary number of $q$ latent variables. These $q$ latent variables are denoted:

$$\theta_1, \theta_2, \ldots, \theta_s, \ldots, \theta_q.$$

When these latent variables are again considered discrete, the number of latent nodes is equal to $T_1, T_2, \ldots, T_s, \ldots, T_q$ respectively. Furthermore, a vector $\boldsymbol{\psi}$ is defined whose elements are the ordered combinations of the latent nodes on the $q$ latent variables. The elements in this vector will be subscribed by the index $t$, so $t = 1, 2, \ldots, (T_1 \times T_2 \times \ldots \times T_s \times \ldots \times T_q)$. Finally, because the response probabilities depend on the $q$ different latent traits, a vector $\boldsymbol{a}_{jg}$ is defined consisting of $q$ different discrimination parameters pertaining to category $g$ of item $j$. Thus,

$$\boldsymbol{a}_{jg} = [a_{jg1}, a_{jg2}, \ldots, a_{jgs}, \ldots, a_{jgq}].$$

The response probabilities in the multidimensional model can then be expressed as

$$p_{jg|\psi_t} = \frac{\exp(\boldsymbol{a}'_{jg} \cdot \boldsymbol{\theta} + c_{jg})}{\sum_h \exp(\boldsymbol{a}'_{jh} \cdot \boldsymbol{\theta} + c_{jh})}. \tag{5.5}$$

The model in Equation 5.5 is the discretized multidimensional version of Bock's Nominal Response model. The continuous version of this model was studied by Reckase (1985) for dichotomous items under the heading *multidimensional two-parameter logistic* and by Bock et al. (1988), who also described a parametric MML method for estimating the parameters in this model. The same model is examined here using discretized latent traits, again in order to facilitate the comparisons with latent class models. Discretizing the latent traits again will lead to semi-parametric MML with fixed latent nodes.

**Estimation in the multidimensional Nominal Response model.** Estimation of the parameters in this model can be carried out using the EM algorithm. Because of the shift in notation that results from the multidimensional perspective, the main steps to be set with this procedure are outlined once again.

In the E-step the frequencies for the "complete" data-matrix, i.e., $e_{\nu\psi_t}$ are estimated. These frequencies pertain to the number of individuals responding with pattern $\nu$ and belonging to combination $t$ at the joint distribution formed by the $q$ latent variables; they are calculated in the E-step by

$$\hat{e}_{\nu\psi_t} = f_\nu \cdot \hat{p}_{\psi_t|\nu}.$$

The observed number of individuals responding with pattern $\nu$ is denoted by $f_\nu$ and the estimated probability that these individuals belong to cell $t$ in the joint latent distribution is represented by $\hat{p}_{\psi_t|\nu}$. Once the frequencies for the complete table are estimated in the E-step, new estimates for the parameters of the model are obtained in an M-step. ML estimates for the elements of $\boldsymbol{a}_{jg}$ as well as the category parameters $c_{jg}$ are found by maximizing the log-likelihood function. This problem can be solved numerically by a standard multivariate Newton-Raphson procedure. This estimation procedure is similar to the procedure used in logistic regression analysis. The first and second derivatives of the log-likelihood with respect to the parameters to be estimated are

$$
\begin{aligned}
\frac{\partial \ln L}{\partial \boldsymbol{a}_{jg}} &= \sum_t [q_{jg\psi_t} - e_{+\psi_t} \cdot p_{jg|\psi_t}] \cdot \boldsymbol{\theta} \\
\frac{\partial \ln L}{\partial c_{jg}} &= \sum_t [q_{jg\psi_t} - e_{+\psi_t} \cdot p_{jg|\psi_t}] \\
\frac{\partial^2 \ln L}{\partial \boldsymbol{a}_{jg} \partial \boldsymbol{a}'_{jg}} &= -\sum_t e_{+\psi_t} \cdot p_{jg|\psi_t} \cdot (1 - p_{jg|\psi_t}) \cdot (\boldsymbol{\theta}\boldsymbol{\theta}') \\
\frac{\partial^2 \ln L}{\partial c^2_{jg}} &= -\sum_t e_{+\psi_t} \cdot p_{jg|\psi_t} \cdot (1 - p_{jg|\psi_t}) \\
\frac{\partial^2 \log L}{\partial \boldsymbol{a}_{jg} \partial \boldsymbol{a}'_{jh}} &= \sum_t e_{+\psi_t} \cdot p_{jg|\psi_t} \cdot p_{jh|\psi_t} \cdot (\boldsymbol{\theta}\boldsymbol{\theta}') \\
\frac{\partial^2 \log L}{\partial c_{jg} \partial c_{jh}} &= \sum_t e_{+\psi_t} \cdot p_{jg|\psi_t} \cdot p_{jh|\psi_t} \\
\frac{\partial^2 \log L}{\partial \boldsymbol{a}_{jg} \partial c_{jg}} &= -\sum_t e_{+\psi_t} \cdot p_{jg|\psi_t} \cdot (1 - p_{jg|\psi_t}) \cdot \boldsymbol{\theta} \\
\frac{\partial^2 \log L}{\partial \boldsymbol{a}_{jg} \partial c_{jh}} &= \sum_t e_{+\psi_t} \cdot p_{jg|\psi_t} \cdot p_{jh|\psi_t} \cdot \boldsymbol{\theta}.
\end{aligned}
$$

In these expressions, the number of individuals belonging to latent class $\psi_t$ is denoted by $e_{+\psi_t}$ and the number of individuals in latent class $\psi_t$ responding in category $g$ of item $j$ is denoted by $q_{jg\psi_t}$.

With these first and second derivatives the provisional estimates for $a_{jg}$ and $c_{jg}$ can be improved at each M-step. Note that because each item has its own set of $a_{jg}$ and $c_{jg}$ parameters, the estimation of the item parameters can be carried out separately for each item. This is not the case for a number of other item response models. When equality constraints are imposed on parameters pertaining to different items, the parameters for these items have to be estimated simultaneously. After the estimates for the $a_{jg}$ and $c_{jg}$ parameters have been improved in the M-step, another E-step follows.

**An example of a multidimensional Nominal Response model.** In order to present an example of a multidimensional Nominal Response model, the data on women's liberation was analyzed again. For this example, the data for the total sample (both male and female respondents) were used. The reason is that there were problems estimating the parameters of the unidimensional Nominal response model for the male and female subsamples and these problems could only be worse in the multidimensional case.

A model with two latent variables was specified. For each latent variable, two latent nodes were used with values equal to $-1$ and $+1$. Table 5.1 presents the estimated values of the parameters in this model.

Table 5.1: Estimated parameters for a two-dimensional Nominal Response model with fixed latent nodes; items on women's liberation for the total sample

| | | $\hat{c}_{jg}$ | $\hat{a}_{jg}^{(1)}$ | $\hat{a}_{jg}^{(2)}$ |
|---|---|---|---|---|
| Item 1 | Cat. 1 | 0.20 | 1.28 | $-0.21$ |
| | Cat. 2 | 0.56 | 1.58 | 0.56 |
| Item 2 | Cat. 1 | 0.56 | 1.24 | $-0.17$ |
| | Cat. 2 | 1.71 | 2.36 | 1.24 |
| Item 3 | Cat. 1 | $-1.07$ | 1.44 | 0.26 |
| | Cat. 2 | $-1.01$ | 2.09 | 1.24 |
| Item 4 | Cat. 1 | $-5.33$ | $-0.05$ | $-5.86$ |
| | Cat. 2 | 3.15 | 1.14 | 1.08 |
| Item 5 | Cat. 1 | $-1.22$ | 1.15 | 0.40 |
| | Cat. 2 | $-0.90$ | 1.52 | 1.35 |

It should be noted that the solution presented in Table 5.1 is not identified because the matrix with $a$-parameters can be rotated. Therefore, it is difficult to give a substantive interpretation of the estimates in Table 5.1.

Table 5.2 gives the estimated joint distribution for the two latent variables. There is a moderate positive association between the two latent variables.

Not surprisingly, this model had a much better fit than the corresponding unidimensional model that was presented in Chapter 3. The value of the log-likelihood ratio was 280.19 with 229 degrees of freedom for the unidimensional model; the two-dimensional model yielded a value of $G^2 = 182.18$ with 209 degrees of freedom. However, the interpretation of the estimated parameters in the multidimensional model

Table 5.2: Estimated joint probability distribution for the latent variables

| | $\theta^{(2)} = -1$ | $\theta^{(2)} = +1$ | |
|---|---|---|---|
| $\theta_{(1)} = -1$ | .093 | .118 | .211 |
| $\theta_{(1)} = +1$ | .310 | .480 | .790 |
| | .403 | .598 | 1.0 |

was more complex than in the corresponding unidimensional model, in particular because the multidimensional solution was not uniquely identified.

**Generalization of the multidimensional Nominal Response model.** The semi-parametric MML estimation scheme is quite flexible and can be used to estimate the parameters in a great variety of different multidimensional latent trait models. These models can be obtained both by generalizing and by restricting the models in Equations 5.4 and 5.5. The generalization has to do with the possibility of including certain higher-order interactions in the model:

$$p_{jg|\theta_r^{(1)};\theta_s^{(2)}} = \frac{\exp(a_{jg}^{(1)} \cdot \theta_r^{(1)} + a_{jg}^{(2)} \cdot \theta_s^{(2)} + a_{jg}^{(1\times 2)} \cdot \theta_r^{(1)} \cdot \theta_s^{(2)} + c_{jg})}{\sum_h \exp(a_{jh}^{(1)} \cdot \theta_r^{(1)} + a_{jh}^{(2)} \cdot \theta_s^{(2)} + a_{jh}^{(1\times 2)} \cdot \theta_r^{(1)} \cdot \theta_s^{(2)} + c_{jh})}. \tag{5.6}$$

This model assumes that the response probability for item $j$ depends not only on the two main effects of the two latent variables $\theta^{(1)}$ and $\theta^{(2)}$, but also on a combined interaction effect of these latent variables. This three-variable log-linear interaction is denoted by $a_{jg}^{(1\times 2)}$.

Of course, with more than two latent variables, still more complex interactions can be included in the model. The possibility of generalizing the multidimensional latent trait model in this fashion was pointed out by Bock et al. (1988). This generalization is also straightforward within the context of the log-linear formulation as presented in Equation 5.1. This latter expression suggests other ways of generalizing the multidimensional model, for example, by including external variables. This topic was already briefly touched upon in the example presented in this section. It will be dealt with more extensively later on.

The model in 5.4 and 5.5 can also be restricted in a number of ways.

Some of these restrictions lead to models that have been suggested in the literature, but little credence has been given to the fact that these models can be seen as restricted versions of the general multidimensional latent trait model discussed in this section. Restricted multidimensional latent trait models are discussed in the next section.

### 5.2.2 Restricted multidimensional latent trait models

The model depicted in Figure 5.1 pertains to two common factors and four indicators. A natural method of restricting this model is to convert it from a exploratory factor model into a confirmatory factor model. In this type of model, indicators are no longer necessarily connected to all latent factors. In other words, some factor loadings are restricted to 0. An example of such a confirmatory factor model is depicted in Figure 5.2.

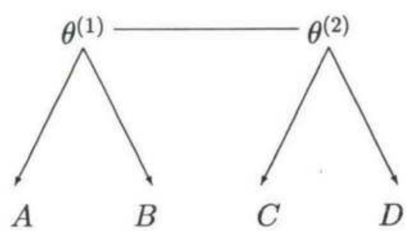


Figure 5.2: Restricted model with two latent variables

This model can be derived from the general model in 5.4 by restricting certain $a_{jg}$ parameters to equal 0. When it is assumed that the four manifest indicators each have three categories, Table 5.3 indicates which $a_{jg}$ parameters are set equal to zero and which parameters are to be estimated. The first are indicated by 0 and the latter by the $\times$-sign.

It is clear that by restricting certain $a_{jg}$ parameters to 0, a large number of potentially interesting models can be developed. In the example in Table 5.3, the restrictions regarding the discrimination parameters are applied in a manner which guarantees that each item is connected to exactly one latent variable. Another scheme of restrictions that could be applied is given in Table 5.4. This set of restrictions links the latent

Table 5.3: Restricted and free $a_{jg}$ parameters in the confirmatory factor model in Figure 5.3

| | Item $A$ | | | Item $B$ | | | Item $C$ | | | Item $D$ | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 |
| $\theta^{(1)}$ | 0 | $\times$ | $\times$ | 0 | $\times$ | $\times$ | 0 | 0 | 0 | 0 | 0 | 0 |
| $\theta^{(2)}$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | $\times$ | $\times$ | 0 | $\times$ | $\times$ |

Table 5.4: Restricted and free $a_{jg}$-parameters in the multidimensional polychotomous Rasch model

| | Item $A$ | | | Item $B$ | | | Item $C$ | | | Item $D$ | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 |
| $\theta^{(1)}$ | 0 | $\times$ | 0 | 0 | $\times$ | 0 | 0 | $\times$ | 0 | 0 | $\times$ | 0 |
| $\theta^{(2)}$ | 0 | 0 | $\times$ | 0 | 0 | $\times$ | 0 | 0 | $\times$ | 0 | 0 | $\times$ |

dimensions not to specific items, but to specific response categories. Such a model was first proposed by Rasch under the heading *Multidimensional Polychotomous Rasch model* (Rasch, 1961). It should be noted that in the Multidimensional Polychotomous Rasch model some additional assumptions are made. The $a_{jg}$ parameters that are denoted by the $\times$-sign in Table 5.4 are assumed to be equal so that they can be absorbed in the latent variables. This is, of course, not an unusual assumption in the context of Rasch modeling. The expression for the probability that individual $i$ responds in category $g$ of item $j$ becomes, in the Multidimensional Polychotomous Rasch model:

$$p_{jg|\theta_{ig}} = \frac{\exp(\theta_{ig} + c_{jg})}{\sum_h \exp(\theta_{ih} + c_{jh})}. \tag{5.7}$$

The Multidimensional Polychotomous Rasch model links the latent variables to different categories and not to different variables. Because in many practical applications it cannot be assumed that the relation between the latent variables and the probability for responding in a specific category will be the same for all manifest items, the Multidimensional Polychotomous Rasch model has only limited use.

Andersen (1983) generalized this multidimensional model for the case in which the $a_{jg}$ parameters are known but not necessarily equal to each other.

The possibility of assuming known $a_{jg}$ parameters can also be utilized for models other than Rasch' Multidimensional Polychotomous model. For example, by using an integer scoring function, a multidimensional Partial Credit model or a multidimensional Rating Scale model can be constructed. Parameters in those models can be estimated by applying either the semi-parametric MML or the CML. In the latter case, the latent variables are conditioned out of the likelihood by conditioning on the sufficient statistics. The latter are equal to the weighted total-scores, where the weights are assumed to be known (for example, successive integers for polytomous items) and the total-scores are calculated over a specified subset of the manifest items. In the model depicted in Figure 5.2, for example, one total-score is calculated over items $A$ and $B$ and the other over items $C$ and $D$. Kelderman (1988) showed how the parameters in such multidimensional latent trait models can be estimated by fitting specific log-linear models. The procedure is similar to the one for one-dimensional models outlined in the previous chapter. There it was shown that the logarithm of the probability that response pattern $\nu$ will be observed can be written as

$$\ln p_\nu = u_s^* + \sum_j x_{ij} \cdot u_j. \tag{5.8}$$

This log-linear model includes, aside from the $u_j$ item parameters the normalizing parameters $u_s^*$, which insure that the distribution of the total-scores will be reproduced. The multidimensional extension has the same form but now the $u_s^*$ parameters pertain to all possible combinations of the different total-scores which are sufficient statistics of the latent variables. As Kelderman (1988) illustrated, fitting such log-linear models is equivalent to fitting a model for quasi-independence in the incomplete table made up of the manifest items and all the distinct total-scores. These multidimensional extended Rasch models were also studied by Goodman (1990) and termed *Generalized Total-score Non-independence models.*

## 5.3 Relations with external variables

In this section attention is focused on the question of how relationships between the latent trait(s) and certain external variables can be investigated. Basically, there are two methods. Firstly, one can attempt to estimate some value or score upon the latent variable for each individual. Subsequently, these scores can be related to external variables. Another possibility is to include the external variables in the model and relating the latent traits to the manifest indicators. The distinction between these two procedures is similar to the difference between the estimation of factor scores and the correlation of these factor scores to external variables on the one hand, and the inclusion of the external variables in a covariance-structure model (e.g., a LISREL-type model) on the other hand. Both procedures are discussed below.

### 5.3.1 Estimating latent scores

In Chapter 4, a number of topics regarding the estimation of latent scores for individuals were discussed. For the discussion in this section, the distinction between the functional model and the structural model as introduced by de Leeuw and Verhelst (1986) is used again. This terminology was also introduced in Chapter 4. In other words, when the subjects are regarded as levels of a fixed factor, then the latent scores $\theta_i$ can be regarded as parameters that should be estimated. Technically speaking, these latent scores have a one-point distribution. This fixed-effects design is denoted as the functional model. It represents what Holland (1990a) has called the *stochastic subject rationale.* As was mentioned in Chapter 4, a natural method of estimating the latent scores in this context is to use joint maximum likelihood estimation (JML). However, JML estimates are not consistent, and in the case of small samples, they can be heavily biased.

In this book, the main emphasis regarding estimation methods is on MML. This procedure explicitly postulates the existence of some *distribution* of latent scores, and thus belongs in the context of the structural model. It is based on Holland's *random sampling rationale.* Estimation of the latent variable within this framework has a different emphasis than in the context of the structural model. Firstly, it should be noted that within the framework of MML, $\theta$ is merely a variable of integration and cannot be estimated (Holland, 1990a). The only thing that is interesting within the context of MML is the density function of the latent variable. In this context, the parameters of the latent distribution are estimated instead of the individual latent scores. Using parametric MML, the parameters of a known density function can be estimated and using semi-parametric MML, the unknown continuous density function can be approximated by an estimable discrete density function with the same moments as the unknown continuous function. However, it is still possible to estimate what Holland (1990a) calls *ability predictors*. These are functions that map response vectors $\nu$ into values on the latent variable. This is done by using the posterior distribution of $\theta$, given $\nu$. The same idea was already discussed with regard to latent class models in Chapter 2. The allocation probabilities $p_{\theta_t|\nu}$ were introduced in Chapter 2. On the basis of these probabilities, a modal allocation probability, denoted by $p^*_{\theta_t|\nu}$ was defined as the value of the latent class for which the allocation probability takes its maximum:

$$p^*_{\theta_t|\nu} = \max_{\theta_t} \left[ p_{\theta_1|\nu} \;\; p_{\theta_2|\nu} \;\; \cdots \;\; p_{\theta_T|\nu} \right].$$

This estimator is also called the *modal a posteriori estimator*, abbreviated MAP. The same estimator can be used (and, in fact, is used) within the context of latent trait models. However, because the latent nodes $t$ have numerical values that have meaning on an interval scale, it is also possible to use the *expected a posteriori estimator* (EAP), which is defined as

$$\bar{\theta}_{i|\nu} = \mathcal{E}(\theta_i \mid \nu) = \sum_t \theta_t \cdot p_{\theta_t|\nu}. \tag{5.9}$$

Both ability estimators, the MAP and the EAP, are optimal in a specific sense. Each minimizes its own loss function. The MAP minimizes the number of failures in predicting the latent class score on the basis of the observed response pattern $\nu$. As was seen in Chapter 2, this definition of a loss function can be used to define a measure indicating the strength of the relationship between the latent variable and the manifest indicators, i.e., the proportion misclassified $E$ (see Section 2.3.3). The loss function that is minimized by EAP is the sum of squared deviations between the true $\theta$-values and the estimates $\bar{\theta}_{i|\nu}$. In other words, EAP minimizes the mean-square error of $\mathcal{E}(\theta_i \mid \nu)$. Besides an a posteriori mode or mean, it is also possible to define an a posteriori median. Of course, this ability estimator minimizes the sum of the absolute differences between $\theta_i$ and $\bar{\theta}_{i|\nu}$.

The concept of estimating latent scores on the basis of the a posteriori distribution of $\theta$, given the response pattern $\nu$, seems attractive in the context of the structural model. There are, however, a number of problems as these ability estimators are not always identified. This identifiability problem with regard to the MAP is discussed by Hagenaars (1985). The problem can be sketched as follows. Let $\theta$ denote the latent variable in the population and $\tilde{\theta}$ the MAP estimator of $\theta$, based on the allocation probabilities in the population. Finally, $\tilde{\theta}^*$ indicates the MAP estimator of $\theta$ based on sample values.[2] Now, when $E \neq 0$, there will always be a discrepancy between $\theta$ and $\tilde{\theta}$. Even if all relevant information concerning the population was available, $\tilde{\theta}$ would never be a perfect substitute for $\theta$, unless all individuals were classified correctly (which is, of course, never the case). As Hagenaars showed, when $E \neq 0$, there are several different sets of "estimated" $\theta$-values that can be assigned to individuals.[3] All these sets are in agreement with the model parameters as well as the observed scores, but they will differ from one another, and occasionally they may even be correlated negatively. These sets provide no firm basis for analyzing the relationship between the latent variable and certain external variables. The problem is, however, not as serious as it may appear as the association between the latent and the manifest variables is stronger. In general, however, the latent scores $\theta_i$ cannot be precisely determined by using the allocation probabilities.

The identification problem with respect to EAP was discussed by Lindsay et al. (1991) and Holland (1990b). Holland illustrated what can be known about the posterior distribution of $\theta$ given $p_\nu$ in various situations. The contribution of Lindsay et al. (1991) focuses on the identifiability of the EAP in the Rasch model. When semi-parametric MML without fixed nodepoints is used, the latent distribution cannot be identified for a PD concordant case (see Chapter 4 for further details). Thus, in this same situation, the EAP will also not be identified. The different estimated mixing distributions that are congruent with the observed frequency distribution of the total-scores, will lead to dif-

[2] $\tilde{\theta}$ is the true a posteriori Bayes estimator, while $\tilde{\theta}^*$ is the empirical Bayes estimator.

[3] The term "estimated" refers not to the estimation of population parameters on the basis of sample data, but to the estimation of $\theta$ values on the basis of the population probabilities $p_{\theta_t|\nu}$.

ferent estimates of EAP.

The problem of an unidentified latent distribution can be handled by using fewer than $(n+1)/2$ node-points. This solves the identifiability problems, but it disturbs the relationship between semi-parametric MML estimation and CML estimation in Rasch-type latent trait models. In other words, this solution is restricted to the context of latent class analysis with linearized interactions between the latent variable and the manifest indicators. It does not apply to the semi-parametric estimation of parameters in latent trait models.

As was noted before, these identifiability problems can have a profound impact on the estimated relationship between the latent trait and external variables. Therefore, the option of including the external variables in the model in the manner in which this is done in LISREL models seems preferable. Before surveying this method in more detail, another method of estimating latent scores is discussed. After the item parameters have been estimated, it is possible to estimate $\theta_i$ by employing standard maximum likelihood methods. An ML estimate for $\theta_i$ can be found by maximizing the likelihood function based on $p_{\nu|\theta_i}$, the probability for responding with response pattern $\nu$, given a latent score $\theta_i$. This conditional probability can be expressed in terms of the response probabilities by

$$\prod_{j=1}^{n}\prod_{g=0}^{m_j} p_{jg|\theta_i}^{x_{\nu jg}},$$

and thus the log-likelihood based on $p_{\nu|\theta_i}$ equals

$$\sum_{j=1}^{n}\sum_{g=0}^{m_j}\left[x_{\nu jg}\cdot\log p_{jg|\theta_i}\right].$$

In these expressions, $x_{\nu jg}$ again is an indicator variable, defined by

$x_{\nu jg}=1$ if in response pattern $\nu$ item $j$ is responded to in category $g$,

$x_{\nu jg}=0$ otherwise.

The quantities $p_{jg|\theta_i}$ in these expressions are replaced by estimates for these response probabilities based on the estimates for the item parameters. The $\theta_i$ value which maximizes this log-likelihood can be found by using a standard univariate Newton-Raphson procedure. The first derivative of the log-likelihood with respect to $\theta_i$ is expressed as

$$\frac{\partial \log L}{\partial \theta_i} = \sum_j \sum_g \frac{\partial \log L}{\partial p_{jg|\theta_i}} \cdot \frac{\partial p_{jg|\theta_i}}{\partial \theta_i}.$$

Now:

$$\frac{\partial \log L}{\partial p_{jg|\theta_i}} = \frac{x_{\nu jg}}{p_{jg|\theta_i}}.$$

The expression for the first derivative of $p_{jg|\theta_i}$ to $\theta_i$ is dependent on the specific type of latent trait model under consideration. For the Nominal Response model, for example, the resulting expression is

$$\frac{\partial p_{jg|\theta_i}}{\partial \theta_i} = p_{jg|\theta_i} \left[ a_{jg} - \sum_{h=0}^{m_j} a_{jh} \cdot p_{jh|\theta_i} \right].$$

The first derivative of the log-likelihood with respect to $\theta_i$ thus becomes

$$\frac{\partial \log L}{\partial \theta_i} = \sum_{j=1}^{n} \sum_{g=0}^{m_j} a_{jg} \left( x_{\nu jg} - p_{jg|\theta_i} \right).$$

And the second derivative is equal to

$$\frac{\partial^2 \log L}{\partial \theta_i^2} = -\sum_j \left\{ \sum_g a_{jg}^2 \cdot p_{jg|\theta_i} - \left[ \sum_g a_{jg} \cdot p_{jg|\theta_i} \right]^2 \right\}.$$

The expressions for the first and second derivatives of the log-likelihood with respect to $\theta_i$ for the other latent trait models presented in Chapter 3 are presented in Appendix E.

Maximum likelihood estimators for $\theta_i$ are widely applied in the context of item response theory. These estimators do not use any information concerning the distribution of latent ability. In this sense, this method seems to belong more to the framework of the functional model than to that of the structural model. However, the method is employed in a number of computer programs using MML (for example, BILOG and MULTILOG). The fact that no information regarding the latent distribution is used can also be interpreted as an advantage, given the identification problems in estimating the latent distribution with semi-parametric MML. However, ML estimation of $\theta_i$ also has some drawbacks. Firstly, it is not possible to estimate $\theta_i$ for individuals with the two most extreme response patterns (i.e., in terms of dichotomous items: response patterns with all items answered incorrectly or all items answered correct). This problem could be circumvented by using parametric MML and including some assumptions concerning the latent population distribution in the ML estimation procedure for $\theta_i$. Most researchers, however, are not willing to make such restrictive assumptions with regard to the distribution of $\theta_i$. Secondly, the properties of ML estimators are asymptotic. Since the ML estimation of $\theta_i$ is dependent on the estimated item parameters, these asymptotic results will only be valid only if the number of *items* (and not individuals) is large. For research concerned with attitudinal data, this is rarely the case.

### 5.3.2 Including external variables in the model

In the past, efforts have been made to include external variables in the latent trait model in order to estimate directly the relation between the latent trait and these external variables. In this section, an overview of the main developments in this area is presented. Firstly, the possibility of relating the latent variable to external variables by way of a regression model is explored. Afterwards, attention is given to the situation in which the relations between the latent trait and the external variables are modeled through log- linear models. This type of analysis is particularly useful when the external variables are discrete. A special case of relating the latent variables to discrete external variables is the multiple-group analysis. In such an analysis, each category of the discrete external variable is regarded as a distinct group of subjects and the multiple-group analysis focusses on questions such as: Does the same measurement model hold in the various groups?; Are the relations between the latent variable and the manifest indicators the same in all groups?, etc. It is clear that these questions have a lot in common with the problem of relating the latent trait to external variables. Finally, some remarks are made concerning the situation in which the distinction between the different groups is not directly observed. In other words, one or more of the external variables is also latent.

#### 5.3.2.1 Relating the latent trait to external variables using a regression model

When external variables are included in a latent trait model, there are basically two routes one can follow. Firstly, the latent variable can be related to these external variables through ordinary regression methods. This results in LISREL-type models in which the measurement models are based on item response theory instead of factor analysis. This option for including external variables is appealing if the external variables are continuous. The other alternative is to use log-linear models in order to model the relations between the latent variable and the external variables. This alternative is dealt with in the next subsection.

In the context of latent class analysis, Dayton and Macready (1988) proposed a number of models which relate external variables to the latent proportions and/or the conditional response probabilities through logistic regression. In item response theory, one of the attempts to relate the external variables to the latent trait with the aid of regression methods, stems from Zwinderman (1991b). The model proposed by Zwinderman is an extension of the Rasch model, but Zwinderman claims that this extension can easily be applied to the two and three-parameter logistic models. Assuming that the latent trait is related to $m$ external variables $x_s$ $(s = 1, \ldots, m)$ through the linear model:

$$\theta_i = \boldsymbol{\beta}'\boldsymbol{x}_i + \epsilon_i,$$

in which $\boldsymbol{\beta}$ is a vector of length $m$ containing the unknown regression weights, $\boldsymbol{x}_i$ is a vector of length $m$ consisting of the scores of subject $i$ on the $m$ external variables, and $\epsilon_i$ is the error-term for subject $i$. The expression for the response probabilities becomes

$$p_{j1|\theta_i} = \frac{\exp(\beta_0 + \beta_1 \cdot x_{i1} + \cdots + \beta_m \cdot x_{im} + \epsilon_i + c_j)}{1 + \exp(\beta_0 + \beta_1 \cdot x_{i1} + \cdots + \beta_m \cdot x_{im} + \epsilon_i + c_j)}. \tag{5.10}$$

The regression parameters $\beta_s$, the item easiness parameters $c_j$ and the standard deviation $\sigma$ for the error terms $\epsilon_i$, can be estimated with an EM algorithm first proposed by Zwinderman, provided that some identification constraints are imposed on the $c_j$ parameters and the mean of the distribution of the $\epsilon_i$. As is common practice in regression analysis, the $\epsilon_i$ are assumed to be independent, identically distributed normal variates. The estimation procedure provides correct estimates both for the sampling design with fixed $x$ variables, as for the case where the $x$ variables are not fixed in advance but are sampled from some population distribution.

It is clear that this model can, in principle, be extended to cases with more than one distinct latent trait, as well as those with polytomous indicators. The further development of these logistic regression models makes possible covariance structure models with latent variables, in which the relations between the latent variables and its manifest indicators are modelled according to item response theory. Further research in this area looks very promising.

### 5.3.2.2 Relating the latent trait to external variables through log-linear models

If the external variables that are to be related to the latent trait are discrete or measured on a ordinal or nominal level, modelling the relations between the latent trait and the external variables can be handled comfortably by using log-linear models. This is especially true if the latent trait is discretized and the parameters in the model are estimated using semi-parametric MML. A simple model in which the latent variable $\theta$ is related to a given external variable $E$, is sketched in Figure 5.3. When

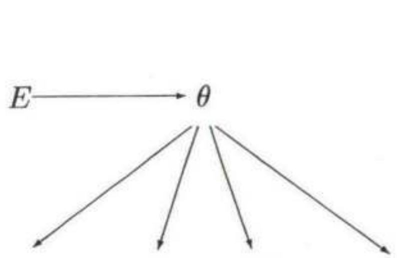


Figure 5.3: Relating an external variable to the latent trait $\theta$

$\theta$ is regarded as a discretized latent trait and the parameters describing the relations between $\theta$ on the one hand and the manifest indicators $A$, $B$, $C$ and $D$ on the other are linearized, then the model is equivalent to the following log-linear model:

$$\begin{aligned}\ln p_{gk\ell mte}^{ABCD\theta E} &= u + u_g^A + u_k^B + u_\ell^C + u_m^D + u_t^\theta + u_e^E \\ &\quad + u_g^{A^*} \cdot \theta_t + u_k^{B^*} \cdot \theta_t + u_\ell^{C^*} \cdot \theta_t + u_m^{D^*} \cdot \theta_t \\ &\quad + u_{et}^{E\theta}. \end{aligned} \tag{5.11}$$

Suggestions for using this log-linear formulation in order to relate the latent variable to external variables were already given by Haberman (1978). The parameters in this model can be estimated by including $E$ as an extra indicator. If, as is the case of Equation 5.11, the parameter describing the relationship between $\theta$ and the external variable $E$ is not linearized, the likelihood equation for estimating the parameter $u_{et}^{E\theta}$ in the M-step is different from the standard likelihood equations used in the Nominal Response model (see Chapter 4 for details). The likelihood equation that must be used in this case is equal to the standard equation used for log-linear models. These can be found in standard textbooks on log-linear modeling (see, for example, Agresti, 1990).

If, however, the $u_{et}^{E\theta}$ parameter is linearized, then all the parameters can be estimated in the same way, i.e., by using the estimation procedure and likelihood equations given in Chapter 4. Formally, there is no distinction between the manifest indicators ($A$, $B$, $C$ and $D$) and the external variable $E$. The only difference between the two types of variables in Figure 5.4, is that the external variable $E$ is assumed to have a causal influence on $\theta$, while the manifest indicators are seen as *caused by* $\theta$. This distinction is, however, irrelevant for the estimation of the parameters. Only the interpretation of the estimated parameters will be different for variable $E$ on the one hand and the manifest indicators on the other hand.

As an example, the data on women's liberation were analyzed again for both the male and the female samples.[4] Table 5.5 presents results concerning the fit of the Nominal Response model for the simultaneous group analysis of the male and female samples. Both with 3 and 4 latent nodes, the model gave an excellent fit according to the $G^2$ and an acceptable fit when the Pearson statistic was used. This means that the same measurement model can be used for both male and female respondents. In other words, this model assumes that there is no *differential*

[4]Only respondents fulfilling the educational requirements, were included in the analysis.

Table 5.5: Semi-parametric MML with fixed nodes; Nominal Response model for a simultaneous group analysis of the male and female samples

| Number of latent nodes | $G^2$ | $p$ | Pearson's $\chi^2$ | $p$ | $df$ |
|---|---|---|---|---|---|
| $T = 3$ | 445.74 | 0.687 | 510.77 | 0.054 | 461 |
| $T = 4$ | 442.30 | 0.715 | 501.80 | 0.087 | 460 |

*item functioning* (DIF).[5] The fact that the items on woman's liberation seemed to have the same characteristics for both male and female subjects could be seen as somewhat surprising, given the differences in difficulty order that could be observed between the two subsamples in the previous chapters.[6] The values of the estimated parameters again indicated that it was difficult to make a distinction between response categories 0 and 1 for item 4. This phenomenon was already observed in the different analyses for the male subsample in Chapter 3. The values of these estimated parameters are not reported here. Results regarding the relationship between the latent variable and the variable gender are, however, presented. Table 5.6 presents the distribution of both the males and the females over the four latent node points.

Table 5.6: Estimated latent distribution of the males and females (percentages)

| Latent nodes | Male | Female |
|---|---|---|
| −2.1 | 34.9 | 16.2 |
| −0.7 | 28.3 | 22.1 |
| 0.7 | 21.5 | 28.1 |
| 2.1 | 15.3 | 33.6 |

These relative distributions can easily be derived from the response probabilities for the variable gender. As in this model gender is treated as just another indicator, computer programs can provide response

[5] This expression is used to indicate a situation in which individuals with the same ability, but belonging to different groups, have systematically different scores. The term DIF originated in the field of testing research and is widely used in the context of cross-cultural research, for obvious reasons.

[6] It should, however, be remembered that these differences in the difficulty order were found in analyses of the dichotomous items. In the analysis reported here, the coding scheme with 3 categories was used.

probabilities for falling in the category female (or male), given specific latent node point positions. By renormalizing these response probabilities to percentages within the categories of the variable gender, the relative distributions over the four latent nodes are obtained. As one can see from the results in Table 5.6, women are more prone to belong to the higher values of the latent variable, indicating that women are more likely to have positive attitudes regarding women's liberation. The fact that the percentage of women with positive attitudes increases (and the percentage of males with positive attitudes therefore decreases), reflects the restriction that the parameter representing the relationship between $\theta$ and gender is linearized.

Including an external variable $E$ as an extra indicator leads to a rather restrictive model. Both the structure of the measurement model and the values of the parameters in this model are the same for all categories of the external variable. In other words, the relationship between the external variables and the manifest indicators is mediated entirely by the latent variable. This model is interesting because it excludes differential item functioning. This is a highly relevant hypothesis in the context of cross-cultural research. However, if this hypothesis has to be rejected, more flexible models are required to explain or describe the structure in the observed data.

Figure 5.4 shows three models that can be relevant in examining the relationship between an external variable ($E$) on the one hand, and a latent variable ($\theta$) and its indicator ($A$) on the other. For the sake of convenience, only one indicator is included in these figures.

In all three of these models, it is assumed that there is a relationship between the latent variable $\theta$ and the external variable $E$. As it is not very interesting at this point to make a distinction on the basis of whether $\theta$ influences $E$ or vice versa, these two variables are connected not by an arrow, but by a line. The presence of the relation between $\theta$ and $E$ has a clear interpretation: the distribution of the latent scores will be different in all subpopulations. It is also possible to formulate models in which there is no relation between $\theta$ and $E$, but these are not included here. Examples of such models can be found in, among others, Hagenaars (1990).

Model 5.4.a is the model described above. Both the structure of the model and the parameter values are the same for all categories of

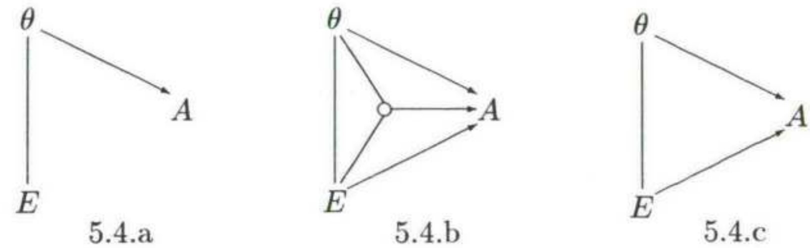


Figure 5.4: A number of models for the inclusion of an external variable in the latent trait model

$E$. The second model, model 5.4.b, allows for all possible interactions between the three kinds of variables ($\theta$, $E$ and $A$). The combined arrow emanating from $\theta$ and $E$ and pointing to $A$ indicates the presence of the three-variable interaction in the corresponding log-linear model. This is the most flexible model that can be postulated in the situation given. The only thing that is assumed to be constant across the different groupings of $E$ is the structure of the measurement model. Both the distributions of the latent and manifest variables and the relations between the latent and manifest variables vary over the different categories of $E$. Model 5.4.b can easily be tested by estimating this model separately for the various subgroups (i.e., the different levels of variable $E$) and summing the obtained values of the $G^2$. The resulting $G^2$ value can be used to test the null hypothesis that the specified model (i.e., model 5.4.b) is valid in the population. The degrees of freedom is equal to the sum of the separate numbers of degrees of freedom for the distinct tests in the subpopulations. In this example, this test cannot be carried out because the Nominal Response model could not be fitted for the male and female subsamples separately (see Chapter 3 for details). However, because the model in Figure 5.4.b is much more flexible than the model in 5.4.a and the latter model provided an acceptable fit, it can safely be concluded that model 5.4.b will also hold.

Models 5.4.a and 5.4.b can be regarded as the two most extreme models that could be postulated if an external variable is included in the latent trait model. Model 5.4.a postulates the complete absence of differential item functioning and is therefore the most restrictive. In model 5.4.b everything but the structure of the measurement model is allowed to vary over the different categories of the external variable. Within the spectrum of these two models, several other competing models can be proposed. One such model is illustrated in Figure 5.4.c. This model states that the relationship between the latent variable $\theta$ and the manifest indicator $A$ is the same for the different categories of $E$. In this model, the relationship between $\theta$ and $A$ is not influenced by $E$, indicating that the discrimination parameters (i.e., the degree of association between $\theta$ and $A$) are the same within the different subgroups. If the relations between the latent trait and the manifest indicators could be described by, for example, a Rasch model, then this Rasch model would hold in all subgroups. However, there is a direct relationship between $E$ and the indicators. This means that the item difficulties will vary over the different levels of $E$. Thus, model 5.4.c specifies the presence of differential item functioning. As stated earlier, there is also a relationship between the latent variable and the external variable, meaning that the distribution of the latent scores is different in the various subpopulations.

The estimation of the parameters in model 5.4.c takes place along the same lines as the estimation of the parameters in the multidimensional latent trait model. The response probabilities for model 5.4.c can be expressed as

$$p_{jg|\theta_t;E_e} = \frac{\exp(a_{jg}^{(1)} \cdot \theta_t + a_{jg}^{(2)} \cdot E_e + c_{jg})}{\sum_h \exp(a_{jh}^{(1)} \cdot \theta_t + a_{jh}^{(2)} \cdot E_e + c_{jh})}. \tag{5.12}$$

The logistic regression weight of the external variable $E$ is indicated here by $a_{jg}^{(2)}$, while the original discrimination parameter belonging to the latent variable $\theta$ is denoted by $a_{jg}^{(1)}$. $c_{jg}$ stands for the easiness parameter belonging to category $g$ of item $j$, while the subscription of the latent variable $\theta$ by the letter $t$ makes clear that the latent trait is discretized. This expression for the response probabilities is formally equivalent to the one for the multidimensional general latent trait model (see Equation 5.4). The expressions for the first and second derivatives are the same as those given in Section 5.2.1, but now the role of the second latent variable has been taken over by the external variable $E$.

The model in 5.12 can, of course, be extended to any number of external variables. Suppose there are $k$ external variables $E_r$ ($r = 1, \ldots k$) and the logistic regression weights for these variables are denoted by $q_{jgr}$. The observed score for individual $i$ on variable $E_r$ is written as $E_{ir}$ and the vector of observed scores on the $k$ external variables for subject $i$ is denoted by $\boldsymbol{E}_i$. In this case, the response probabilities can be written as

$$p_{jg|\theta_t;\mathbf{E_i}} = \frac{\exp(a_{jg} \cdot \theta_t + c_{jg} + \sum_{r=1}^{k} q_{jgr} \cdot E_{ir})}{\sum_h \exp(a_{jh} \cdot \theta_t + c_{jh} + \sum_{r=1}^{k} q_{jhr} \cdot E_{ir})}. \tag{5.13}$$

Assuming that all manifest items are dichotomous and that the $a_j$ discrimination parameters are all equal to each other and therefore can be incorporated in the latent variable $\theta$, the resulting model is a Rasch-type model, incorporating external variables. The expression for the response probabilities becomes

$$p_{j|\theta_t;\mathbf{E_i}} = \frac{\exp(\theta_t + c_j + \sum_{r=1}^{k} q_{jr} \cdot E_{ir})}{1 + \exp(\theta_t + c_j + \sum_{r=1}^{k} q_{jr} \cdot E_{ir})}. \tag{5.14}$$

It is interesting to compare this model with the *linear logistic test model* (LLTM) proposed by Fischer (1974, 1983). For this model the response probabilities are written as

$$p_{j|\theta_i;\mathbf{E_i}} = \frac{\exp(\theta_i - \sum_{r=1}^{k} q_{jr} \cdot E_{ir})}{1 + \exp(\theta_i - \sum_{r=1}^{k} q_{jr} \cdot E_{ir})}. \tag{5.15}$$

There are a number of differences between 5.14 and 5.15. Firstly, Fischer treats the latent variable $\theta$ as continuous, so it is subscribed by the subject index $i$, instead of $t$ as is the case in 5.14. Secondly, and more importantly, 5.15 has no separate $c_j$, item parameter while 5.14 does have such a parameter. The LLTM does assume that all differences in item difficulty can be explained by differences in the "external" variables $E_r$. These variables are thought of as representing the factors underlying the cognitive complexity of the items. In other examples in which the LLTM (or models that are derived from the LLTM) is used these $E_r$ variables embody the absence/presence of treatments subjects received. The values of these variables $E_r$ are therefore assumed to be known. The basic idea behind the LLTM is that all differences in item difficulties are due to differences in the structural characteristics of items or to differences in treatments subjects received that are implied by the research design. If the item difficulties are denoted by $b_j$, the LLTM assumes that

$$b_j = \sum_{r=1}^{k} q_{jr} \cdot E_r.$$

Thus, an important difference between the LLTM and the model defined by Equation 5.14 is that in the LLTM the $E_r$ variables denote differences between the *items*, while in model 5.14 the $E_{ir}$ variables indicate differences between individuals.

The fact that it is the item difficulty parameter $b_j$ rather than the easiness parameter $c_j$ that is reparameterized in the LLTM also explains why the $+$-sign in 5.14 is replaced by the $-$-sign in 5.15. This difference is irrelevant.

The LLTM was proposed by Fischer in order to make it possible to study the cognitive complexity of the items. The LLTM should help the researcher formulate theories regarding the elementary cognitive operations that are involved in responding to items, in essence, a micro-theory of learning. If the Rasch model is suited to identify items with DIF, the LLTM was to make it clear *why* these items show DIF. In practice, however, the LLTM has not proven successful. The reasons are obvious. In many situations it is not possible to explain all variance in item difficulty parameters by the linear decomposition in terms of cognitive factors or experimental treatments. Another reason, pointed out by Zwinderman (1991a), is that the cognitive operations often do not function according to the simple unidimensional model that is assumed in the LLTM.

The fact the LLTM uses a linear decomposition of the item difficulty parameters suggests that the LLTM is a special case of the Rasch model. Within this interpretation, the LLTM should have been surveyed in Chapter 3. However, as Fischer (1989) claims, the likelihood equations and uniqueness conditions of the LLTM make it, in a formal sense, more general than the Rasch model.

The differences between the LLTM, on the one hand, and the discretized latent trait models which incorporate external variables, on the other, are far from obvious. The theoretical foundation of the LLTM makes this model rather special. It is impossible to argue that the LLTM should be regarded as simply a special case of 5.14 in which the $c_j$ item parameters have been set equal to 0. However, the major drawback of the LLTM, i.e., the assumption that all variability within the items is due solely to the basic variables $E_r$, makes the models in 5.13 and 5.14 interesting competitors for describing and explaining the relationships between external variables and latent traits in cases in which items exhibit differential item functioning.

The differences between LLTM and the discretized latent trait model in Equation 5.14 are now clear. However, it is possible to develop models that combine the characteristics of the LLTM and the model in Equation 5.14. The details of such latent trait models will not be discussed here.

#### 5.3.2.3 Relating the latent trait to unobserved external variables

In the preceding section it was argued that relating a discrete external variable to the latent trait model is equivalent to performing a multi-group analysis, in which the distinction between the groups is ruled by the categorization of the external variable. At times, this idea of multi-group analysis may seem very attractive, even when the grouping variable is not observed. Introducing latent grouping variables makes possible the study of DIF in situations in which no grouping variable is available; it also makes it possible to study DIF with a correction for measurement error in observed grouping variables. The idea of accounting for heterogenity in the values of the model parameters among different subpopulations by introducing a latent variable is not new. It was applied in, for example, hazard models (Heckman and Singer, 1982), Markov models (van de Pol and de Leeuw, 1986) and the analysis of rankings (Croon, 1989; Croon and Luijkx, 1992). Recently, several suggestions have been made as to how to incorporate the idea of latent grouping variables in the latent trait models. All of these proposals focus on the Rasch model or the Partial Credit model because of the presence of known sufficient statistics in these models. This presence makes the estimation of parameters in these models much easier.

Mislevy and Verhelst (1990) proposed a latent trait model which takes into account the different solution strategies individuals can employ when responding to test items. The model is based on the logistic latent test model. It assumes that the item parameters can be expressed as functions of smaller numbers of parameters that pertain to these solution strategies. Furthermore, it is assumed that an individual uses the same strategy for all items, but it is unknown which strategy is used by a specific individual. It is necessary to make some assumptions with respect to the item difficulties in order to relate them to the basic parameters characterizing the different solution strategies. Mislevy and Verhelst indicated how these basic parameters, the proportions of individuals employing the different strategies as well as the latent distribution within each solution class, can be estimated using an MML method. The necessity of explicit assumptions concerning the way in which the basic strategy parameters are related to the item difficulties is both the strength and the weakness of this model. If an explicit theory on solution strategies and the way in which the item difficulties depend on these strategies is available, this model allows for the testing of quite specific hypotheses. However, in the absence of a priori theoretical notions, the model would appear to be not very useful.

Another attempt at incorporating unobserved grouping variables in latent trait models was made by Rost (1990, 1991). Rost's models assume the existence of several latent classes and within each latent class a latent trait model is specified. The latent trait models in these latent classes are formally equivalent (in each class the model is a Rasch model or in each class the response probabilities follow a Partial Credit model), but the item parameters in each class differ. The models proposed by Rost do not require a priori assumptions regarding the item or category parameters. The expression for the response probabilities in this model, if a Partial Credit model is specified, within each class is

$$p_{jg|\theta_i;u_r} = \frac{\exp(g \cdot \theta_{ir} + c_{jgr})}{\sum_h \exp(h \cdot \theta_{ir} + c_{jhr})}. \tag{5.16}$$

In this expression, the latent grouping variable is denoted by $u$ and an arbitrary category of this variable with $r$. The latent trait is subscribed by both the subject index $i$ and by the latent class index $r$ though this is not necessary as each subject belongs to just one latent class. As the category parameters vary over the latent classes, these parameters are also subscribed by $r$. In this model, the total-scores are the sufficient statistics for the parameters $\theta_{ir}$, provided an equidistant integer scoring system is used for the manifest items. By conditioning on these total-scores, a likelihood function can be obtained in which the $\theta_{ir}$ parameters have disappeared. The CML estimation procedure proposed by Rost uses the EM algorithm and provides estimates for the latent class specific parameters $c_{jgr}$, for the proportions of subjects in the latent classes $p_r$, and for the total-score distribution within each latent class. Rost claimed that this estimation method could be used with relatively large numbers of items. A drawback of this procedure is that no estimates can be obtained for the proportions scoring the two most extreme total-scores in each latent class. This is a direct consequence of the proposed CML procedure.

Finally, Kelderman and Macready (1990) suggested the use of log-linear and latent class models for examining the influence of both observed and unobserved grouping variables on differential item functioning. In Chapter 3, the log-linear Rasch model was dealt with extensively. The log-linear formulation of the Rasch model is equal to

$$\ln p_\nu = u_s^* + \sum_j x_{ij} \cdot u_j. \tag{5.17}$$

The $u_s^*$ parameters serve to reproduce the distribution of the total-score. The item difficulties are equal to $-u_j$, and the $x_{ij}$ quantities represent the observed score of subject $i$ responding to item $j$ (thus $x_{ij}$ equal eithers 0 or 1). The most attractive property of the log-linear formulation of the Rasch model is that the latent variable $\theta$ is eliminated entirely because of the conditioning on the sufficient statistic, i.e., the total-score variable. A manifest and/or latent grouping variable can easily be added to the log-linear model in 5.16. To illustrate this, the sole focus is on the addition of a latent grouping variable. Details concerning log- linear Rasch models with manifest grouping variables were extensively reviewed in Kelderman and Macready (1990). Two examples of log-linear Rasch models with a latent grouping variable are denoted in Table 5.7. The $n$ manifest indicators are denoted by $X_1$ through $X_n$, while the total-score and the latent grouping variable are indicated by $S$ and $U$, respectively. The notation used here is

Table 5.7: Log-linear Rasch models with a latent grouping variable

| Model | Fitted Marginals |
|---|---|
| Model I | $\{X_1\}, \ldots, \{X_n\}, \{SU\}$ |
| Model II | $\{X_1U\}, \ldots, \{X_nU\}, \{SU\}$ |

common in the literature on log-linear models. The marginals of the joint $X_1 \times \cdots X_n \times S \times U$ contingency table that are fitted under the model are indicated in Table 5.7. Fitting these marginals implies that the corresponding log-linear interactions as well as all interactions of lower-order that are contained in these marginals, are fitted by the model.[7]

The first model in Table 5.7 assumes that there is a relationship between the total-score and the latent grouping variable, but that there are no relations between this grouping variable and the manifest indicators. This model is, thus, equal to the model in Figure 5.4. The expressions for the logarithms of the joint probabilities can be shown to be a special case of the expression in Equation 5.11. This model, therefore, assumes a Rasch model with the same parameters in each latent class. There is no differential item functioning.

Model II in Table 5.7 takes into account the direct effects of the latent grouping variable on the manifest items. This model assumes a Rasch model within each latent class, but the item parameters differ in the individual latent groups. This model is closely related to the Rost model in 5.15 (the only difference, of course, is the fact that the model proposed by Rost assumes the more general Partial credit model, while Kelderman and Macready restricted their models to cases involving dichotomous items, i.e., the Rasch model). The model is also equivalent to the one shown in Figure 5.4.c, though it must be pointed out that the grouping variable considered in model II is latent.

It should be noted that the joint contingency table $X_1 \times \cdots X_n \times S \times U$ is incomplete because of the presence of structural zeros. The expected frequencies for this joint contingency table for the two models can be estimated using Hagenaars' program LCAG. Once these frequencies are

[7]Some one-variable parameters must be set equal to zero, in order to impose the necessary identifiability constraints. This is not indicated in Table 5.7, as this would needlessly complicate things. See Chapter 4 for details.

calculated, the estimation of the log-linear parameters, the item difficulties, and the total-score distribution in the latent classes is straightforward.

The procedure proposed by Kelderman and Macready is very flexible, but like the methods suggested by Rost, can only be applied to item response models with known sufficient statistics for the person parameters. The advantage of the log-linear methods over the CML procedures proposed by Rost is that the complete total-score distribution within the latent classes can be estimated. Rost's estimation scheme does not allow for the estimation of the total-score frequencies for the two most extreme categories within the latent groups. However, the drawback of the log-linear approach is also evident. If the latent trait $\theta$ is considered a random variable in each latent class, the log-linear models do not guarantee that the estimated item parameters will be compatible with the existence of a proper density function of $\theta$ within the various latent classes.

## 5.4 Causal models for latent and external variables

In both the multidimensional latent trait models and the models in which external variables are related to the latent traits it is possible to specify causal models for the relations between the latent and the manifest external variables. It can be assumed, for example, that the latent variables are independent or that the relation between the latent variables is linear. It is also possible to specify specific causal models for the relation between latent traits and manifest external variables. These discretized "LISREL"-type models have been proposed by Hagenaars et al. (1980) and Hagenaars (1988). Causal relationships between latent traits can be modeled using the *modified path analysis approach* proposed by Goodman (1973b). This type of model has been frequently applied by Hagenaars (1992a, 1992b), but in these applications the relations between the latent variables and the manifest indicators were governed by the classical latent class models; the latent variables were variables on a nominal scale. Theoretically, however, there is no problem whatsoever with applying the same kind of models to situations in which the latent variables are variables on an interval scale, and can be regarded as discretized latent traits. The same kind of causal models can be used when the relations between the latent trait and a number of given external variables are investigated. The specific adjustments that are needed in the M-step are clarified here through the discussion of an example with manifest external variables.

The estimation of parameters in such models requires a modified EM procedure. The model specified requires that certain restrictions be imposed on the joint distribution of the latent and manifest external variables before the next E-step is carried out. This can be dealt with by fitting the required log-linear model on the frequencies for the joint distribution, estimated in the former E-step. The latent×manifest frequencies adjusted by fitting the appropriate log-linear model, together with the improved estimates for the $a_{jg}$ and $c_{jg}$ parameters, are the input for the next E-step.[8]

A possible model that could be interesting is depicted in Figure 5.5. Supposing that the attitudes towards woman's liberation were measured during a particular number of years the model in Figure 5.2 could be analyzed if, besides the variable "time", also the variable "gender" were introduced in the analysis. One crucial assumption in this model

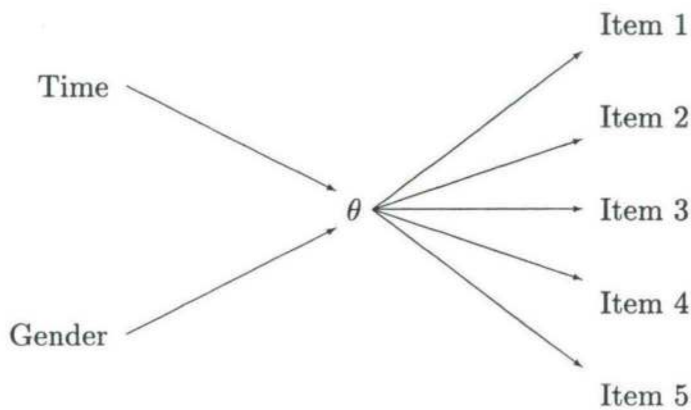


Figure 5.5: A model for analyzing change incorporating the variable gender

is that neither the variable "time", nor the variable "gender" has a direct effect on any of the manifest indicators. The same measurement model (including the same values for the parameters pertaining to this

[8]Note that the joint latent×manifest distribution only pertains to the manifest external variables and not the manifest indicators.

measurement model) is assumed for males and females, and this measurement model is also assumed to remain unchanged over a period of several years. Furthermore, the model assumes that the (latent) opinion towards woman's liberation will change over the years, hence, the relationship between the variable "time" and $\theta$. It is also assumed that there is a difference between males and females' opinions on woman's liberation. This is indicated by the arrow from "gender" towards $\theta$. Finally, this difference between males and females is not allowed to change over time and, therefore, the three-variable interaction between "gender", "time", and $\theta$ is absent. Because it is not very likely that the distribution of the variable "gender" will change over the years, no relation is assumed to exist between "gender" and "time". The latter assumption is essential.

Estimation of the parameters in this model requires a modification of the EM algorithm as described in Chapter 4 and in Section 5.2.1. For the sake of convenience the following notation is introduced. The latent variable is denoted by $\theta$ and an arbitrary category of this latent variable by $t$. The number of latent node points is equal to $T$. The variable "gender" is indicated by $\mathcal{G}$ and a category of this variable by $g$. The number of categories $\mathcal{G}$ is equal to 2. Finally, the variable "time" is symbolized by $\mathcal{R}$ and an arbitrary point of time by $r$ ($r = 1, \ldots, R$).

The joint frequency distribution for the latent and the manifest external variables has $T \times 2 \times R$ categories. The response probabilities that relate the five manifest items to all the categories of this joint distribution depend only on the value of $\theta$. Because $\theta$ only has a direct effect on the manifest items, the usual expressions for the response probabilities as they were presented in Chapter 3 can be used in the present example. Of course, when a model was used in which $\mathcal{G}$ and/or $\mathcal{R}$ was also directly related to the manifest indicators, then expressions for the response probabilities similar to those in the multidimensional latent trait models were relevant (see Equations 5.4 and 5.5).

In the E-step, the frequencies for the complete table were estimated using the observed counts for the (Item 1) $\times \cdots \times$ (Item 5) $\times\, \mathcal{G} \times \mathcal{R}$ contingency table, the initial estimates for $p_{\theta_t}$, and the initial estimates for the category-specific discrimination and easiness parameters (the $a_{jg}$ and $c_{jg}$ parameters). The procedure followed in this E-step is the one outlined in Chapter 4.

The initial estimates for the category-specific discrimination and easiness parameters can be improved in the next M-step using the usual likelihood equations. In the example presented in Figure 5.5 the manifest items depend only on $\theta$ and, therefore, the first and second derivatives for the unidimensional latent trait models can be used. These derivatives are presented in Appendix D. If direct effects from $\mathcal{G}$ and/or $\mathcal{R}$ are also assumed, expressions analogous to the ones for the multidimensional latent trait models must be used. These expressions for the Nominal Response model are given in section 5.2.1.

At the end of the M-step a new cycle is built in, in order to estimate the parameters pertaining to the causal model for the variables $\theta$, $\mathcal{G}$ and $\mathcal{R}$. It is necessary to include this new cycle because the ordinary EM algorithm will lead to estimates for the expected frequencies in the three-way table $\theta \times \mathcal{G} \times \mathcal{R}$ that will not satisfy the causal relations between the three variables sketched in Figure 5.5. Generally, a saturated model is needed to describe the relations between the three variables if no correction is made after the M-step, because all log-linear interactions can be present in the estimated expected frequencies. The extra cycle at the end of the M-step is meant to adjust the expected frequencies for the three-way table of the joint distribution of the latent and the manifest external variables to the specified causal model.

It is assumed that $\mathcal{G}$ and $\mathcal{R}$ are mutually independent and that $\theta$ depends on $\mathcal{G}$ and $\mathcal{R}$. Therefore, the independence between $\mathcal{G}$ and $\mathcal{R}$ should hold in the two-way table $\mathcal{G} \times \mathcal{R}$, and not in the three-way table $\theta \times \mathcal{G} \times \mathcal{R}$. It is incorrect to fit the log-linear model $\{\mathcal{G}\theta\}\{\mathcal{R}\theta\}$ for the three-way table. If there is no association between the two explanatory latent variables (no association in the two-way table $\mathcal{G} \times \mathcal{R}$), there will almost certainly be some interaction between these two latent variables in the three-way table. Goodman (1973b, 1973c) indicated that the parameters in such models, as well as the expected frequencies, can be estimated step-by-step by solving a system of log-linear equations. The suggestion that this modified path analysis be applied to systems with latent variables was made by Hagenaars (1985) and Hagenaars et al. (1980).

In short, the procedure is as follows. The expected frequencies for the joint latent variable that have been estimated in the $m$-th E-step are denoted by $\mathcal{F}_{tgr}(m)$. By summation over $\theta$, the expected frequencies for the two-way table $\mathcal{G} \times \mathcal{R}$, denoted by $\mathcal{F}_{gr}(m)$, are easily obtained. Because it is assumed that $\mathcal{G}$ and $\mathcal{R}$ are independent for the two-way table, improved estimates $\mathcal{F}_{gr}(m\dagger)$ for this two-way table can be found by fitting the log-linear model $\{\mathcal{G}\}\{\mathcal{R}\}$ for this table. For the three-way table, a simple log-linear model is assumed to hold with the only main effects being from $\mathcal{G}$ and $\mathcal{R}$ on $\theta$. There is no combined interaction of the two independent variables on $\theta$. It can be shown (Goodman, 1973b, 1973c) that the expected frequencies for this log-linear model can be found by fitting the log-linear model $\{\mathcal{G}\theta\}\{\mathcal{R}\theta\}\{\mathcal{G}\mathcal{R}\}$. This leads to improved estimates of the expected frequencies in the three-way table $\mathcal{F}_{tgr}(m\dagger)$. The causal model in Figure 5.5 assumes that both the model specified for the two-way table and the model for the three-way table are correct. The expected frequencies for the three-way table under the assumption that both models hold (indicated by $\mathcal{F}^*_{tgr}(m\dagger)$), can be calculated by

$$\mathcal{F}^*_{tgr}(m\dagger) = \mathcal{F}_{gr}(m\dagger) \cdot \frac{\mathcal{F}_{tgr}(m\dagger)}{\mathcal{F}_{gr}(m)}. \tag{5.18}$$

These frequencies $\mathcal{F}^*_{tgr}(m\dagger)$ are input for E-step $m+1$.

This example was presented in order to illustrate the possible strength of the incorporation of external variables in models for trend analysis. Of course, more complex models are also possible, for instance, models with several latent traits or models in which the manifest indicators are directly affected by the external variables. All these situations can be dealt with using the adjusted EM algorithm discussed here. For examples in which this procedure was used in the context of latent class analysis, the reader is referred to Hagenaars (1985, 1990, 1992a).

## 5.5 Evaluation

In this chapter, two important extensions of latent trait models were discussed. Firstly, multidimensional latent trait models were surveyed. The generalization of unidimensional to multidimensional models is straightforward, certainly if the latent trait is discretized and the analogy with log-linear models including latent variables can be established. The link with log-linear modeling also draws attention to the possibility of testing models with causal relations between the latent variables.

The methodology of the "modified path analysis" introduced by Goodman (1973b, 1973c) for log-linear models and applied to latent class analysis by Hagenaars (1985) can be useful in item response theory, provided that the latent traits are discretized.

Another important extension of latent trait models concerns the relations between latent traits and external variables. One way to deal with this problem is to estimate latent scores and to relate these scores to the external variables afterwards. This procedure has as its major drawback that the estimation of latent scores is hampered by identification problems. Besides, adequate estimation of latent scores depends heavily on the strength of the association between the latent variable and the manifest indicators.

An alternative procedure is to include the external variables in the model in the same fashion that this is done in LISREL-type models. The log-linear framework has proven to be quite useful in this respect. If the latent traits are discretized, it is easy to relate external variables to latent traits using the log- linear formulation. This approach does, however, lean heavily on the modified path analysis methodology mentioned above. One problem for applied researchers is that, at present, no standard software is available that can be used to test multidimensional latent trait models either with or without external variables. The development of such software warrants attention in the near future.

# Appendix A

# Maximum-likelihood estimation for log-linear and latent class models

## A.1 Maximum-likelihood estimation for log-linear models

In the context of log-linear models, the logarithm of the probability of observing a given arbitrary response pattern $\nu$ can be written as

$$\ln p_\nu = u + \sum_{s=1}^{q} u_s \cdot x_{\nu s}.$$

The probability of observing a given response pattern $\nu$ can be expressed as

$$\begin{aligned} p_\nu &= \exp\left(u + \sum_{s=1}^{q} u_s \cdot x_{\nu s}\right) \\ &= \exp(u) \exp\left(\sum_{s=1}^{q} u_s \cdot x_{\nu s}\right). \end{aligned}$$

Because the $\sum_\nu p_\nu = 1$, $p_\nu$ can also be written as

$$p_\nu = \frac{\exp\left(\sum_{s=1}^{q} u_s \cdot x_{\nu s}\right)}{\sum_\nu \exp\left(\sum_{s=1}^{q} u_s \cdot x_{\nu s}\right)}.$$

The first derivative of $p_\nu$ to a given parameter $u_s$ is, therefore,

$$\frac{\partial p_\nu}{\partial u_s} = p_\nu \left(x_{\nu s} - \sum_\nu x_{\nu s} \cdot p_\nu\right).$$

The likelihood can, using the multinomial distribution, be expressed as

$$L = \frac{N!}{\prod_\nu f_\nu} \prod_\nu p_\nu^{f_\nu},$$

resulting in the log-likelihood:

$$\ln L = C + \sum_\nu f_\nu \ln p_\nu.$$

The first derivative of the log-likelihood to a given $u_s$ parameter is equal to

$$\begin{aligned} \frac{\partial \ln L}{\partial u_s} &= \sum_\nu \frac{\partial \ln L}{\partial p_\nu} \cdot \frac{\partial p_\nu}{\partial u_s} \\ &= \sum_\nu \frac{f_\nu}{p_\nu} \cdot \frac{\partial p_\nu}{\partial u_s} \\ &= \sum_\nu f_\nu \left( x_{\nu s} - \sum_\nu x_{\nu s} \cdot p_\nu \right) \\ &= \sum_\nu (f_\nu - N \cdot p_\nu)\, x_{\nu s}. \end{aligned} \tag{A.1}$$

For the second derivatives, the following results can then easily be obtained:

$$\begin{aligned} \frac{\partial^2 \ln L}{\partial u_s \partial u_r} &= -N \sum_\nu x_{\nu s} \frac{\partial p_\nu}{\partial u_r} \\ &= -N \left[ \sum_\nu x_{\nu s} \cdot x_{\nu r} \cdot p_\nu - \left( \sum_\nu x_{\nu s} \cdot p_\nu \right) \cdot \left( \sum_\nu x_{\nu r} \cdot p_\nu \right) \right]. \end{aligned} \tag{A.2}$$

The first and second derivatives, as given in this section, can be used for the Newton-Raphson procedure to estimate the parameters in the log-linear model.

## A.2 Estimating the variance-covariance matrix for the parameters in the log-linear model

The expected information matrix or Fisher information matrix is defined as the expected value of the negative of the matrix of second-order derivatives of the log-likelihood with respect to the unknown model parameters. A general result can be derived from Equation A.1 which also holds in the context of latent class and latent trait models:

$$\frac{\partial^2 \ln L}{\partial u_s \partial u_r} = \sum_\nu \frac{f_\nu}{p_\nu} \cdot \frac{\partial^2 p_\nu}{\partial u_s \partial u_r} - \sum_\nu \frac{f_\nu}{(p_\nu)^2} \cdot \frac{\partial p_\nu}{\partial u_s} \cdot \frac{\partial p_\nu}{\partial u_r}.$$

The expected value of the first term on the right is

$$\begin{aligned} E\left[\sum_\nu \frac{f_\nu}{p_\nu} \cdot \frac{\partial^2 p_\nu}{\partial u_s \partial u_r}\right] &= N \sum_\nu \frac{p_\nu}{p_\nu} \cdot \frac{\partial^2 p_\nu}{\partial u_s \partial u_r} \\ &= N \frac{\partial^2 \left(\sum_\nu p_\nu\right)}{\partial u_k \partial u_\ell} = 0, \end{aligned}$$

since $\sum_\nu p_\nu = 1$ and the derivative of a constant is zero. For the second term in the expression for the information matrix, it is obvious that

$$-E\left[\sum_\nu \frac{f_\nu}{(p_\nu)^2} \cdot \frac{\partial p_\nu}{\partial u_s} \cdot \frac{\partial p_\nu}{\partial u_r}\right] = -N \sum_\nu \frac{1}{p_\nu} \cdot \frac{\partial p_\nu}{\partial u_s} \cdot \frac{\partial p_\nu}{\partial u_r}. \qquad \text{(A.3)}$$

The negative of this last result is a general expression for the expected information matrix. The inverse of this information matrix evaluated at the final estimates for the parameters provides estimates of the asymptotical variance-covariance matrix for the parameter estimates. However, the variance-covariance matrix for the parameter estimates can also be estimated by using the observed information matrix. The elements in this matrix are obtained by taking the negative of the expressions for the second derivatives of the log-likelihood with respect to the parameters to be estimated (see Equation A.2). The inverse of this matrix, evaluated for the final estimates of the $u$-terms, also provides estimates for the asymptotical variance-covariance matrix of the parameter estimates. Of course, the expected information matrix gives asymptotically the same results as the observed information matrix. In the case of log-linear models in which all of the variables are directly observed, there is no difference between the observed and the expected information matrix (see Haberman, 1977a and 1977b). This can be verified by substituting the results for the first derivatives of $p_\nu$ to $u_s$, as given in the preceding subsection, with the expression for the expected information matrix given above. It can also be seen that in the expression for the observed information matrix for log-linear models no terms appear that are subject to sampling fluctuations. Because the expected value of a given constant is equal to that constant, there is no difference between the expected and the observed information matrix in the case of log-linear models. There is, however, a difference between the two types of information matrices when one or more latent variables appear in the model, as is shown in the next section.

## A.3 Estimating the variance-covariance matrix for the parameters in the latent class model

In the latent class model, the variance-covariance matrix for the parameters can be estimated by using either the observed information matrix or the expected information matrix. The methods do provide somewhat different results. It should also be noted that the computation of the estimated variance-covariance matrix depends on the parameterization chosen. One possibility is to use the parameterization given by Lazarsfeld, which is also used by Goodman. In this case, the parameters to be estimated are the latent proportions and the conditional response probabilities. Another possibility is log-linear parameterization. Log-linear parameterization makes possinle the estimation of the variance-covariance matrix for the latent proportions and the log-linear parameters $u_{jg}$ and $u_{jg\theta_t}$.[1] If the estimated variance-covariance matrix of the parameters is to be based on the observed information matrix, it is necessary to derive the first and second derivatives of the log-likelihood to the parameters. The log-likelihood in the latent class model can be written as

$$\ln L = \sum_{\nu} f_\nu \cdot \ln p_\nu,$$

in which the probabilities for observing response pattern $\nu$ are given by

$$p_\nu = \sum_{t=1}^{T} p_{\nu|\theta_t} \cdot p_{\theta_t},$$

[1] One could, of course, also estimate the variance-covariance matrix for the log-linear parameters $u_{jg}$ and $u_{jg\theta_t}$ and the log-linear parameters $u_{\theta_t}$, instead of the latent proportions. However, in most practical applications the standard errors for the latent proportions are considered more important than the standard errors for the log-linear parameters $u_{\theta_t}$.

and the conditional probabilities $p_{\nu|\theta_t}$ are defined by

$$p_{\nu|\theta_t} = \prod_j \prod_g p_{jg|\theta_t}{}^{x_{\nu jg}}.$$

The conditional response probabilities $p_{jg|\theta_t}$ can be expressed as a function of the log-linear parameters $u_{jg}$ and $u_{jg\theta_t}$:

$$p_{jg|\theta_t} = \frac{\exp(u_{jg} + u_{jg\theta_t})}{\sum_g \exp(u_{jg} + u_{jg\theta t})}.$$

The way in which the indicator variables $x_{\nu jg}$ are defined is explained in Chapter 2. Furthermore, some notation is required for the estimated complete data. The estimated frequencies in the complete table are written as

$$e_{\nu\theta_t} = f_\nu \cdot p_{\theta_t|\nu},$$

in which the allocation probabilities $p_{\theta_t}$ can be found by using Bayes's theorem:

$$p_{\theta_t|\nu} = \frac{p_{\nu|\theta} \cdot p_{\theta_t}}{p_\nu}.$$

Occasionally, an expression is required for the expected frequency in latent class $\theta_t$ denoted by $e_{+\theta_t}$ and for the expected marginal frequency found in the two-way table made up of the latent variable $\theta$ and one manifest indicator $j$. The expected number of subjects responding in category $g$ of item $j$ and belonging to latent class $\theta_t$ is designated by $q_{jg\theta_t}$. These two expected frequencies can be found by

$$\begin{aligned} e_{+\theta_t} &= \sum_\nu e_{\nu\theta_t}, \\ q_{jg\theta_t} &= \sum_\nu e_{\nu\theta_t} \cdot x_{\nu jg}. \end{aligned}$$

Arbitrary categories for two items $j$ and $j'$ are denoted by $g$ and $h$, while two arbitrary latent classes are indicated by $t$ and $t'$. Using this notation, the following results are found for the first and second partial derivatives of the log-likelihood to the log-linear parameters $u_{jg}$ and $u_{jg\theta_t}$:

$$\frac{\partial \ln L}{\partial u_{jg}} = \frac{\partial \ln L}{\partial u_{jg\theta_t}} = \sum_t \left[ q_{jg\theta_t} - e_{+\theta_t} \cdot p_{jg|\theta_t} \right],$$

$$
\begin{aligned}
\frac{\partial \ln L}{\partial p_{\theta_t}} &= \sum_{\nu} \frac{f_\nu}{p_\nu}\left[p_{\nu|\theta_t} - p_{\nu|\theta_1}\right] \text{ for } t = 2, \ldots, T, \\
\frac{\partial \ln L}{\partial p_{jg|\theta_t}} &= \sum_{\nu} f_\nu \cdot p_{\theta_t|\nu} \cdot \frac{x_{\nu jg}}{p_{jg|\theta_t}}, \\
\frac{\partial^2 \ln L}{\partial u_{jg}^2} = \frac{\partial^2 \ln L}{\partial u_{jg\theta_t}^2} &= \frac{\partial^2 \ln L}{\partial u_{jg}\partial u_{jg\theta_t}} \\
&= -\sum_t e_{+\theta_t} \cdot p_{jg|\theta_t}(1 - p_{jg|\theta_t}), \\
\frac{\partial^2 \ln L}{\partial u_{jg}\partial u_{jh}} = \frac{\partial^2 \ln L}{\partial u_{jg}\partial u_{jg\theta_t}} &= \frac{\partial^2 \ln L}{\partial u_{jg\theta_t}\partial u_{jh\theta_t}} \\
&= \sum_t e_{+\theta_t} \cdot p_{jg|\theta_t} \cdot p_{jh|\theta_t}, \\
\frac{\partial^2 \ln L}{\partial p_{\theta_t}\partial p_{\theta_{t'}}} &= -\sum_{\nu} f_\nu (p_{\nu|\theta_t} - p_{\nu|\theta_1})(p_{\nu|\theta_{t'}} - p_{\nu|\theta_1}), \\
\frac{\partial^2 \ln L}{\partial {p_{jg|\theta_t}}^2} &= -\sum_{\nu} f_\nu \cdot \frac{x_{\nu jg}^2}{{p_{jg|\theta_t}}^2} \cdot p_{\theta_t|\nu}^2, \\
\frac{\partial^2 \ln L}{\partial p_{jg|\theta_t}\partial p_{j'h|\theta_t}} &= \sum_{\nu}\left[f_\nu \cdot \frac{x_{\nu jg} \cdot x_{\nu j'h}}{p_{jg|\theta_t} \cdot p_{j'h|\theta_t}} \cdot p_{\theta_t|\nu}(1 - p_{\theta_t|\nu})\right].
\end{aligned}
$$

These second order derivatives can be used to construct the observed information matrix for any of the parameterizations mentioned above.

In order to estimate the variance-covariance matrix for the parameters by the inverse of the expected information matrix, it is necessary to derive the first derivatives of $p_\nu$ to the parameters (see Equation A.3). The following results can be obtained:

$$
\begin{aligned}
\frac{\partial p_\nu}{\partial p_{jg|\theta_t}} &= p_{\theta_t} \cdot p_{\nu|\theta_t} \cdot \frac{x_{\nu jg}}{p_{jg|\theta_t}}, \\
\frac{\partial p_\nu}{\partial p_{\theta_t}} &= p_{\nu|\theta_t} - p_{\nu|\theta_1} \text{ for all } \theta_t \text{ except } \theta_1, \\
\frac{\partial p_\nu}{\partial u_{jg}} = \frac{\partial p_\nu}{\partial u_{jg\theta_t}} &= \sum_t p_{\theta_t} \cdot p_{\nu|\theta_t}\left[x_{\nu jg} - p_{jg|\theta_t}\right].
\end{aligned}
$$

# Appendix B

# Marginal distributions and response patterns with frequencies

## B.1 Wording of the items and category-scoring

The items used in the analyses of Chapters 2 and 3 were phrased as follows:

1. Women's liberation sets women against men.
2. It's better for a wife not to have a job because that always poses problems in the household, especially if there are children.
3. The most natural situation occurs when the man is the breadwinner and the woman runs the household and takes care of the children.
4. It isn't really as important for a girl to get a good education as it is for a boy.
5. A woman is better suited to raise small children than a man.

In the analyses in which three response categories were used, the meaning of these categories was as follows:

- category 0: agree entirely or agree,
- category 1: do not agree, do not disagree,
- category 2: do not agree or do not agree at all.

## B.2 Marginal distributions

Below are the marginal distributions for the five items concerning women's liberation with three categories. The marginal distributions are given for both male and female respondents who fullfilled the educational requirements mentioned in Chapter 2. The frequencies for the grand total of these two subsamples are also reported.

Table B.1: Women's liberation sets women against men

| | | Male | | Female | | Total | |
|---|---|---|---|---|---|---|---|
| | | *f* | % | *f* | % | *f* | % |
| | Category 0 | 98 | 18.1 | 96 | 16.2 | 194 | 17.1 |
| Item 1 | Category 1 | 171 | 31.5 | 142 | 24.0 | 313 | 27.6 |
| | Category 2 | 273 | 50.4 | 354 | 59.8 | 627 | 55.3 |
| | Sum | 542 | 100.0 | 592 | 100.0 | 1134 | 100.0 |

Table B.2: It's better for a wife not to have a job because that always poses problems in the household, especially if there are children

| | | Male | | Female | | Total | |
|---|---|---|---|---|---|---|---|
| | | *f* | % | *f* | % | *f* | % |
| | Category 0 | 67 | 12.4 | 51 | 8.6 | 118 | 10.4 |
| Item 2 | Category 1 | 87 | 16.1 | 81 | 13.7 | 168 | 14.8 |
| | Category 2 | 388 | 71.6 | 460 | 77.7 | 848 | 74.8 |
| | Sum | 542 | 100.0 | 592 | 100.0 | 1134 | 100.0 |

Table B.3: The most natural situation occurs when the man is the breadwinner and the woman runs the household and takes care of the children

| | | Male | | Female | | Total | |
|---|---|---|---|---|---|---|---|
| | | *f* | % | *f* | % | *f* | % |
| | Category 0 | 183 | 33.8 | 175 | 29.6 | 358 | 31.6 |
| Item 3 | Category 1 | 106 | 19.6 | 111 | 18.8 | 217 | 19.1 |
| | Category 2 | 253 | 46.7 | 306 | 51.7 | 559 | 49.3 |
| | Sum | 542 | 100.0 | 592 | 100.0 | 1134 | 100.0 |

Table B.4: It isn't really as important for a girl to get a good education as it is for a boy

| | | Male | | Female | | Total | |
|---|---|---|---|---|---|---|---|
| | | $f$ | % | $f$ | % | $f$ | % |
| | Category 0 | 22 | 4.1 | 19 | 3.2 | 41 | 3.6 |
| Item 4 | Category 1 | 37 | 6.8 | 19 | 3.2 | 56 | 4.9 |
| | Category 2 | 483 | 89.1 | 554 | 93.6 | 1037 | 91.4 |
| | Sum | 542 | 100.0 | 592 | 100.0 | 1134 | 100.0 |

Table B.5: A woman is better suited to raise small children than a man

| | | Male | | Female | | Total | |
|---|---|---|---|---|---|---|---|
| | | $f$ | % | $f$ | % | $f$ | % |
| | Category 0 | 266 | 49.1 | 145 | 24.5 | 411 | 36.2 |
| Item 5 | Category 1 | 92 | 17.0 | 109 | 18.4 | 201 | 17.7 |
| | Category 2 | 184 | 33.9 | 338 | 57.1 | 522 | 46.0 |
| | Sum | 542 | 100.0 | 592 | 100.0 | 1134 | 100.0 |

## B.3 Response patterns with frequencies

The frequencies for the response patterns for the same samples for which the marginal distributions were presented in the previous section, are given on the following pages. Again, data are provided separately for male and female respondents, and for the total sample.

Table B.6: Response patterns with frequencies; male respondents

| Item 1=0 | | | | | | Item 1=1 | | | | | | Item 1=2 | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Item 2=0 | | Item 2=1 | | Item 2=2 | | Item 2=0 | | Item 2=1 | | Item 2=2 | | Item 2=0 | | Item 2=1 | | Item 2=2 | |
| Item 345 | | Item 345 | | Item 345 | | Item 345 | | Item 345 | | Item 345 | | Item 345 | | Item 345 | | Item 345 | |
| 000 | 8 | 000 | 2 | 000 | 2 | 000 | 0 | 000 | 1 | 000 | 1 | 000 | 1 | 000 | 0 | 000 | 0 |
| 001 | 0 | 001 | 1 | 001 | 0 | 001 | 0 | 001 | 0 | 001 | 0 | 001 | 0 | 001 | 0 | 001 | 0 |
| 002 | 0 | 002 | 0 | 002 | 0 | 002 | 0 | 002 | 0 | 002 | 0 | 002 | 0 | 002 | 1 | 002 | 0 |
| 010 | 6 | 010 | 2 | 010 | 0 | 010 | 4 | 010 | 2 | 010 | 3 | 010 | 1 | 010 | 3 | 010 | 0 |
| 011 | 1 | 011 | 0 | 011 | 0 | 011 | 1 | 011 | 0 | 011 | 2 | 011 | 0 | 011 | 0 | 011 | 1 |
| 012 | 0 | 012 | 0 | 012 | 0 | 012 | 1 | 012 | 0 | 012 | 0 | 012 | 0 | 012 | 0 | 012 | 0 |
| 020 | 13 | 020 | 7 | 020 | 13 | 020 | 6 | 020 | 8 | 020 | 20 | 020 | 9 | 020 | 6 | 020 | 20 |
| 021 | 0 | 021 | 0 | 021 | 2 | 021 | 0 | 021 | 0 | 021 | 5 | 021 | 2 | 021 | 1 | 021 | 6 |
| 022 | 3 | 022 | 1 | 022 | 6 | 022 | 0 | 022 | 2 | 022 | 3 | 022 | 1 | 022 | 2 | 022 | 3 |
| 100 | 0 | 100 | 0 | 100 | 0 | 100 | 0 | 100 | 0 | 100 | 0 | 100 | 0 | 100 | 0 | 100 | 0 |
| 101 | 0 | 101 | 0 | 101 | 0 | 101 | 0 | 101 | 0 | 101 | 0 | 101 | 0 | 101 | 0 | 101 | 0 |
| 102 | 0 | 102 | 0 | 102 | 0 | 102 | 0 | 102 | 0 | 102 | 0 | 102 | 0 | 102 | 1 | 102 | 0 |
| 110 | 0 | 110 | 1 | 110 | 0 | 110 | 0 | 110 | 1 | 110 | 0 | 110 | 0 | 110 | 0 | 110 | 1 |
| 111 | 0 | 111 | 0 | 111 | 0 | 111 | 0 | 111 | 0 | 111 | 0 | 111 | 0 | 111 | 0 | 111 | 1 |
| 112 | 0 | 112 | 0 | 112 | 1 | 112 | 0 | 112 | 0 | 112 | 0 | 112 | 0 | 112 | 0 | 112 | 0 |
| 120 | 0 | 120 | 1 | 120 | 1 | 120 | 1 | 120 | 7 | 120 | 19 | 120 | 0 | 120 | 7 | 120 | 17 |
| 121 | 0 | 121 | 1 | 121 | 3 | 121 | 0 | 121 | 4 | 121 | 5 | 121 | 0 | 121 | 3 | 121 | 8 |
| 122 | 0 | 122 | 0 | 122 | 1 | 122 | 0 | 122 | 1 | 122 | 9 | 122 | 0 | 122 | 0 | 122 | 12 |
| 200 | 0 | 200 | 0 | 200 | 0 | 200 | 0 | 200 | 0 | 200 | 0 | 200 | 0 | 200 | 1 | 200 | 0 |
| 201 | 0 | 201 | 0 | 201 | 0 | 201 | 0 | 201 | 0 | 201 | 0 | 201 | 0 | 201 | 0 | 201 | 1 |
| 202 | 0 | 202 | 0 | 202 | 0 | 202 | 0 | 202 | 0 | 202 | 0 | 202 | 0 | 202 | 0 | 202 | 2 |
| 210 | 0 | 210 | 0 | 210 | 1 | 210 | 0 | 210 | 0 | 210 | 1 | 210 | 0 | 210 | 0 | 210 | 1 |
| 211 | 0 | 211 | 0 | 211 | 0 | 211 | 0 | 211 | 0 | 211 | 0 | 211 | 0 | 211 | 0 | 211 | 0 |
| 212 | 0 | 212 | 0 | 212 | 0 | 212 | 0 | 212 | 0 | 212 | 2 | 212 | 0 | 212 | 0 | 212 | 0 |
| 220 | 1 | 220 | 2 | 220 | 9 | 220 | 2 | 220 | 4 | 220 | 11 | 220 | 0 | 220 | 4 | 220 | 35 |
| 221 | 0 | 221 | 0 | 221 | 2 | 221 | 0 | 221 | 1 | 221 | 14 | 221 | 2 | 221 | 1 | 221 | 24 |
| 222 | 1 | 222 | 0 | 222 | 6 | 222 | 0 | 222 | 0 | 222 | 30 | 222 | 3 | 222 | 8 | 222 | 84 |

Table B.7: Response patterns with frequencies; female respondents

| Item 1=0 | | | | | | Item 1=1 | | | | | | Item 1=2 | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Item 2=0 | | Item 2=1 | | Item 2=2 | | Item 2=0 | | Item 2=1 | | Item 2=2 | | Item 2=0 | | Item 2=1 | | Item 2=2 | |
| Item 345 | | Item 345 | | Item 345 | | Item 345 | | Item 345 | | Item 345 | | Item 345 | | Item 345 | | Item 345 | |
| 000 | 5 | 000 | 1 | 000 | 0 | 000 | 0 | 000 | 0 | 000 | 0 | 000 | 0 | 000 | 1 | 000 | 0 |
| 001 | 0 | 001 | 0 | 001 | 0 | 001 | 0 | 001 | 0 | 001 | 1 | 001 | 0 | 001 | 1 | 001 | 0 |
| 002 | 0 | 002 | 0 | 002 | 0 | 002 | 0 | 002 | 0 | 002 | 0 | 002 | 1 | 002 | 0 | 002 | 0 |
| 010 | 4 | 010 | 3 | 010 | 0 | 010 | 1 | 010 | 2 | 010 | 1 | 010 | 0 | 010 | 0 | 010 | 0 |
| 011 | 0 | 011 | 0 | 011 | 0 | 011 | 0 | 011 | 1 | 011 | 0 | 011 | 0 | 011 | 1 | 011 | 1 |
| 012 | 0 | 012 | 0 | 012 | 0 | 012 | 0 | 012 | 0 | 012 | 0 | 012 | 0 | 012 | 0 | 012 | 0 |
| 020 | 10 | 020 | 6 | 020 | 15 | 020 | 5 | 020 | 9 | 020 | 7 | 020 | 2 | 020 | 4 | 020 | 15 |
| 021 | 2 | 021 | 2 | 021 | 2 | 021 | 2 | 021 | 4 | 021 | 3 | 021 | 1 | 021 | 2 | 021 | 14 |
| 022 | 2 | 022 | 1 | 022 | 4 | 022 | 0 | 022 | 6 | 022 | 6 | 022 | 1 | 022 | 4 | 022 | 22 |
| 100 | 1 | 100 | 0 | 100 | 0 | 100 | 0 | 100 | 1 | 100 | 0 | 100 | 1 | 100 | 0 | 100 | 0 |
| 101 | 0 | 101 | 0 | 101 | 0 | 101 | 0 | 101 | 0 | 101 | 1 | 101 | 0 | 101 | 0 | 101 | 1 |
| 102 | 0 | 102 | 0 | 102 | 0 | 102 | 0 | 102 | 0 | 102 | 0 | 102 | 0 | 102 | 0 | 102 | 1 |
| 110 | 2 | 110 | 1 | 110 | 0 | 110 | 0 | 110 | 1 | 110 | 0 | 110 | 0 | 110 | 0 | 110 | 0 |
| 111 | 0 | 111 | 0 | 111 | 0 | 111 | 0 | 111 | 0 | 111 | 0 | 111 | 0 | 111 | 0 | 111 | 0 |
| 112 | 0 | 112 | 0 | 112 | 0 | 112 | 0 | 112 | 0 | 112 | 0 | 112 | 0 | 112 | 0 | 112 | 0 |
| 120 | 2 | 120 | 2 | 120 | 3 | 120 | 1 | 120 | 0 | 120 | 4 | 120 | 0 | 120 | 3 | 120 | 6 |
| 121 | 0 | 121 | 1 | 121 | 1 | 121 | 0 | 121 | 3 | 121 | 10 | 121 | 0 | 121 | 1 | 121 | 10 |
| 122 | 0 | 122 | 0 | 122 | 4 | 122 | 2 | 122 | 3 | 122 | 15 | 122 | 0 | 122 | 5 | 122 | 25 |
| 200 | 0 | 200 | 0 | 200 | 0 | 200 | 0 | 200 | 0 | 200 | 0 | 200 | 0 | 200 | 0 | 200 | 0 |
| 201 | 0 | 201 | 0 | 201 | 0 | 201 | 0 | 201 | 0 | 201 | 1 | 201 | 0 | 201 | 0 | 201 | 0 |
| 202 | 0 | 202 | 0 | 202 | 0 | 202 | 0 | 202 | 0 | 202 | 0 | 202 | 0 | 202 | 1 | 202 | 1 |
| 210 | 0 | 210 | 0 | 210 | 0 | 210 | 0 | 210 | 0 | 210 | 1 | 210 | 0 | 210 | 0 | 210 | 0 |
| 211 | 0 | 211 | 0 | 211 | 0 | 211 | 0 | 211 | 0 | 211 | 0 | 211 | 0 | 211 | 0 | 211 | 0 |
| 212 | 0 | 212 | 0 | 212 | 0 | 212 | 0 | 212 | 0 | 212 | 0 | 212 | 0 | 212 | 0 | 212 | 0 |
| 220 | 1 | 220 | 1 | 220 | 0 | 220 | 0 | 220 | 0 | 220 | 4 | 220 | 1 | 220 | 3 | 220 | 15 |
| 221 | 0 | 221 | 1 | 221 | 2 | 221 | 0 | 221 | 1 | 221 | 9 | 221 | 1 | 221 | 2 | 221 | 27 |
| 222 | 1 | 222 | 2 | 222 | 14 | 222 | 0 | 222 | 0 | 222 | 37 | 222 | 2 | 222 | 1 | 222 | 177 |

Table B.8: Response patterns with frequencies; male and female respondents

| Item 1=0 | | | | | | Item 1=1 | | | | | | Item 1=2 | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Item 2=0 | | Item 2=1 | | Item 2=2 | | Item 2=0 | | Item 2=1 | | Item 2=2 | | Item 2=0 | | Item 2=1 | | Item 2=2 | |
| Item 345 | | Item 345 | | Item 345 | | Item 345 | | Item 345 | | Item 345 | | Item 345 | | Item 345 | | Item 345 | |
| 000 | 13 | 000 | 3 | 000 | 2 | 000 | 0 | 000 | 1 | 000 | 1 | 000 | 1 | 000 | 1 | 000 | 0 |
| 001 | 0 | 001 | 1 | 001 | 0 | 001 | 0 | 001 | 0 | 001 | 1 | 001 | 0 | 001 | 1 | 001 | 0 |
| 002 | 0 | 002 | 0 | 002 | 0 | 002 | 0 | 002 | 0 | 002 | 0 | 002 | 1 | 002 | 1 | 002 | 0 |
| 010 | 10 | 010 | 5 | 010 | 0 | 010 | 5 | 010 | 4 | 010 | 4 | 010 | 1 | 010 | 3 | 010 | 0 |
| 011 | 1 | 011 | 0 | 011 | 0 | 011 | 1 | 011 | 1 | 011 | 2 | 011 | 0 | 011 | 1 | 011 | 2 |
| 012 | 0 | 012 | 0 | 012 | 0 | 012 | 1 | 012 | 0 | 012 | 0 | 012 | 0 | 012 | 0 | 012 | 0 |
| 020 | 23 | 020 | 13 | 020 | 28 | 020 | 11 | 020 | 17 | 020 | 27 | 020 | 11 | 020 | 10 | 020 | 35 |
| 021 | 2 | 021 | 2 | 021 | 4 | 021 | 2 | 021 | 4 | 021 | 8 | 021 | 3 | 021 | 3 | 021 | 20 |
| 022 | 5 | 022 | 2 | 022 | 10 | 022 | 0 | 022 | 8 | 022 | 9 | 022 | 2 | 022 | 6 | 022 | 25 |
| 100 | 1 | 100 | 0 | 100 | 0 | 100 | 0 | 100 | 1 | 100 | 0 | 100 | 1 | 100 | 0 | 100 | 0 |
| 101 | 0 | 101 | 0 | 101 | 0 | 101 | 0 | 101 | 0 | 101 | 1 | 101 | 0 | 101 | 0 | 101 | 1 |
| 102 | 0 | 102 | 0 | 102 | 0 | 102 | 0 | 102 | 0 | 102 | 0 | 102 | 0 | 102 | 1 | 102 | 1 |
| 110 | 2 | 110 | 2 | 110 | 0 | 110 | 0 | 110 | 2 | 110 | 0 | 110 | 0 | 110 | 0 | 110 | 1 |
| 111 | 0 | 111 | 0 | 111 | 0 | 111 | 0 | 111 | 0 | 111 | 0 | 111 | 0 | 111 | 0 | 111 | 1 |
| 112 | 0 | 112 | 0 | 112 | 1 | 112 | 0 | 112 | 0 | 112 | 0 | 112 | 0 | 112 | 0 | 112 | 0 |
| 120 | 2 | 120 | 3 | 120 | 4 | 120 | 2 | 120 | 7 | 120 | 23 | 120 | 0 | 120 | 10 | 120 | 23 |
| 121 | 0 | 121 | 2 | 121 | 4 | 121 | 0 | 121 | 7 | 121 | 15 | 121 | 0 | 121 | 4 | 121 | 18 |
| 122 | 0 | 122 | 0 | 122 | 5 | 122 | 2 | 122 | 4 | 122 | 24 | 122 | 0 | 122 | 5 | 122 | 37 |
| 200 | 0 | 200 | 0 | 200 | 0 | 200 | 0 | 200 | 0 | 200 | 0 | 200 | 0 | 200 | 1 | 200 | 0 |
| 201 | 0 | 201 | 0 | 201 | 0 | 201 | 0 | 201 | 0 | 201 | 1 | 201 | 0 | 201 | 0 | 201 | 1 |
| 202 | 0 | 202 | 0 | 202 | 0 | 202 | 0 | 202 | 0 | 202 | 0 | 202 | 0 | 202 | 1 | 202 | 3 |
| 210 | 0 | 210 | 0 | 210 | 1 | 210 | 0 | 210 | 0 | 210 | 2 | 210 | 0 | 210 | 0 | 210 | 1 |
| 211 | 0 | 211 | 0 | 211 | 0 | 211 | 0 | 211 | 0 | 211 | 0 | 211 | 0 | 211 | 0 | 211 | 0 |
| 212 | 0 | 212 | 0 | 212 | 0 | 212 | 0 | 212 | 0 | 212 | 2 | 212 | 0 | 212 | 0 | 212 | 0 |
| 220 | 2 | 220 | 3 | 220 | 9 | 220 | 2 | 220 | 4 | 220 | 15 | 220 | 1 | 220 | 7 | 220 | 50 |
| 221 | 0 | 221 | 1 | 221 | 4 | 221 | 0 | 221 | 2 | 221 | 23 | 221 | 3 | 221 | 3 | 221 | 51 |
| 222 | 2 | 222 | 2 | 222 | 20 | 222 | 0 | 222 | 0 | 222 | 67 | 222 | 5 | 222 | 9 | 222 | 261 |

# Appendix C

# Derivatives of $p_\nu$ with respect to the model parameters[1]

The numbering of the models is identical to that used in Figure 3.3 in Chapter 3.

## C.1 The Nominal Response model

**Response probabilities:**

$$p_{jg|\theta_t} = \frac{\exp[a_{jg} \cdot \theta_t + c_{jg}]}{\sum_{h=0}^{m_j} \exp[a_{jh} \cdot \theta_t + c_{jh}]}$$

**First derivatives:**

$$\begin{aligned}
\frac{\partial p_\nu}{\partial a_{jg}} &= \sum_t p_{\theta_t} \cdot p_{\nu|\theta_t} \cdot \theta_t (x_{\nu jg} - p_{jg|\theta_t}) \\
\frac{\partial p_\nu}{\partial c_{jg}} &= \sum_t p_{\theta_t} \cdot p_{\nu|\theta_t} (x_{\nu jg} - p_{jg|\theta_t})
\end{aligned}$$

## C.2 The Nominal Response model with equality constraints on the discrimination parameters

**Response probabilities:**

$$p_{jg|\theta_t} = \frac{\exp[a_g \cdot \theta_t + c_{jg}]}{\sum_{h=0}^{m} \exp[a_h \cdot \theta_t + c_{jh}]}$$

[1]These results are necessary for calculating the expected information matrix; see Appendix A for further information.

**First derivatives:**

$$\begin{aligned}\frac{\partial p_\nu}{\partial a_g} &= \sum_t \sum_j p_{\theta_t} \cdot p_{\nu|\theta_t} \cdot \theta_t (x_{\nu jg} - p_{jg|\theta_t}) \\ \frac{\partial p_\nu}{\partial c_{jg}} &= \sum_t p_{\theta_t} \cdot p_{\nu|\theta_t} \cdot (x_{\nu jg} - p_{jg|\theta_t})\end{aligned}$$

## C.3 The Partial Credit model with item-specific scaling parameters

**Response probabilities:**

$$p_{jg|\theta_t} = \frac{\exp[\alpha_j \cdot g \cdot \theta_t + c_{jg}]}{\sum_h \exp[\alpha_j \cdot h \cdot \theta_t + c_{jh}]}$$

**First derivatives:**

$$\begin{aligned}\frac{\partial p_\nu}{\partial \alpha_j} &= \sum_t \sum_g p_{\theta_t} \cdot p_{\nu|\theta_t} \cdot g \cdot \theta_t (x_{\nu jg} - p_{jg|\theta_t}) \\ \frac{\partial p_\nu}{\partial c_{jg}} &= \sum_t p_{\theta_t} \cdot p_{\nu|\theta_t} \cdot (x_{\nu jg} - p_{jg|\theta_t})\end{aligned}$$

## C.4 The Partial Credit model without item-specific scaling parameters

**Response probabilities:**

$$p_{jg|\theta_t} = \frac{\exp[\alpha \cdot g \cdot \theta_t + c_{jg}]}{\sum_h \exp[\alpha \cdot h \cdot \theta_t + c_{jh}]}$$

**First derivatives:**

$$\begin{aligned}\frac{\partial p_\nu}{\partial \alpha} &= \sum_t \sum_j \sum_g p_{\theta_t} \cdot p_{\nu|\theta_t} \cdot g \cdot \theta_t (x_{\nu jg} - p_{jg|\theta_t}) \\ \frac{\partial p_\nu}{\partial c_{jg}} &= \sum_t p_{\theta_t} \cdot p_{\nu|\theta_t} \cdot (x_{\nu jg} - p_{jg|\theta_t})\end{aligned}$$

## C.5 The Unidimensional Polychotomous Rasch model

**Response probabilities:**

$$p_{jg|\theta_t} = \frac{\exp[a_g \cdot (\theta_t - \delta_j) + c_g]}{\sum_{h=0}^{m} \exp[a_h \cdot (\theta_t - \delta_j) + c_h]}$$

**First derivatives:**

$$\begin{aligned}
\frac{\partial p_\nu}{\partial a_g} &= \sum_t \sum_j p_{\theta_t} \cdot p_{\nu|\theta_t} \cdot (\theta_t - \delta_j) \cdot (x_{\nu jg} - p_{jg|\theta_t}) \\
\frac{\partial p_\nu}{\partial c_g} &= \sum_t \sum_j p_{\theta_t} \cdot p_{\nu|\theta_t} \cdot (x_{\nu jg} - p_{jg|\theta_t}) \\
\frac{\partial p_\nu}{\partial \delta_j} &= -\sum_t \sum_g p_{\theta_t} \cdot p_{\nu|\theta_t} \cdot a_g (x_{\nu jg} - p_{jg|\theta_t})
\end{aligned}$$

## C.6 The Rating Scale model without item-specific scaling parameters

**Response probabilities:**

$$p_{jg|\theta_t} = \frac{\exp[g \cdot (\alpha \cdot \theta_t - \delta_j) + c_g]}{\sum_{h=0}^{m} \exp[h \cdot (\alpha \cdot \theta_t - \delta_j) + c_h]}$$

**First derivatives:**

$$\begin{aligned}
\frac{\partial p_\nu}{\partial \alpha} &= \sum_t \sum_j \sum_g p_{\theta_t} \cdot p_{\nu|\theta_t} \cdot g \cdot \theta_t (x_{\nu jg} - p_{jg|\theta_t}) \\
\frac{\partial p_\nu}{\partial c_g} &= \sum_t \sum_j p_{\theta_t} \cdot p_{\nu|\theta_t} \cdot (x_{\nu jg} - p_{jg|\theta_t}) \\
\frac{\partial p_\nu}{\partial \delta_j} &= -\sum_t \sum_g p_{\theta_t} \cdot p_{\nu|\theta_t} \cdot g (x_{\nu jg} - p_{jg|\theta_t})
\end{aligned}$$

## C.7 The Rating Scale model with item-specific scaling parameters

**Response probabilities:**

$$p_{jg|\theta_t} = \frac{\exp[g \cdot (\alpha_j \cdot \theta_t - \delta_j) + c_g]}{\sum_{h=0}^{m} \exp[h \cdot (\alpha_j \cdot \theta_t - \delta_j) + c_h]}$$

**First derivatives:**

$$\begin{aligned}
\frac{\partial p_\nu}{\partial \alpha_j} &= \sum_t \sum_g p_{\theta_t} \cdot p_{\nu|\theta_t} \cdot g \cdot \theta_t (x_{\nu jg} - p_{jg|\theta_t}) \\
\frac{\partial p_\nu}{\partial c_g} &= \sum_t \sum_j p_{\theta_t} \cdot p_{\nu|\theta_t} \cdot (x_{\nu jg} - p_{jg|\theta_t}) \\
\frac{\partial p_\nu}{\partial \delta_j} &= -\sum_t \sum_g p_{\theta_t} \cdot p_{\nu|\theta_t} \cdot g (x_{\nu jg} - p_{jg|\theta_t})
\end{aligned}$$

## C.8 The Unidimensional Polychotomous Rasch model with equality constraints on the item difficulties

**Response probabilities:**

$$p_{jg|\theta_t} = \frac{\exp[a_g \cdot \theta_t + c_g]}{\sum_{h=0}^{m} \exp[a_h \cdot \theta_t + c_h]}$$

**First derivatives:**

$$\begin{aligned}
\frac{\partial p_\nu}{\partial a_g} &= \sum_t \sum_j p_{\theta_t} \cdot p_{\nu|\theta_t} \cdot \theta_t \cdot (x_{\nu jg} - p_{jg|\theta_t}) \\
\frac{\partial p_\nu}{\partial c_g} &= \sum_t \sum_j p_{\theta_t} \cdot p_{\nu|\theta_t} \cdot (x_{\nu jg} - p_{jg|\theta_t})
\end{aligned}$$

## C.9 The Rating Scale model without item-specific scaling parameters and with equality constraints on the item difficulties

**Response probabilities:**

$$p_{jg|\theta_t} = \frac{\exp(\alpha \cdot g \cdot \theta_t + c_g)}{\sum_{h=0}^{m} \exp(\alpha \cdot h \cdot \theta_t + c_h)}$$

**First derivatives:**

$$\begin{aligned}
\frac{\partial p_\nu}{\partial \alpha} &= \sum_t \sum_j \sum_g p_{\theta_t} \cdot p_{\nu|\theta_t} \cdot g \cdot \theta_t (x_{\nu jg} - p_{jg|\theta_t}) \\
\frac{\partial p_\nu}{\partial c_g} &= \sum_t \sum_j p_{\theta_t} \cdot p_{\nu|\theta_t} \cdot (x_{\nu jg} - p_{jg|\theta_t})
\end{aligned}$$

## C.10 The Rating Scale model with item-specific scaling parameters and equality constraints on the item difficulties

**Response probabilities:**

$$p_{jg|\theta_t} = \frac{\exp(\alpha_j \cdot g \cdot \theta_t + c_g)}{\sum_{h=0}^{m} \exp(\alpha_j \cdot h \cdot \theta_t + c_h)}$$

**First derivatives:**

$$\begin{aligned}
\frac{\partial p_\nu}{\partial \alpha_j} &= \sum_t \sum_g p_{\theta_t} \cdot p_{\nu|\theta_t} \cdot g \cdot \theta_t (x_{\nu jg} - p_{jg|\theta_t}) \\
\frac{\partial p_\nu}{\partial c_g} &= \sum_t \sum_j p_{\theta_t} \cdot p_{\nu|\theta_t} \cdot (x_{\nu jg} - p_{jg|\theta_t})
\end{aligned}$$

## C.11 The Rating Scale model without item-specific scaling parameters and with equality constraints on the category parameters

**Response probabilities:**

$$p_{jg|\theta_t} = \frac{\exp[\alpha \cdot g \cdot \theta_t - g \cdot \delta_j]}{\sum_{h=0}^{m} \exp[\alpha \cdot h \cdot \theta_t - h \cdot \delta_j]}$$

**First derivatives:**

$$\begin{aligned}
\frac{\partial p_\nu}{\partial \alpha} &= \sum_t \sum_j \sum_g p_{\theta_t} \cdot p_{\nu|\theta_t} \cdot g \cdot \theta_t (x_{\nu jg} - p_{jg|\theta_t}) \\
\frac{\partial p_\nu}{\partial \delta_j} &= -\sum_t \sum_g p_{\theta_t} \cdot p_{\nu|\theta_t} \cdot g (x_{\nu jg} - p_{jg|\theta_t})
\end{aligned}$$

## C.12 The Rating Scale model with item-specific scaling parameters and equality constraints on the category parameters

**Response probabilities:**

$$p_{jg|\theta_t} = \frac{\exp[\alpha_j \cdot g \cdot \theta_t - g \cdot \delta_j]}{\sum_h \exp[\alpha_j \cdot h \cdot \theta_t - h \cdot \delta_j]}$$

**First derivatives:**

$$\begin{aligned}
\frac{\partial p_\nu}{\partial \alpha_j} &= \sum_t \sum_g p_{\theta_t} \cdot p_{\nu|\theta_t} \cdot g \cdot \theta_t (x_{\nu j g} - p_{jg|\theta_t}) \\
\frac{\partial p_\nu}{\partial \delta_j} &= -\sum_t \sum_g p_{\theta_t} \cdot p_{\nu|\theta_t} \cdot g (x_{\nu j g} - p_{jg|\theta_t})
\end{aligned}$$

# Appendix D

# First and second derivatives of the log-likelihood with respect to the item parameters for a number of discretized latent trait models

In this appendix, the first and second derivatives of the log-likelihood with respect to the item parameters are given. The part of the log-likelihood that needs to be maximized in order to obtain ML estimates for the item parameters can be expressed as (see Section 4.2.3.1)

$$\ln L = \sum_t \sum_j \sum_g q_{jg\theta_t} \cdot \ln p_{jg|\theta_t}.$$

The expression for $p_{jg|\theta_t}$ depends on the latent trait model specified. The symbol $q_{jg\theta_t}$ denotes the number of individuals belonging to latent "class" $\theta_t$ who respond in category $g$ of item $j$. Furthermore, the number of individuals belonging to latent "class" $\theta_t$ is denoted by $\theta_t$. The numbering of the models is the same as that used in Figure 3.3 in Chapter 3.

## D.1 The Nominal Response model

**Response probabilities:**

$$p_{jg|\theta_t} = \frac{\exp[a_{jg} \cdot \theta_t + c_{jg}]}{\sum_{h=0}^{m_j} \exp[a_{jh} \cdot \theta_t + c_{jh}]}$$

**First derivatives:**

$$\begin{aligned}\frac{\partial \ln L}{\partial a_{jg}} &= \sum_t [q_{jg\theta_t} - e_{+t} \cdot p_{jg|\theta_t}] \cdot \theta_t \\ \frac{\partial \ln L}{\partial c_{jg}} &= \sum_t [q_{jg\theta_t} - e_{+t} \cdot p_{jg|\theta_t}]\end{aligned}$$

**Second derivatives:**

$$\begin{aligned}\frac{\partial^2 \ln L}{\partial a_{jg}^2} &= -\sum_t e_{+t} \cdot p_{jg|\theta_t} \cdot (1 - p_{jg|\theta_t}) \cdot \theta_t^2 \\ \frac{\partial^2 \ln L}{\partial c_{jg}^2} &= -\sum_t e_{+t} \cdot p_{jg|\theta_t} \cdot (1 - p_{jg|\theta_t}) \\ \frac{\partial^2 \ln L}{\partial a_{jg} \partial a_{jh}} &= \sum_t e_{+t} \cdot p_{jg|\theta_t} \cdot p_{jh|\theta_t} \cdot \theta_t^2 \\ \frac{\partial^2 \ln L}{\partial c_{jg} \partial c_{jh}} &= \sum_t e_{+t} \cdot p_{jg|\theta_t} \cdot p_{jh|\theta_t} \\ \frac{\partial^2 \ln L}{\partial a_{jg} \partial c_{jg}} &= -\sum_t e_{+t} \cdot p_{jg|\theta_t} \cdot (1 - p_{jg|\theta_t}) \cdot \theta_t \\ \frac{\partial^2 \ln L}{\partial a_{jg} \partial c_{jh}} &= \sum_t e_{+t} \cdot p_{jg|\theta_t} \cdot p_{jh|\theta_t} \cdot \theta_t\end{aligned}$$

## D.2 The Nominal Response model with equality constraints on the discrimination parameters

**Response probabilities:**

$$p_{jg|\theta_t} = \frac{\exp[a_g \cdot \theta_t + c_{jg}]}{\sum_{h=0}^{m} \exp[a_h \cdot \theta_t + c_{jh}]}$$

**First derivatives:**

$$\begin{aligned}\frac{\partial \ln L}{\partial a_g} &= \sum_t \sum_j [q_{jg\theta_t} - e_{+t} \cdot p_{jg|\theta_t}] \cdot \theta_t \\ \frac{\partial \ln L}{\partial c_{jg}} &= \sum_t [q_{jg\theta_t} - e_{+t} \cdot p_{jg|\theta_t}]\end{aligned}$$

**Second derivatives:**

$$
\begin{aligned}
\frac{\partial^2 \ln L}{\partial a_g^2} &= -\sum_t \sum_j e_{+t} \cdot p_{jg|\theta_t} \cdot (1 - p_{jg|\theta_t}) \cdot \theta_t^2 \\
\frac{\partial^2 \ln L}{\partial a_g \partial a_h} &= \sum_t \sum_j e_{+t} \cdot p_{jg|\theta_t} \cdot p_{jh|\theta_t} \cdot \theta_t^2 \\
\frac{\partial^2 \ln L}{\partial a_g \partial c_{jg}} &= -\sum_t e_{+t} \cdot p_{jg|\theta_t} \cdot (1 - p_{jg|\theta_t}) \cdot \theta_t \\
\frac{\partial^2 \ln L}{\partial a_g \partial c_{jh}} &= \sum_t e_{+t} \cdot p_{jg|\theta_t} \cdot p_{jh|\theta_t} \cdot \theta_t \\
\frac{\partial^2 \ln L}{\partial c_{jg}^2} &= -\sum_t e_{+t} \cdot p_{jg|\theta_t} \cdot (1 - p_{jg|\theta_t}) \\
\frac{\partial^2 \ln L}{\partial c_{jg} \partial c_{jh}} &= \sum_t e_{+t} \cdot p_{jg|\theta_t} \cdot p_{jh|\theta_t}
\end{aligned}
$$

## D.3 The Partial Credit model with item-specific scaling parameters

**Response probabilities:**

$$p_{jg|\theta_t} = \frac{\exp[\alpha_j \cdot g \cdot \theta_t + c_{jg}]}{\sum_h \exp[\alpha_j \cdot h \cdot \theta_t + c_{jh}]}$$

**First derivatives:**

$$
\begin{aligned}
\frac{\partial \ln L}{\partial \alpha_j} &= \sum_t \sum_g \left[q_{jg\theta_t} - e_{+t} \cdot p_{jg|\theta_t}\right] g \cdot \theta_t \\
\frac{\partial \ln L}{\partial c_{jg}} &= \sum_t [q_{jg\theta_t} - e_{+t} \cdot p_{jg|\theta_t}]
\end{aligned}
$$

**Second derivatives:**

$$
\begin{aligned}
\frac{\partial^2 \ln L}{\partial \alpha_j^2} &= -\sum_t \left\{ e_{+t} \cdot \theta_t^2 \left[ \sum_g g^2 \cdot p_{jg|\theta_t} - \left( \sum_g g \cdot p_{jg|\theta_t} \right)^2 \right] \right\} \\
\frac{\partial^2 \ln L}{\partial \alpha_j \partial c_{jg}} &= -\sum_t e_{+t} \left[ g \cdot \theta_t \cdot p_{jg|\theta_t} - p_{jg|\theta_t} \left( \sum_g g \cdot \theta_t \cdot p_{jg|\theta_t} \right) \right]
\end{aligned}
$$

$$\frac{\partial^2 \ln L}{\partial c_{jg}^2} = -\sum_t e_{+t} \cdot p_{jg|\theta_t} \cdot (1 - p_{jg|\theta_t})$$
$$\frac{\partial^2 \ln L}{\partial c_{jg} \partial c_{jh}} = \sum_t e_{+t} \cdot p_{jg|\theta_t} \cdot p_{jh|\theta_t}$$

## D.4 The Partial Credit model without item-specific scaling parameters

**Response probabilities:**

$$p_{jg|\theta_t} = \frac{\exp[\alpha \cdot g \cdot \theta_t + c_{jg}]}{\sum_h \exp[\alpha \cdot h \cdot \theta_t + c_{jh}]}$$

**First derivatives:**

$$\frac{\partial \ln L}{\partial \alpha} = \sum_t \sum_j \sum_g \left[q_{jg\theta_t} - e_{+t} \cdot p_{jg|\theta_t}\right] \cdot g \cdot \theta_t$$
$$\frac{\partial \ln L}{\partial c_{jg}} = \sum_t [q_{jg\theta_t} - e_{+t} \cdot p_{jg|\theta_t}]$$

**Second derivatives:**

$$\frac{\partial^2 \ln L}{\partial \alpha^2} = -\sum_t \sum_j \left\{ e_{+t} \cdot \theta_t^2 \left[ \sum_g g^2 \cdot p_{jg|\theta_t} - \left( \sum_g g \cdot p_{jg|\theta_t} \right)^2 \right] \right\}$$
$$\frac{\partial^2 \ln L}{\partial \alpha \partial c_{jg}} = -\sum_t e_{+t} \left[ g \cdot \theta_t \cdot p_{jg|\theta_t} - p_{jg|\theta_t} \left( \sum_g g \cdot \theta_t \cdot p_{jg|\theta_t} \right) \right]$$
$$\frac{\partial^2 \ln L}{\partial c_{jg}^2} = -\sum_t e_{+t} \cdot p_{jg|\theta_t} \cdot (1 - p_{jg|\theta_t})$$
$$\frac{\partial^2 \ln L}{\partial c_{jg} \partial c_{jh}} = +\sum_t e_{+t} \cdot p_{jg|\theta_t} \cdot p_{jh|\theta_t}$$

## D.5 The Unidimensional Polychotomous Rasch model

**Response probabilities:**

$$p_{jg|\theta_t} = \frac{\exp[a_g \cdot (\theta_t - \delta_j) + c_g]}{\sum_{h=0}^{m} \exp[a_h \cdot (\theta_t - \delta_j) + c_h]}$$

**First derivatives:**

$$\frac{\partial \ln L}{\partial \delta_j} = -\sum_t \sum_g \left[q_{jg\theta_t} - e_{+t} \cdot p_{jg|\theta_t}\right] a_g$$

$$\frac{\partial \ln L}{\partial a_g} = \sum_t \sum_j \left[q_{jg\theta_t} - e_{+t} \cdot p_{jg|\theta_t}\right] (\theta_t - \delta_j)$$

$$\frac{\partial \ln L}{\partial c_g} = \sum_t \sum_j \left[q_{jg\theta_t} - e_{+t} \cdot p_{jg|\theta_t}\right]$$

**Second derivatives:**

$$\frac{\partial^2 \ln L}{\partial a_g^2} = -\sum_t \sum_j \left\{\left[(\theta_t - \delta_j)^2 \cdot p_{jg|\theta_t} \cdot (1 - p_{jg|\theta_t})\right] \cdot e_{+t}\right\}$$

$$\frac{\partial^2 \ln L}{\partial \delta_j^2} = -\sum_t \left[\sum_g a_g^2 \cdot p_{jg|\theta_t} - \left(\sum_g a_g \cdot p_{jg|\theta_t}\right)^2\right] \cdot e_{+t}$$

$$\frac{\partial^2 \ln L}{\partial c_g^2} = -\sum_t \sum_j \left[e_{+t} \cdot p_{jg|\theta_t} \cdot (1 - p_{jg|\theta_t})\right]$$

$$\frac{\partial^2 \ln L}{\partial a_g \partial a_h} = \sum_t \sum_j \left[e_{+t} \cdot (\theta_t - \delta_j)^2 \cdot p_{jg|\theta_t} \cdot p_{jh|\theta_t}\right]$$

$$\frac{\partial^2 \ln L}{\partial c_g \partial c_h} = \sum_t \sum_j \left[e_{+t} \cdot p_{jg|\theta_t} \cdot p_{jh|\theta_t}\right]$$

$$\frac{\partial^2 \ln L}{\partial a_g \partial c_g} = -\sum_t \sum_j \left[e_{+t} \cdot (\theta_t - \delta_j) \cdot p_{jg|\theta_t} \cdot (1 - p_{jg|\theta_t})\right]$$

$$\frac{\partial^2 \ln L}{\partial a_g \partial c_h} = \sum_t \sum_j \left[e_{+t} \cdot (\theta_t - \delta_j) \cdot p_{jg|\theta_t} \cdot p_{jh|\theta_t}\right]$$

$$\begin{aligned}\frac{\partial^2 \ln L}{\partial a_g \partial \delta_j} = \sum_t \Big[ & e_{+t} \cdot (\theta_t - \delta_j \cdot p_{jg|\theta_t} \cdot \\ & \left\{a_g - \left(\sum_g a_g \cdot p_{jg|\theta_t}\right)\right\} + e_{+t} \cdot p_{jg|\theta_t} - q_{jg\theta_t} \Big]\end{aligned}$$

$$\frac{\partial^2 \ln L}{\partial c_g \partial \delta_j} = \sum_t e_{+t} \cdot p_{jg|\theta_t} \left[a_g - \left(\sum_g a_g \cdot p_{jg|\theta_t}\right)\right]$$

## D.6 The Rating Scale model without item-specific scaling parameters

**Response probabilities:**

$$p_{jg|\theta_t} = \frac{\exp[g \cdot (\alpha \cdot \theta_t - \delta_j) + c_g]}{\sum_{h=0}^{m} \exp[h \cdot (\alpha \cdot \theta_t - \delta_j) + c_h]}$$

**First derivatives:**

$$\begin{aligned}
\frac{\partial \ln L}{\partial \delta_j} &= -\sum_t \sum_g \left[q_{jg\theta_t} - e_{+t} \cdot p_{jg|\theta_t}\right] g \\
\frac{\partial \ln L}{\partial \alpha} &= \sum_t \sum_j \sum_g \left[q_{jg\theta_t} - e_{+t} \cdot p_{jg|\theta_t}\right] \cdot g \cdot \theta_t \\
\frac{\partial \ln L}{\partial c_g} &= \sum_t \sum_j \left[q_{jg\theta_t} - e_{+t} \cdot p_{jg|\theta_t}\right]
\end{aligned}$$

**Second derivatives:**

$$\begin{aligned}
\frac{\partial^2 \ln L}{\partial \alpha^2} &= -\sum_t \sum_j \left\{ e_{+t} \cdot \theta_t^2 \left[ \sum_g g^2 \cdot p_{jg|\theta_t} - \left( \sum_g g \cdot p_{jg|\theta_t} \right)^2 \right] \right\} \\
\frac{\partial^2 \ln L}{\partial \delta_j^2} &= -\sum_t \left[ \sum_g g^2 \cdot p_{jg|\theta_t} - \left( \sum_g g \cdot p_{jg|\theta_t} \right)^2 \right] \cdot e_{+t} \\
\frac{\partial^2 \ln L}{\partial c_g^2} &= -\sum_t \sum_j \left[ e_{+t} \cdot p_{jg|\theta_t} \cdot (1 - p_{jg|\theta_t}) \right] \\
\frac{\partial^2 \ln L}{\partial c_g \partial \delta_j} &= \sum_t e_{+t} \cdot p_{jg|\theta_t} \left[ g - \sum_g g \cdot p_{jg|\theta_t} \right] \\
\frac{\partial^2 \ln L}{\partial c_g \partial c_h} &= \sum_t \sum_j \left[ e_{+t} \cdot p_{jg|\theta_t} \cdot p_{jh|\theta_t} \right] \\
\frac{\partial^2 \ln L}{\partial \alpha \partial \delta_j} &= \sum_t \theta_t \cdot e_{+t} \left[ \sum_g g^2 \cdot p_{jg|\theta_t} - \left( \sum_g g \cdot p_{jg|\theta_t} \right)^2 \right] \\
\frac{\partial^2 \ln L}{\partial \alpha \partial c_g} &= -\sum_t e_{+t} \cdot \theta_t \sum_j \left[ g \cdot p_{jg|\theta_t} - p_{jg|\theta_t} \sum_h h \cdot p_{jh|\theta_t} \right]
\end{aligned}$$

## D.7 The Rating Scale model with item-specific scaling parameters

**Response probabilities:**

$$p_{jg|\theta_t} = \frac{\exp[g \cdot (\alpha_j \cdot \theta_t - \delta_j) + c_g]}{\sum_{h=0}^{m} \exp[h \cdot (\alpha_j \cdot \theta_t - \delta_j) + c_h]}$$

**First derivatives:**

$$\begin{aligned}
\frac{\partial \ln L}{\partial \delta_j} &= -\sum_t \sum_g \left[q_{jg\theta_t} - e_{+t} \cdot p_{jg|\theta_t}\right] g \\
\frac{\partial \ln L}{\partial \alpha_j} &= \sum_t \sum_g \left[q_{jg\theta_t} - e_{+t} \cdot p_{jg|\theta_t}\right] \cdot g \cdot \theta_t \\
\frac{\partial \ln L}{\partial c_g} &= \sum_t \sum_j \left[q_{jg\theta_t} - e_{+t} \cdot p_{jg|\theta_t}\right]
\end{aligned}$$

**Second derivatives:**

$$\begin{aligned}
\frac{\partial^2 \ln L}{\partial \alpha_j^2} &= -\sum_t \left\{ e_{+t} \cdot \theta_t^2 \left[ \sum_g g^2 \cdot p_{jg|\theta_t} - \left( \sum_g g \cdot p_{jg|\theta_t} \right)^2 \right] \right\} \\
\frac{\partial^2 \ln L}{\partial \delta_j^2} &= -\sum_t \left[ \sum_g g^2 \cdot p_{jg|\theta_t} - \left( \sum_g g \cdot p_{jg|\theta_t} \right)^2 \right] \cdot e_{+t} \\
\frac{\partial^2 \ln L}{\partial c_g^2} &= -\sum_t \sum_j \left[ e_{+t} \cdot p_{jg|\theta_t} \cdot (1 - p_{jg|\theta_t}) \right] \\
\frac{\partial^2 \ln L}{\partial c_g \partial \delta_j} &= \sum_t e_{+t} \cdot p_{jg|\theta_t} \left[ g - \sum_g g \cdot p_{jg|\theta_t} \right] \\
\frac{\partial^2 \ln L}{\partial c_g \partial c_h} &= \sum_t \sum_j \left[ e_{+t} \cdot p_{jg|\theta_t} \cdot p_{jh|\theta_t} \right] \\
\frac{\partial^2 \ln L}{\partial \alpha_j \partial \delta_j} &= \sum_t \theta_t \cdot e_{+t} \left[ \sum_g g^2 \cdot p_{jg|\theta_t} - \left( \sum_g g \cdot p_{jg|\theta_t} \right)^2 \right] \\
\frac{\partial^2 \ln L}{\partial \alpha_j \partial c_g} &= -\sum_t e_{+t} \cdot \theta_t \left[ g \cdot p_{jg|\theta_t} - p_{jg|\theta_t} \sum_h h \cdot p_{jh|\theta_t} \right]
\end{aligned}$$

## D.8 The Unidimensional Polychotomous Rasch model with equality constraints on the item difficulties

**Response probabilities:**

$$p_{jg|\theta_t} = \frac{\exp[a_g \cdot \theta_t + c_g]}{\sum_{h=0}^{m} \exp[a_h \cdot \theta_t + c_h]}$$

**First derivatives:**

$$\frac{\partial \ln L}{\partial a_g} = \sum_t \sum_j \left[q_{jg\theta_t} - e_{+t} \cdot p_{jg|\theta_t}\right] \theta_t$$

$$\frac{\partial \ln L}{\partial c_g} = \sum_t \sum_j \left[q_{jg\theta_t} - e_{+t} \cdot p_{jg|\theta_t}\right]$$

**Second derivatives:**

$$\frac{\partial^2 \ln L}{\partial a_g^2} = -\sum_t \sum_j \left[\theta_t^2 \cdot p_{jg|\theta_t} \cdot (1 - p_{jg|\theta_t}) \cdot e_{+t}\right]$$

$$\frac{\partial^2 \ln L}{\partial c_g^2} = -\sum_t \sum_j \left[e_{+t} \cdot p_{jg|\theta_t} \cdot (1 - p_{jg|\theta_t})\right]$$

$$\frac{\partial^2 \ln L}{\partial a_g \partial a_h} = \sum_t \sum_j \left[e_{+t} \cdot \theta_t^2 \cdot p_{jg|\theta_t} \cdot p_{jh|\theta_t}\right]$$

$$\frac{\partial^2 \ln L}{\partial c_g \partial c_h} = \sum_t \sum_j \left[e_{+t} \cdot p_{jg|\theta_t} \cdot p_{jh|\theta_t}\right]$$

$$\frac{\partial^2 \ln L}{\partial a_g \partial c_g} = -\sum_t \sum_j \left[e_{+t} \cdot \theta_t \cdot p_{jg|\theta_t} \cdot (1 - p_{jg|\theta_t})\right]$$

$$\frac{\partial^2 \ln L}{\partial a_g \partial c_h} = \sum_t \sum_j \left[e_{+t} \cdot \theta_t \cdot p_{jg|\theta_t} \cdot p_{jh|\theta_t}\right]$$

## D.9 The Rating Scale model without item-specific scaling parameters and with equality constraints on the item difficulties

**Response probabilities:**

$$p_{jg|\theta_t} = \frac{\exp(\alpha \cdot g \cdot \theta_t + c_g)}{\sum_{h=0}^{m} \exp(\alpha \cdot h \cdot \theta_t + c_h)}$$

**First derivatives:**

$$\frac{\partial \ln L}{\partial \alpha} = \sum_t \sum_j \sum_g \left[q_{jg\theta_t} - e_{+t} \cdot p_{jg|\theta_t}\right] \cdot g \cdot \theta_t$$

$$\frac{\partial \ln L}{\partial c_g} = \sum_t \sum_j \left[q_{jg\theta_t} - e_{+t} \cdot p_{jg|\theta_t}\right]$$

**Second derivatives:**

$$\frac{\partial^2 \ln L}{\partial \alpha^2} = -\sum_t \sum_j \left\{e_{+t} \cdot \theta_t^2 \left[\sum_g g^2 \cdot p_{jg|\theta_t} - \left(\sum_g g \cdot p_{jg|\theta_t}\right)^2\right]\right\}$$

$$\frac{\partial^2 \ln L}{\partial c_g^2} = -\sum_t \sum_j \left[e_{+t} \cdot p_{jg|\theta_t} \cdot (1 - p_{jg|\theta_t})\right]$$

$$\frac{\partial^2 \ln L}{\partial c_g \partial c_h} = \sum_t \sum_j \left[e_{+t} \cdot p_{jg|\theta_t} \cdot p_{jh|\theta_t}\right]$$

$$\frac{\partial^2 \ln L}{\partial \alpha \partial c_g} = -\sum_t e_{+t} \cdot \theta_t \sum_j \left[g \cdot p_{jg|\theta_t} - p_{jg|\theta_t} \sum_h h \cdot p_{jh|\theta_t}\right]$$

## D.10 The Rating Scale model with item-specific scaling parameters and equality constraints on the item difficulties

**Response probabilities:**

$$p_{jg|\theta_t} = \frac{\exp(\alpha_j \cdot g \cdot \theta_t + c_g)}{\sum_{h=0}^{m} \exp(\alpha_j \cdot h \cdot \theta_t + c_h)}$$

**First derivatives:**

$$\frac{\partial \ln L}{\partial \alpha_j} = \sum_t \sum_g \left[q_{jg\theta_t} - e_{+t} \cdot p_{jg|\theta_t}\right] \cdot g \cdot \theta_t$$

$$\frac{\partial \ln L}{\partial c_g} = \sum_t \sum_j \left[q_{jg\theta_t} - e_{+t} \cdot p_{jg|\theta_t}\right]$$

**Second derivatives:**

$$\frac{\partial^2 \ln L}{\partial \alpha_j^2} = -\sum_t \left\{e_{+t} \cdot \theta_t^2 \left[\sum_g g^2 \cdot p_{jg|\theta_t} - \left(\sum_g g \cdot p_{jg|\theta_t}\right)^2\right]\right\}$$

$$\frac{\partial^2 \ln L}{\partial c_g^2} = -\sum_t \sum_j \left[e_{+t} \cdot p_{jg|\theta_t} \cdot (1 - p_{jg|\theta_t})\right]$$

$$\frac{\partial^2 \ln L}{\partial c_g \partial c_h} = \sum_t \sum_j \left[e_{+t} \cdot p_{jg|\theta_t} \cdot p_{jh|\theta_t}\right]$$

$$\frac{\partial^2 \ln L}{\partial \alpha_j \partial c_g} = -\sum_t e_{+t} \cdot \theta_t \left[g \cdot p_{jg|\theta_t} - p_{jg|\theta_t} \sum_h h \cdot p_{jh|\theta_t}\right]$$

## D.11 The Rating Scale model without item-specific scaling parameters and with equality constraints on the category parameters

**Response probabilities:**

$$p_{jg|\theta_t} = \frac{\exp[\alpha \cdot g \cdot \theta_t - g \cdot \delta_j]}{\sum_{h=0}^{m} \exp[\alpha \cdot h \cdot \theta_t - h \cdot \delta_j]}$$

**First derivatives:**

$$\frac{\partial \ln L}{\partial \alpha} = \sum_t \sum_j \sum_g \left[q_{jg\theta_t} - e_{+t} \cdot p_{jg|\theta_t}\right] \cdot g \cdot \theta_t$$

$$\frac{\partial \ln L}{\partial \delta_j} = -\sum_t \sum_g \left[q_{jg\theta_t} - e_{+t} \cdot p_{jg|\theta_t}\right] g$$

**Second derivatives:**

$$\frac{\partial^2 \ln L}{\partial \alpha^2} = -\sum_t \sum_j \left\{ e_{+t} \cdot \theta_t^2 \left[\sum_g g^2 \cdot p_{jg|\theta_t} - \left(\sum_g g \cdot p_{jg|\theta_t}\right)^2\right]\right\}$$

$$\frac{\partial^2 \ln L}{\partial \delta_j^2} = -\sum_t \left[\sum_g g^2 \cdot p_{jg|\theta_t} - \left(\sum_g g \cdot p_{jg|\theta_t}\right)^2\right] \cdot e_{+t}$$

$$\frac{\partial^2 \ln L}{\partial \alpha \partial \delta_j} = \sum_t \theta_t \cdot e_{+t} \left[\sum_g g^2 \cdot p_{jg|\theta_t} - \left(\sum_g g \cdot p_{jg|\theta_t}\right)^2\right]$$

## D.12 The Rating Scale model with item-specific scaling parameters and equality constraints on the category parameters

**Response probabilities:**

$$p_{jg|\theta_t} = \frac{\exp[\alpha_j \cdot g \cdot \theta_t - g \cdot \delta_j]}{\sum_h \exp[\alpha_j \cdot h \cdot \theta_t - h \cdot \delta_j]}$$

**First derivatives:**

$$\frac{\partial \ln L}{\partial \alpha_j} = \sum_t \sum_g \left[q_{jg\theta_t} - e_{+t} \cdot p_{jg|\theta_t}\right] g \cdot \theta_t$$

$$\frac{\partial \ln L}{\partial \delta_j} = -\sum_t \sum_g \left[q_{jg\theta_t} - e_{+t} \cdot p_{jg|\theta_t}\right] \cdot g$$

**Second derivatives:**

$$\frac{\partial^2 \ln L}{\partial \alpha_j^2} = -\sum_t \left\{ e_{+t} \cdot \theta_t^2 \left[ \sum_g g^2 \cdot p_{jg|\theta_t} - \left( \sum_g g \cdot p_{jg|\theta_t} \right)^2 \right] \right\}$$

$$\frac{\partial^2 \ln L}{\partial \delta_j^2} = -\sum_t \left\{ e_{+t} \left[ \sum_g g^2 \cdot p_{jg|\theta_t} - \left( \sum_g g \cdot p_{jg|\theta_t} \right)^2 \right] \right\}$$

$$\frac{\partial^2 \ln L}{\partial \alpha_j \partial \delta_j} = \sum_t \left\{ e_{+t} \cdot \theta_t \left[ \sum_g g^2 \cdot p_{jg|\theta_t} - \left( \sum_g g \cdot p_{jg|\theta_t} \right)^2 \right] \right\}$$

# Appendix E

# First and second derivatives of the log-likelihood with respect to $\theta_i$ for a number of discretized latent trait models

## E.1 The Nominal Response model

**Response probabilities:**

$$p_{jg|\theta_t} = \frac{\exp[a_{jg} \cdot \theta_t + c_{jg}]}{\sum_{h=0}^{m_j} \exp[a_{jh} \cdot \theta_t + c_{jh}]}$$

**First derivative:**

$$\frac{\partial \log L}{\partial \theta_i} = \sum_{j=1}^{n} \sum_{g=0}^{m_j} a_{jg} \left( x_{\nu jg} - p_{jg|\theta_i} \right)$$

**Second derivative:**

$$\frac{\partial^2 \log L}{\partial \theta_i^2} = -\sum_j \left\{ \sum_g a_{jg}^2 \cdot p_{jg|\theta_i} - \left[ \sum_g a_{jg} \cdot p_{jg|\theta_i} \right]^2 \right\}$$

## E.2 The Nominal Response model with equality constraints on the discrimination parameters

**Response probabilities:**

$$p_{jg|\theta_t} = \frac{\exp[a_g \cdot \theta_t + c_{jg}]}{\sum_{h=0}^{m} \exp[a_h \cdot \theta_t + c_{jh}]}$$

**First derivative:**

$$\frac{\partial \log L}{\partial \theta_i} = \sum_j \sum_g a_g \cdot (x_{\nu jg} - p_{jg|\theta_i})$$

**Second derivative:**

$$\frac{\partial^2 \log L}{\partial \theta_i^2} = -\sum_j \left\{ \sum_g a_g^2 \cdot p_{jg|\theta_i} - \left[ \sum_g a_g \cdot p_{jg|\theta_i} \right]^2 \right\}$$

## E.3 The Partial Credit model with item-specific scaling parameters

**Response probabilities:**

$$p_{jg|\theta_t} = \frac{\exp[\alpha_j \cdot g \cdot \theta_t + c_{jg}]}{\sum_h \exp[\alpha_j \cdot h \cdot \theta_t + c_{jh}]}$$

**First derivative:**

$$\frac{\partial \log L}{\partial \theta_i} = \sum_j \sum_g \alpha_j \cdot g \cdot (x_{\nu jg} - p_{jg|\theta_i})$$

**Second derivative:**

$$\frac{\partial^2 \log L}{\partial \theta_i^2} = -\sum_j \left\{ \sum_g (\alpha_j \cdot g)^2 \cdot p_{jg|\theta_i} - \left[ \sum_g \alpha_j \cdot g \cdot p_{jg|\theta_i} \right]^2 \right\}$$

## E.4 The Partial Credit model without item-specific scaling parameters

**Response probabilities:**

$$p_{jg|\theta_t} = \frac{\exp[\alpha \cdot g \cdot \theta_t + c_{jg}]}{\sum_h \exp[\alpha \cdot h \cdot \theta_t + c_{jh}]}$$

**First derivative:**

$$\frac{\partial \log L}{\partial \theta_i} = \sum_j \sum_g \alpha \cdot g \cdot (x_{\nu jg} - p_{jg|\theta_i})$$

**Second derivatives:**

$$\frac{\partial^2 \log L}{\partial \theta_i^2} = -\sum_j \left\{ \sum_g (\alpha \cdot g)^2 \cdot p_{jg|\theta_i} - \left[ \sum_g \alpha \cdot g \cdot p_{jg|\theta_i} \right]^2 \right\}$$

## E.5 The Unidimensional Polychotomous Rasch model

**Response probabilities:**

$$p_{jg|\theta_t} = \frac{\exp[a_g \cdot (\theta_t - \delta_j) + c_g]}{\sum_{h=0}^{m} \exp[a_h \cdot (\theta_t - \delta_j) + c_h]}$$

**First derivative:**

$$\frac{\partial \log L}{\partial \theta_i} = \sum_j \sum_g a_g \cdot (x_{\nu jg} - p_{jg|\theta_i})$$

**Second derivative:**

$$\frac{\partial^2 \log L}{\partial \theta_i^2} = -\sum_j \left\{ \sum_g a_g^2 \cdot p_{jg|\theta_i} - \left[ \sum_g a_g \cdot p_{jg|\theta_i} \right]^2 \right\}$$

## E.6 The Rating Scale model without item-specific scaling parameters

**Response probabilities:**

$$p_{jg|\theta_t} = \frac{\exp[g \cdot (\alpha \cdot \theta_t - \delta_j) + c_g]}{\sum_{h=0}^{m} \exp[h \cdot (\alpha \cdot \theta_t - \delta_j) + c_h]}$$

**First derivative:**

$$\frac{\partial \log L}{\partial \theta_i} = \sum_j \sum_g \alpha \cdot g \cdot (x_{\nu jg} - p_{jg|\theta_i})$$

**Second derivative:**

$$\frac{\partial^2 \log L}{\partial \theta_i^2} = -\sum_j \left\{ \sum_g (\alpha \cdot g)^2 \cdot p_{jg|\theta_i} - \left[ \sum_g \alpha \cdot g \cdot p_{jg|\theta_i} \right]^2 \right\}$$

## E.7 The Rating Scale model with item-specific scaling parameters

**Response probabilities:**

$$p_{jg|\theta_t} = \frac{\exp[g \cdot (\alpha_j \cdot \theta_t - \delta_j) + c_g]}{\sum_{h=0}^{m} \exp[h \cdot (\alpha_j \cdot \theta_t - \delta_j) + c_h]}$$

**First derivative:**

$$\frac{\partial \log L}{\partial \theta_i} = \sum_j \sum_g \alpha_j \cdot g \cdot (x_{\nu jg} - p_{jg|\theta_i})$$

**Second derivative:**

$$\frac{\partial^2 \log L}{\partial \theta_i^2} = -\sum_j \left\{ \sum_g (\alpha_j \cdot g)^2 \cdot p_{jg|\theta_i} - \left[ \sum_g \alpha_j \cdot g \cdot p_{jg|\theta_i} \right]^2 \right\}$$

## E.8 The Unidimensional Polychotomous Rasch model with equality constraints on the item difficulties

**Response probabilities:**

$$p_{jg|\theta_t} = \frac{\exp[a_g \cdot \theta_t + c_g]}{\sum_{h=0}^{m} \exp[a_h \cdot \theta_t + c_h]}$$

**First derivative:**

$$\frac{\partial \log L}{\partial \theta_i} = \sum_j \sum_g a_g \cdot (x_{\nu jg} - p_{jg|\theta_i})$$

**Second derivative:**

$$\frac{\partial^2 \log L}{\partial \theta_i^2} = -\sum_j \left\{ \sum_g a_g^2 \cdot p_{jg|\theta_i} - \left[ \sum_g a_g \cdot p_{jg|\theta_i} \right]^2 \right\}$$

## E.9 The Rating Scale model without item-specific scaling parameters and with equality constraints on the item difficulties

**Response probabilities:**

$$p_{jg|\theta_t} = \frac{\exp(\alpha \cdot g \cdot \theta_t + c_g)}{\sum_{h=0}^{m} \exp(\alpha \cdot h \cdot \theta_t + c_h)}$$

**First derivative:**

$$\frac{\partial \log L}{\partial \theta_i} = \sum_j \sum_g \alpha \cdot g \cdot (x_{\nu jg} - p_{jg|\theta_i})$$

**Second derivative:**

$$\frac{\partial^2 \log L}{\partial \theta_i^2} = -\sum_j \left\{ \sum_g (\alpha \cdot g)^2 \cdot p_{jg|\theta_i} - \left[ \sum_g \alpha \cdot g \cdot p_{jg|\theta_i} \right]^2 \right\}$$

## E.10 The Rating Scale model with item-specific scaling parameters and equality constraints on the item difficulties

**Response probabilities:**

$$p_{jg|\theta_t} = \frac{\exp(\alpha_j \cdot g \cdot \theta_t + c_g)}{\sum_{h=0}^{m} \exp(\alpha_j \cdot h \cdot \theta_t + c_h)}$$

**First derivative:**

$$\frac{\partial \log L}{\partial \theta_i} = \sum_j \sum_g \alpha_j \cdot g \cdot (x_{\nu jg} - p_{jg|\theta_i})$$

**Second derivative:**

$$\frac{\partial^2 \log L}{\partial \theta_i^2} = -\sum_j \left\{ \sum_g (\alpha_j \cdot g)^2 \cdot p_{jg|\theta_i} - \left[ \sum_g \alpha_j \cdot g \cdot p_{jg|\theta_i} \right]^2 \right\}$$

## E.11 The Rating Scale model without item-specific scaling parameters and with equality constraints on the category parameters

**Response probabilities:**

$$p_{jg|\theta_t} = \frac{\exp[\alpha \cdot g \cdot \theta_t - g \cdot \delta_j]}{\sum_{h=0}^{m} \exp[\alpha \cdot h \cdot \theta_t - h \cdot \delta_j]}$$

**First derivative:**

$$\frac{\partial \log L}{\partial \theta_i} = \sum_j \sum_g \alpha \cdot g \cdot (x_{\nu jg} - p_{jg|\theta_i})$$

**Second derivative:**

$$\frac{\partial^2 \log L}{\partial \theta_i^2} = -\sum_j \left\{ \sum_g (\alpha \cdot g)^2 \cdot p_{jg|\theta_i} - \left[ \sum_g \alpha \cdot g \cdot p_{jg|\theta_i} \right]^2 \right\}$$

## E.12 The Rating Scale model with item-specific scaling parameters and equality constraints on the category parameters

**Response probabilities:**

$$p_{jg|\theta_t} = \frac{\exp[\alpha_j \cdot g \cdot \theta_t - g \cdot \delta_j]}{\sum_h \exp[\alpha_j \cdot h \cdot \theta_t - h \cdot \delta_j]}$$

**First derivative:**

$$\frac{\partial \log L}{\partial \theta_i} = \sum_j \sum_g \alpha_j \cdot g \cdot (x_{\nu jg} - p_{jg|\theta_i})$$

**Second derivative:**

$$\frac{\partial^2 \log L}{\partial \theta_i^2} = -\sum_j \left\{ \sum_g (\alpha_j \cdot g)^2 \cdot p_{jg|\theta_i} - \left[ \sum_g \alpha_j \cdot g \cdot p_{jg|\theta_i} \right]^2 \right\}$$

# Appendix F

# The DILTRAN program

Parameters can be estimated for all discretized latent trait models mentioned in Figure 3.3 and discussed in Chapter 3 using the DILTRAN program. It is also possible to estimate parameters in mixed models. At present, the program is only available for VAX/VMS but the program will be ported to other platforms. It is delivered with a manual and examples.

The program is driven by commands, either directly from the keyboard, indirectly from so-called COMMAND files, or from general BATCH files. The user must define the (mixed) model and the number and values of the latent nodes. The data can be entered as raw data (i.e., a vector with responses for each individual) or as frequency data. Initial estimates for the parameters can be specified by the user or generated randomly. Stop-criteria and the number of iterations are determined by the user. After a run, intermediate results can be displayed and the program can be restarted with or without altered model parameters and/or stop-criteria.

The program produces output concerning the estimated item parameters and latent proportions, as well as a number of test statistics. Optional output is available for estimated expected frequencies (the full table or tables collapsed over one or more variables), estimates for the latent scores (ML estimates, MAP and EAP), the information matrix, the estimated variance/covariance matrix for the estimated parameters, and the eigenvalues and eigenvectors of this matrix.

# Summary

Measurement models in which unobserved latent variables are linked to observed manifest indicators have become increasingly popular in recent decades. Both latent trait models and latent class models are examples of such measurement models. In latent class and latent trait models, it is assumed that the manifest indicators are discrete variables. One of the most frequently mentioned differences between these two types of models is that the latent variables in latent class models are also considered discrete, while in latent trait models the latent variables are assumed to be continuous. This difference is, however, not as absolute as is often suggested. In some cases, the continuous latent variable in a latent trait model can be accurately approximated by a discrete distribution. The resulting latent trait model is equivalent to certain restricted latent trait models. In this study, the relationship between (restricted) latent class models and discretized latent trait models is explored in depth.

A classification of latent variable models that can be used for the analysis of discrete manifest indicators is presented in the first introductory chapter. This classification is based on a distinction between the measurement level of the latent and manifest variables.

The latent class model is discussed extensively in Chapter 2. This model can be parameterized in a number of ways. Goodman (1974a) suggested a parameterization in terms of conditional response probabilities and latent proportions, while Haberman (1979) proposed a parameterization based on the log-linear model. In this book, a combination of these two methods of parameterization is used. The latent class model is parameterized in terms of latent proportions and conditional response probabilities; the conditional response probabilities are written as a function of log-linear parameters. A number of restricted latent class models are presented in Chapter 2. The most interesting of

these are models with linear restrictions on the log-linear interactions describing the relation between the latent and the manifest variables. When these restrictions are used, the latent variable is considered to be a metrical variable. The resulting models can be shown to be equivalent to discretized latent trait models.

Maximum likelihood methods for estimating the parameters in latent class models as well as methods for model selection are also discussed in Chapter 2. All maximum likelihood methods discussed are based on the same likelihood function. The differences between these methods are merely numerical.

In Chapter 3, attention is focused on latent trait models. A number of latent trait models are presented, in which the latent variable is discretized. The most general discretized latent trait model is the Nominal Response model proposed by Bock (1972). All other latent trait models discussed in Chapter 3 can be obtained by imposing specific restrictions on the parameters in the Nominal Response model. If the latent variable is treated as discrete, the latent trait models are identical to a number of restricted latent class models that were introduced in Chapter 2.

The parameters in latent trait models can be estimated using several maximum likelihood procedures. These methods are based on different likelihood functions, so the differences between these methods are not merely numerical. These methods are called *joint maximum likelihood* (JML), *conditional maximum likelihood* (CML), and *marginal maximum likelihood* (MML). In Chapter 4, the focus is on CML and MML. In JML, person parameters and item parameters are estimated simultaneously. When CML is used, the person parameters are eliminated from the likelihood function by conditioning on their sufficient statistics. Thus, this method can be used only in cases in which these sufficient statistics are known. In MML, the person parameters are eliminated in a different fashion. By making certain assumptions about the population distribution of the latent variable, the person parameters can be integrated out of the likelihood function. The strictest method of doing this is to assume that the complete population distribution is specified a priori. It is, however, also possible to make less far reaching assumptions about the population distribution. The unknown continuous latent distribution can be approximated by a discrete distribution.

In some cases, it is assumed that only the values of the latent node-points are known in advance. This is called *semi-parametric MML*. In other cases, no a priori assumptions are made concerning this discrete latent distribution. This method is called *fully semi-parametric MML*. In Chapter 4, the relations between fully-semi-parametric MML and CML are explored in depth. It can be shown that when CML is used, the estimated item parameters are not necessarily compatible with a proper latent population distribution. If subjects are considered as a random sample from a population and, therefore, the concept of a population distribution of the latent variable becomes relevant, estimating the parameters using MML is to be preferred over estimation using CML.

In Chapter 5, two types of extension of discrete latent trait models are discussed. The first deals with multidimensional latent trait models. The other extension tackles the problem of relating the latent trait to external variables. This can be done either by estimating latent scores or by including the external variables in the measurement model in a way that is similar to the strategy used in LISREL models. The log-linear path-analysis approach proposed by Goodman (1973b) and extended by Hagenaars (1988) offers a general framework for including both types of extensions discussed in Chapter 5, i.e., multidimensional latent trait models and latent traits models in which the latent variables are related to external variables.

# Samenvatting

In de laatste decennia zijn meetmodellen waarin niet-geobserveerde latente variabelen worden gekoppeld aan geobserveerde manifeste indicatoren meer en meer populair geworden. Zowel latente trek modellen als latente klassen modellen zijn voorbeelden van dergelijke meetmodellen. In latente klassen en latente trek modellen wordt verondersteld dat de manifeste indicatoren discrete variabelen zijn. Een van de meest frequent genoemde verschillen tussen de beide typen van modellen is, dat in latente klassen modellen ten aanzien van de latente variabelen verondersteld wordt dat deze ook discreet zijn, terwijl in latente trek modellen de latente variabelen als continue worden beschouwd. Echter, dit verschil is niet zo absoluut als soms wel wordt gesuggereerd. In een aantal gevallen kan de continue latente variabele in een latente trek model accuraat benaderd worden door een discrete verdeling. Het resulterende latente trek model is dan identiek aan bepaalde gerestricteerde latente klassen modellen. De relatie tussen (gerestricteerde) latente klassen modellen en gediskretiseerde latente trek modellen wordt in dit boek diepgaand onderzocht.

In het eerste, inleidende hoofdstuk wordt een classificatie gegeven van modellen met latente variabelen die gebruikt kunnen worden voor de analyse van discrete manifeste indicatoren. Deze classificatie is gebaseerd op een onderscheid in het meetniveau van de latente en manifeste variabelen.

In het tweede hoofdstuk wordt het latente klassen model uitvoerig besproken. Dit model kan op een aantal verschillende manieren geparametriseerd worden. Goodman (1974) stelde een parametrisering voor in termen van conditionele antwoordkansen en latente proporties, terwijl Haberman (1979) een parametrisering introduceerde die gebaseerd is op het log-lineaire model. In dit boek wordt een combinatie van deze beide manieren om het latente klassen model te parametriseren gebruikt. De gekozen parametrisering maakt gebruik van latente proporties en conditionele antwoordkansen; deze laatste worden echter geschreven als functie van de log-lineaire parameters. In hoofdstuk 2 worden ook een aantal gerestricteerde latente klassen modellen besproken. Het meest interessant zijn modellen waarin gebruik wordt gemaakt van lineaire restricties op de log-lineaire parameters die de relatie tussen de latente en de manifeste variabelen beschrijven. Wanneer dergelijke restricties worden gebruikt, kan de latent variabele beschouwd worden als een variabele die gemeten is op een metrisch niveau. Dergelijke modellen zijn identiek aan gediscretiseerde latente trek modellen.

Zowel maximum likelihood methoden voor het schatten van de parameters in latente klassen modellen, als methoden voor model selectie worden ook besproken in hoofdstuk 2. De verschillende maximum likelihood procedures zijn allemaal gebaseerd op dezelfde likelihood functie. De verschillen tussen deze procedures zijn dus louter numeriek.

In het derde hoofdstuk staan latente trek modellen centraal. Er wordt een aantal latente trek modellen met een gediskretiseerde latente variabele gepresenteerd. Het meest algemene gediskretiseerde latente trek model is het Nominal Response model, voorgesteld door Bock (1972). Alle andere latente trek modellen die in hoofdstuk 3 worden besproken kunnen worden afgeleid van het Nominal Response model door het aanbrengen van bepaalde restricties. Als de latente variabele als een discrete variabele wordt beschouwd, zijn deze latente trek modellen identiek aan sommige latente klassen modellen die in hoofdstuk 2 werden geïntroduceerd.

De parameters in latente trek modellen kunnen worden geschat met behulp van verschillende maximum likelihood methoden. Deze methoden zijn gebaseerd op verschillende likelihood functies, zodat de verschillen tussen deze methoden niet louter numeriek zijn. Deze methoden zijn resp. *joint maximum likelihood* (JML), *conditional maximum likelihood* (CML), en *marginal maximum likelihood* (MML). Het accent ligt in hoofdstuk 4 op CML en MML. Bij JML worden de persoons-parameters en de item-parameters gelijktijdig geschat. Bij CML worden de persoons-parameters geëlimineerd uit de likelihood functie door te conditioneren op de sufficiënte statistieken voor deze persoons-parameters. Daarom kan deze methode alleen worden gebruikt in situaties waarin deze sufficiënte statistieken bekend zijn. Bij

MML worden de persoons-parameters op een andere wijze uit de likelihood functie geëlimineerd. Door bepaalde veronderstellingen te maken over de populatie verdeling van de latente variabele kunnen de persoons-parameters uit de likelihood functie geïntegreerd worden. De meest strikte methode om dat te doen, is de populatieverdeling op voorhand helemaal vast te leggen. Het is echter ook mogelijk om minder vergaande assumpties te maken met betrekking tot de populatie verdeling. De onbekende continue latente verdeling kan worden benaderd door een discrete verdeling. In sommige gevallen hoeven alleen de waarden van de latente knooppunten op voorhand te worden gespecificeerd. Dit wordt *semi-parametrische MML* genoemd. In andere gevallen hoeven helemaal geen assumpties met betrekking tot deze discrete latente verdeling gemaakt te worden. Dit wordt *volledig semi-parametrische MML* genoemd. De relaties tussen volledig semi-parametrische MML en CML worden in hoofdstuk 4 diepgaand bestudeerd. Het kan worden aangetoond dat, wanneer CML wordt gebruikt, de geschatte item parameters niet noodzakelijkerwijze gecombineerd kunnen worden met een latente populatie verdeling. Als de individuen beschouwd worden als een toevals-steekproef uit een populatie, en aldus het idee van een latente populatie verdeling relevant wordt, heeft het schatten van de parameters met MML daarom de voorkeur boven het gebruik van CML.

In het vijfde hoofdstuk worden twee uitbreidingen van het gediskretiseerde latente trekken model besproken. De eerste heeft betrekking op multidimensionele latente trek modellen. De andere uitbreiding gaat in op de vraag hoe latente trekken gerelateerd kunnen worden aan externe variabelen. Dit kan gebeuren door latente scores te schatten of door de externe variabelen op te nemen in het meetmodel, analoog aan de manier waarop dat in LISREL modellen gebeurt. De log-lineaire padanalyse zoals voorgesteld door Goodman (1973b) en veralgemeniseerd door Hagenaars (1988) biedt een algemeen kader voor het opnemen van de beide soort uitbreidingen die in hoofdstuk 5 worden besproken, nl. multidimensionele latente trek modellen en latente trek modellen waarin de latente variabelen worden gerelateerd aan externe variabelen.

# References


Agresti, A. (1984). *Analysis of categorical data.* New York: John Wiley.

Agresti, A. (1990). *Categorical data analysis.* New York: John Wiley.

Agresti, A. (1992). A survey of exact inference for contingency tables. *Statistical Science, 7*, 131-177.

Aitkin, M., Anderson, D. & Hinde, J. (1981). Statistical modelling of data on teaching styles. *Journal of the Royal Statistical Society. Series A. 144*, 419-461.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control, 19*, 716-723.

Akaike, H. (1987). Factor analysis and AIC. *Psychometrika, 52*, 317-332.

Andersen, E.B. (1970). Asymptotic properties of conditional likelihood estimates. *Journal of the Royal Statistical Society. Series B. 32*, 283-301.

Andersen, E.B. (1973a). Conditional inference for multiple-choice questionnaires. *British Journal of Mathematical and Statistical Psychology, 26*, 31-44.

Andersen, E.B. (1973b). A goodness of fit test for the Rasch model. *Psychometrika, 38*, 123-140.

Andersen, E.B. (1977). Sufficient statistics and latent trait models. *Psychometrika, 42*, 69-81.

Andersen, E.B. (1983). A general latent structure model for contingency table data. In H. Wainer & S. Messick (Eds.), *Principals of modern psychological measurement; a festschrift for Frederic M. Lord* (pp. 415-494) Hillsdale: LEA.

Andersen, E.B. (1988). Comparison of latent structure models. In R. Langeheine & J. Rost (Eds.), *Latent trait and latent class models* (pp. 207-229). New York: Plenum Press.

Andersen, E.B. (1990). *The statistical analysis of categorical data.* Berlin: Springer-Verlag.

Andersen, E.B. & Madsen, M. (1977). Estimating the parameters of the latent population distribution. *Psychometrika, 42*, 357-374.

Andrich, D. (1978a). A rating formulation for ordered response categories. *Psychometrika, 43*, 561-573.

Andrich, D. (1978b). Application of a psychometric rating model to ordered categories which are scored with successive integers. *Applied Psychological Measurement, 2*, 581-594.

Andrich, D. (1978c). A binomial latent trait model for the study of Likert-style attitude questionnaires. *British Journal of Mathematical and Statistical Psychology, 31*, 84-98.

Andrich, D. (1979). A model for contingency tables having an ordered response classification. *Biometrics, 35*, 404-415.

Andrich, D. (1982). An extension of the Rasch model for ratings providing both location and dispersion parameters. *Psychometrika, 47*, 105-113.

Andrich, D. (1985). An elaboration of Guttman scaling with Rasch models. In N.B. Tuma (Ed.), *Sociological Methodology* (pp. 33-80). San Francisco: Jossey-Bass.

Andrich, D. (1992). On the function of fundamental measurement in the social sciences: the objective measurement of subjective meaning. Paper presented as a keynote address on the International Conference on Social Science Methodology. Trento, Italy, June 1992.

Ansley, T.N. & Forsyth, R.A. (1985). An examination of the characteristics of unidimensional IRT parameter estimates derived from two-dimensional data. *Applied Psychological Measurement, 9*, 37-48.

Baker, F.B. (1987). Methodology review: item parameter estimation under the one-, two-, and three-parameter logistic models. *Applied Psychological Measurement, 11*, 111-141.

Bartholomew, D.J. (1987). *Latent variable models and factor analysis.* London: Charles Griffin & Company Ltd.

Bentler, P.M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107*, 238-246.

Bentler, P.M. & Bonnet, D.G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin, 88*, 588-606.

Birch, M.W. (1963). Maximum likelihood in three-way contingency tables. *Journal of the Royal Statistical Society. Series B. 25*, 220-223.

Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F.M. Lord & M.R. Novick, *Statistical theories of mental test scores.* Reading, MA: Addison-Wesley.

Bishop, Y.M.M., Fienberg, S.E. & Holland, P.W. (1975). *Discrete multivariate analysis.* Cambridge, MA: MIT Press.

Bock, R.D. (1972). Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika, 37*, 29-51.

Bock, R.D. & Jones, L. (1968). *The measurement and prediction of judgement and choice.* San Francisco: Holden-Day.

Bock, R.D. & Lieberman, M. (1970). Fitting a response model for $n$ dichotomously scored items. *Psychometrika, 35*, 179-197.

Bock, R.D. & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters : application of an EM algorithm. *Psychometrika, 46*, 443-459.

Bock, R.D., Gibbons, R. & Muraki, E. (1988). Full-information item factor analysis. *Applied Psychological Measurement, 12*, 261-280.

Bollen, K.A. (1989). A new incremental fit index for general structural equation models. *Sociological Methods & Research, 16*, 492-503.

Bollen, K.A. (1990). Overall fit in covariance structure models: two types of sample size effects. *Psychological Bulletin, 107*, 256-259.

Bonnet, D.G. & Bentler, P.M. (1983). Goodness-of-fit procedures for the evaluation and selection of log-linear models. *Psychological Bulletin, 93*, 149-166.

Bozdogan, H. (1987). Model selection and Akaike's information criterion: the general theory and its analytical extension. *Psychometrika, 52*, 345-370.

Clogg, C.C. (1977). Some latent structure models for the analysis of Likert-type data. *Social Science Research, 8*, 287-301.

Clogg, C.C. (1981). New developments in latent structure analysis. In D.J. Jackson & E.F. Borgatta (Eds.), *Factor analysis and measurement in sociological research* (pp. 215-246). Sage Studies in International Sociology 21. Beverly Hills, CA: Sage.

Clogg, C.C. (1982). Some models for the analysis of association in multiway cross-classifications having ordered categories. *Journal of the American Statistical Association, 77*, 803-815.

Clogg, C.C. (1988). Latent class models for measuring. In R. Langeheine & J. Rost (Eds.), *Latent trait and latent class models* (pp. 173-205). New York: Plenum Press.

Clogg, C.C. & Sawyer, D.O. (1981). A comparison of alternative models for analyzing the scalability of response patterns. In S. Leinhardt (Ed.), *Sociological Methodology* (pp. 240-280). San Francisco: Jossey-Bass.

Clogg, C.C. & Goodman, L.A. (1984). Latent structure analysis of a set of multidimensional contingency tables. *Journal of the American Statistical Association, 79*, 762-771.

Clogg, C.C. & Goodman, L.A. (1985). Simultaneous latent structure analysis in several groups. In N.B. Tuma (Ed.), *Sociological Methodology 1985* (pp. 18-110). San Francisco: Jossey-Bass.

Clogg, C.C. & Goodman, L.A. (1986). On scaling models applied to data from several groups. *Psychometrika, 51*, 123-135.

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences.* New York: Academic Press.

Cressie, N. & Holland, P.W. (1983). Characterizing the manifest probabilities of latent trait models. *Psychometrika, 48*, 129-141.

Croon, M. (1989). Latent class models for the analysis of rankings. In G. de Soete, H. Feger & K.C. Klauer (Eds.), *New developments in Psychological Choice Modeling* (pp. 99-121). Amsterdam, The Netherlands: Elsevier.

Croon, M. (1990). Latent class analysis with ordered classes. *British Journal of the Mathematical Statistical Society, 43*, 171-192.

Croon, M. (1991a). The problem of ordered classes: a hierarchy of solutions and a new algorithm. Unpublished paper.

Croon, M. (1991b). Investigating Mokken scalability of dichotomous items by means of ordinal latent class analysis. *British Journal of the Mathematical Statistical Society, 44*, 315-331.

Croon, M. & Luijkx, R. (1992). Latent structure models for ranking data. In M.A. Fligner & J.S. Verducci (Eds.), *Probability models for ranking data* (pp. 53-74). New York: Springer Verlag.

Croon, M. (1993). Ordinal latent class analysis for single-peaked items. *Kwantitatieve Methoden, 14*, 127-142.

Cudeck, R. & Browne, M.W. (1983). Cross-validation of covariance structures. *Multivariate Behavioral Research, 18*, 147-167.

Darroch, J.N. & Ratcliff, D. (1972). Generalized iterative scaling for log-linear models. *Annals of Mathematical Statistics, 43*, 1470-1480.

Davis, J.A. (1974). Hierarchical models for significance tests in multivariate contingency tables. In H.L. Costner (Ed.), *Sociological Methodology 1973-1974* (pp. 189-231). San Francisco: Jossey-Bass.

Dayton, C.M. & Macready, G.B. (1976). A probabilistic model for the validation of behavioral hierarchies. *Psychometrika, 41*, 189-204.

Dayton, C.M. & Macready, G.B. (1980). A scaling model with response errors and intrinsically unscalable individuals. *Psychometrika, 45*, 343-356.

Dayton, C.M. & Macready, G.B. (1988). A latent class covariate model with applications to criterion-referenced testing. In R. Langeheine & J. Rost (Eds.), *Latent trait and latent class models* (pp. 129-143). New York: Plenum Press.

De Leeuw, J. & Verhelst, N. (1986). Maximum likelihood estimation in generalized Rasch models. *Journal of Educational Statistics, 11*, 183-196.

Dempster, A.P., Laird, N.M. & Rubin, D.B. (1977). Maximum likelihood from incomplete data via the EM- algorithm. *Journal of the Royal Statistical Society. Series B. 39*, 1-38.

Duncan, O.D. (1975). Partitioning polytomous variables in multiway contingency analysis. *Social Science Research, 4*, 167-182.

Duncan, O.D. (1984a). The latent trait approach in survey research. In C.F. Turner & E. Martin (Eds.), *Surveying subjective phenomena*, vol. 1 (pp. 210-229). New York: Russell Sage Foundation.

Duncan, O.D. (1984b). Rasch measurement: further examples and discussion. In C.F. Turner & E. Martin (Eds.), *Surveying subjective phenomena*, vol. 2 (pp. 367-403). New York: Russell Sage Foundation.

Embretson, S.E. (1984). A general latent trait model for response process. *Psychometrika, 49*, 175-186.

Engelen, R.J.H. (1989). *Parameter Estimation in the logistic item response model.* Unpublished doctoral dissertation. Enschede, The Netherlands.

Efron, B. & Hinkley, D.V. (1978). Assessing the accuracy of the maximum likelihood estimator: observed versus expected Fisher information. *Biometrika, 65*, 457-487.

Evers, M. & Namboodiri, N.K. (1978). On the design matrix strategy in the analysis of categorical data. In K.F. Schuessler (Ed.), *Sociological Methodology 1979* (pp. 86-111). San Francisco: Jossey-Bass.

Felling, A., Peters, J. & Schreuder, O. (1987). *Religion in Dutch society 85; documentation of a national survey on religious and secular attitudes in 1985.* Amsterdam: Steinmetz Archive.

Fischer, G.H. (1970). A further note on estimation in Rasch's measurement model with two categories of answers. Wien: Research Bulletin Nr. 3/70, Psychologisch Institut Universität Wien.

Fischer, G.H. (1974). *Einführung in die Theorie psychologischer Tests.* Wien: Verlag Hans Huber.

Fischer, G.H. (1983). Logistic latent trait models with linear constraints. *Psychometrika, 48*, 3-26.

Fischer, G.H. (1987). Applying the principle of specific objectivity and of generalizabilty to the measurement of change. *Psychometrika, 52*, 565-587.

Fischer, G.H. (1989). An IRT-based model for dichotomous longitudinal data. *Psychometrika, 54*, 599-624.

Fischer, G.H. & Scheiblechner, H.H. (1970). Algorithmen und Programmen für das probabilistische Testmodell von Rasch. *Psychologische Beitrage, 12*, 23-51.

Fischer, G.H. & Parzer, P. (1991). An extension of the rating scale model with an application to the measurement of change. *Psychometrika, 56*, 637-651.

Fienberg, S.E. (1977, 1980). *The analysis of cross-classified categorical data.* Cambridge MA: MIT Press.

Follman, D. (1988). Consistent estimation in the Rasch model based on nonparametric margins. *Psychometrika, 53*, 553-562.

Formann, A.K. (1984). *Die Latent-Class-Analyse. Einführung in Theorie und Anwendung.* Weinheim: Beltz.

Formann, A.K. (1985). Constrained latent class analysis. *British Journal of Mathematical and Statistical Psychology, 38*, 87-111.

Formann, A.K. (1988). Latent class models for nonmonotone dichotomous items. *Psychometrika, 53*, 45-62.

Formann, A.K. (1992). Linear logistic latent class analysis for polytomous data. *Journal of the American Statistical Association, 87*, 476-486.

Glas, C.A.W. (1989). *Contributions to estimating and testing Rasch models.* Unpublished doctoral dissertation. Enschede, The Netherlands.

Glas, C.A.W. & Verhelst, N.D. (1989). Extensions of the partial credit model. *Psychometrika, 54*, 635-659.

Goodman, L.A. (1964). Simultaneous confidence limits for cross-product ratios in contingency tables. *Journal of the Royal Statistical Society. Series B. 26*, 86-102.

Goodman, L.A. (1968). The analysis of cross-classified data: independence, quasi-independence and interaction in contingency tables with or without missing cells. *Journal of the American Statistical Association, 63*, 1019-1131.

Goodman, L.A. (1971). The analysis of multidimensional contingency tables: stepwise procedures and direct estimation methods for building models for multiple classification. *Technometrics, 13*, 33-61.

Goodman, L.A. (1972a). A modified multiple regression approach to the analysis of dichotomous variables. *American Sociological Review, 37*, 28-46.

Goodman, L.A. (1972b). A general model for the analysis of surveys. *American Journal of Sociology, 77*, 1035-1086.

Goodman, L.A. (1973a). Guided and unguided methods for the selection of models for a set of T multidimensional contingency tables. *Journal of the American Statistical Association, 68*, 165-175.

Goodman, L.A. (1973b). The analysis of multidimensional contingency tables when some variables are posterior to others: A modified path analysis approach. *Biometrika, 60*, 179-192.

Goodman, L.A. (1973c). Causal analysis of panel studies and other kinds of surveys. *American Journal of Sociology, 78*, 1135-1191.

Goodman, L.A. (1974a). The analysis of systems of qualitative variables when some of the variables are unobservable. Part I - a modified latent structure approach. *American Journal of Sociology, 79*, 1179-1259.

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

Goodman, L.A. (1975). The relationship between modified and usual multiple regression approaches to the analysis of dichotomous variables. In D.R. Heise (Ed.), *Sociological Methodology 1976* (pp. 83-110). San Francisco: Jossey-Bass.

Goodman, L.A. (1978). *Analyzing qualitative/categorical data. Log-linear models and latent structure analysis.* Cambridge, MA: Abt Books. and Wesley.

Goodman, L.A. (1979). Simple models for the analysis of association in cross-classifications having ordered categories. *Journal of the American Statistical Association, 74*, 537-552.

Goodman, L.A. (1981). Three elementary views of log-linear models for the analysis of cross-classifications having ordered categories. In S. Leinhardt (Ed.), *Sociological Methodology 1981* (pp. 193-239). San Francisco: Jossey-Bass.

Goodman, L.A. (1983). The analysis of dependence in cross-classifications having ordered categories, using log-linear models for frequencies and loglinear models for odds. *Biometrics, 39*, 149-160.

Goodman, L.A. (1984). *The analysis of cross-classified data having ordered categories.* Cambridge, MA: Harvard University Press.

Goodman, L.A. (1990). Total-score models and Rasch-type models for the analysis of a multidimensional contingency table, or a set of multidimensional contingency tables with specified and/or unspecified order for response categories. In C.C. Clogg (Ed.), *Sociological Methodology 1990* (pp. 249-294). Oxford: Basil Blackwell.

Goodman, L.A. & Kruskal, W.H. (1954). Measures of association for cross-classifications. *Journal of the Royal Statistical Association, 49*, 732-764.

Guttman, L.A. (1950). The basis for scalogram analysis. In S.A. Stouffer et al. (Eds.), *Measurement and Prediction* (pp. 60-90). Princeton (New Yersey): Princeton University Press.

Haberman, S.J. (1974a). Log-linear models for frequency tables derived by indirect observation. *Annals of Statistics, 2*, 911-924.

Haberman, S.J. (1974b). *The analysis of frequency data.* Chicago: University of Chicago Press.

Haberman, S.J. (1976). Iterative scaling procedures for log-linear models for frequency data derived by indirect observation. *Proceedings of the American Statistical Association 1975: Statistical Computing Section*, (pp. 45-50).

Haberman, S.J. (1977a). Log-linear models for frequency tables with small expected cell counts. *Annals of Statistics, 5*, 1148-1169.

Haberman, S.J. (1977b). Maximum likelihood estimates in exponential response models. *Annals of Statistics, 5*, 815-841.

Haberman, S.J. (1978). *Analysis of qualitative data, vol 1. Introductory Topics.* New York: Academic Press.

Haberman, S.J. (1979). *Analysis of qualitative data, vol 2. New Developments.* New York: Academic Press.

Haberman, S.J. (1988). A stabilized Newton-Raphson algorithm for log-linear models for frequency tables derived by indirect observation. In C.C. Clogg (Ed.), *Sociological Methodology 1988* (pp. 193-212). Washington, DC: American Sociological Association.

Haertel, E.H. (1990). Continuous and discrete latent structure models for item response data. *Psychometrika, 55*, 477-494.

Hagenaars, J.A. (1985). *Loglineaire analyse van herhaalde surveys: Panel-, trend- en cohortonderzoek.* Unpublished doctoral dissertation. Tilburg, The Netherlands.

Hagenaars, J.A. (1988). LCAG - loglinear modelling with latent variables: a modified LISREL approach. In W.E.Saris & I.N. Gallhofer (Eds.), *Sociometric research, vol. 2. Data analysis* (pp. 111-131). London, England: MacMillan.

Hagenaars, J.A. (1990). *Categorical longitudinal data; log-linear panel, trend and cohort analysis.* Newbury Park, CA: Sage.

Hagenaars, J.A. (1992a). Exemplifying longitudinal log-linear analysis with latent variables. In P.G.M. van der Heijden, W. Jansen, B. Francis & G.U.H. Seeber (Eds.), *Statistical Modelling* (pp. 105-120). Amsterdam: Elsevier.

Hagenaars, J.A. (1992b). Latent variables in log-linear models of repeated observations, Working Paper Series #67, Tilburg University, Faculty of Social Sciences.

Hagenaars, J.A., Heinen A.G.J. & Hamers, P.A.M. (1980). Causale modellen met diskrete latente variabelen: een variant op de Lisrel-benadering. *MDN, 5*, 38-54.

Hagenaars, J.A. & Luijkx, R. (1987). LCAG; Latent class models and other log-linear models with latent variables. User's manual, Working Paper Series #17, Tilburg University, Department of Sociology.

Hambleton, R.K. & Swaminathan, H. (1985). *Item response theory.* Boston: Kluwer-Nijhoff Publishing.

Heckman, J. & Singer, B. (1982). The identification problem in econometric models for duration data. In: W. Hildebrand (Ed.), *Advances in Econometrics* (pp. 39-77). Cambridge: Cambridge University Press.

Heinen, A., Hagenaars, J. & Croon, M. (1988). Latent trait models in LCA perspective. Paper presented at the SMABS-conference, Groningen.

Heinen, A. & Vermaseren, P. (1992). Diltran users guide. Tilburg University, Department of Social Sciences.

Heinen, A. & Croon, M. (1992). Latent structure measurement models for polytomous items. Paper presented on the International Conference on Social Science Methodology. Trento, Italy, June 1992.

Hoijtink, H. (1990). A latent trait model for dichotomous choice data. *Psychometrika, 55*, 641-656.

Holland, P.W. (1990a). The Dutch identity: a new tool for the study of item response models. *Psychometrika, 55*, 5-18.

Holland, P.W. (1990b). On the sampling theory foundations of item response theory models. *Psychometrika, 55*, 577-601.

Holland, P.W. & Rosenbaum, P.R. (1986). Conditional association and unidimensionality in monotone latent variable models. *Annals of Statistics, 14*, 1523-1543.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics, 6*, 65-70.

Holt, J.A. & MacReady, G.B. (1989). A simulation study of the difference chi-square statistic for comparing latent class models under violation of regularity conditions. *Applied Psychological Measurement, 13*, 221-231.

Hout, M., Duncan, O.D. & Sobel, M.E. (1987). Association and heterogeneity: structural models of similarities and differences. In C.C. Clogg (Ed.), *Sociological Methodology 1987* (pp. 145-184). Washington, DC: American Sociological Association.

Irtel, H. (1987). On specific objectivity as a concept in measurement. In E.E. Roskam & R. Suck (Eds.), *Progress in Mathematical Psychology, 1.* Amsterdam, The Netherlands: Elsevier.

Jansen, P.G.W. (1983). *Rasch analysis of attitudinal data.* 's-Gravenhage, PTT. Unpublished doctoral dissertation.

Jansen, P.G.W. & Roskam, E.E. (1986). Latent trait models and dichotomization of graded responses. *Psychometrika, 51*, 69-91.

Jöreskog, K.G. & Sörbom, D. (1988). *LISREL 7. A guide to the program and applications.* Chicago: SPSS Inc.

Karlin, S. & Studden, W.J. (1966). *Tchebycheff systems: with applications in analysis and statistics.* New York: John Wiley.

Kelderman, H. (1984). Log-linear Rasch model tests. *Psychometrika, 49*, 223-245.

Kelderman, H. (1987). *Quasi-loglinear models for test and item analysis.* Unpublished doctoral dissertation, Enschede.

Kelderman, H. (1988). Loglinear multidimensional IRT models for polytomously scored items. Research report 88-17. Enschede: University of Twente, department of education.

Kelderman, H. (1989). Item bias detection using log-linear IRT. *Psychometrika, 54*, 681-697.

Kelderman, H. & Macready, G.B. (1990). The use of loglinear models for assessing differential item functioning across manifest and latent examinee groups. *Journal of Educational Measurement, 27*, 307-327.

Kerlinger, F.N. & Pedhazur, E.J. (1973). *Multiple regression in behavioral research.* New York: Holt, Rinehart & Winston.

Kiefer, J. & Wolfowitz, J. (1956). Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters. *Annals of Mathematical Statistics, 27*, 887-906.

Kreiner, S. (1987). Analysis of multidimensional contingency tables by exact conditional tests: techniques and strategies. *Scandinavian Journal of Statistics, 14*, 97-112.

Langeheine, R. (1988). New developments in latent class theory. In R. Langeheine & J. Rost (Eds.), *Latent trait and latent class models* (pp. 77-108). New York: Plenum Press.

Langeheine, R. & Rost, J. (1988) Introduction and overview. In R. Langeheine & J. Rost (Eds.), *Latent trait and latent class models* (pp. 1-7). New York: Plenum Press.

Lawley, D.N. (1943). On problems connected with item selection and test construction. *Proceeding of the Royal Statistical Society of Edinburgh, 61*, 273-287.

Lazarsfeld, P.F. (1950a). The logical and mathematical foundation of latent structure analysis. In S. Stouffer (Ed.), *Measurement and prediction* (pp. 362-412), Princeton, NJ: Princeton University Press.

Lazarsfeld, P.F. (1950b). The interpretation and mathematical foundation of latent structure analysis. In S. Stouffer (Ed.), *Measurement and prediction* (pp. 413-472), Princeton, NJ: Princeton University Press.

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

Lindsay, B., Clogg, C.C. & Grego J. (1991). Semiparametric estimation in the Rasch model and related models, including a simple latent class model for item analysis. *Journal of the American Statistical Association, 86*, 96-107.

Long, J.S. (1984). Estimable functions in log-linear models. *Sociological Methods & Research, 12*, 399-432.

Lord, F.M. (1952). A theory of test scores. *Psychometrika Monograph No. 7, 17* (4, Pt. 2).

Lord, F.M. (1967). *An analysis of the Verbal Scholastic Aptitude Test using Birnbaum's three-parametric logistic model* (ETS Research Bulletin RB-67-34). Princeton, NJ: Educational Testing Service.

Lord, F.M. & Novick, M.R. (1968). *Statistical theories of mental test scores.* Reading, MA: Addison-Wesley.

Louis, Th.A. (1982). Finding the observed information matrix when using the EM algorithm. *Journal of the Royal Statistical Society. Series B 44*, 226-233.

Luce, R.D. (1959). *Individual choice behavior.* New York: Wiley.

Martin-Löf, P. (1973). *Statistika Modeller. Anteckningar från seminarier Lasåret 1969-1970 utarbetade av Rolf Sunberg obetydligt ändrat nytryk, oktober 1973.* Stockholm: Institutet för Försäkringsmatematik och Matematisk Statistik vid Stockholms Universitet.

Masters, G.N. (1982). A Rasch model for partial credit scoring. *Psychometrika, 47*, 149-174.

Masters, G.N. (1985). A comparison of latent trait and latent class analyses of Likert-type data. *Psychometrika, 50*, 69-82.

Masters, G.N. & Wright, B.D. (1984). The essential process in a family of measurement models. *Psychometrika, 49*, 529-544.

McCullagh, P. & Nelder, J.A. (1983, 1989). *Generalized linear models.* London, England: Chapman and Hall.

McCutcheon, A.L. (1987). *Sage university paper 64. Latent class analysis.* Newbury Park, CA: Sage.

McCutcheon, A.L. (1993) Logit models with latent response and polytomous effects variables. In A. von Eye & C.C. Clogg (Eds.) *Analysis of Latent Variables in Developmental Research.* New York: Academic Press.

McDonald, R.P. (1967). Non-linear factor analysis. *Psychometric Monographs*, No. 15.

McDonald, R.P. & Marsh, H.W. (1990). Choosing a multivariate model: noncentrality and goodness of fit. *Psychological Bulletin, 107*, 247-255.

McLaghlan, G.J. & Basford, K.E. (1988). *Mixture models: inference and applications to clustering.* New York: M. Dekker, Inc.

Meilijson, I. (1989). A fast improvement to the EM algorithm on its own terms. *Journal of the Royal Statistical Society. Series B. 51*, 127-138.

Mellenbergh, G.J. & Vijn, P. (1981). The Rasch model as a log-linear model. *Applied Psychological Measurement, 5*, 369-376.

Mislevy, R.J. (1984). Estimating latent distributions. *Psychometrika, 49*, 359-381.

Mislevy, R.J. (1986). Bayes modal estimation in item response models. *Psychometrika, 51*, 177-195.

Mislevy, R.J. & Verhelst, N. (1990). Modeling item responses when different subjects employ different solution strategies. *Psychometrika, 55*, 195-215.

Molenaar, I.W. (1983a). Item steps. *Heymans Bulletins* (Report No. HB-83-630-EX). Groningen: Rijks Universiteit Groningen.

Molenaar, I.W. (1983b). Some improved diagnostics for failure of the Rasch model. *Psychometrika, 48*, 49-72.

Mood, A., Graybill, F.A. & Boes, D.C. (1974). *Introduction to the theory of statistics.* Tokyo, Japan: McGraw-Hill.

Mooijaart, A. & van der Heijden, P.G.M. (1992). The EM algorithm for latent class analysis with equality constraints. *Psychometrika, 57*, 261-269.

Muraki, E. (1992). A generalized partial credit model: application of an EM algorithm. *Applied Psychological Measurement, 16*, 159-176.

Nelder, J.A. & Wedderburn, R.W.M. (1972). Generalized linear models. *Journal of the Royal Statistical Society. Series A. 135*, 370-384.

Neyman, J. & Scott, E.L. (1948). Consistent estimates based on partially consistent observations. *Econometrika, 16*, 1-32.

Proctor, C.H. (1970). A probabilistic formulation and statistical analysis of Guttman scaling. *Psychometrika, 35*, 73-78.

Raftery, A.E. (1986a). Choosing models for cross-classifications. *American Sociological Review, 51*, 145-146.

Raftery, A.E. (1986b). A note on Bayes factors for log-linear contingency table models with vague prior information. *Journal of the Royal Statistical Society. Series B. 48*, 249-250.

Rao, C.R. (1973). *Linear statistical inference and its applications.* (2nd edn). New York: John Wiley.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests.* Copenhagen: Danish institute of educational research.

Rasch, G. (1961). On general laws and the meaning of measurement in psychology. *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, 321-333.

Rasch, G. (1977). On specific objectivity: an attempt at formalizing the request for generality and validity of scientific statements. In M. Blegvad (Ed.), *Danish Yearbook of Philosophy.* Copenhagen: Munksgaard.

Read, T.R. & Cressie, N.A. (1988). *Goodness-of-fit statistics for discrete multivariate data.* New York/Berlin: Springer-Verlag.

Reckase, M.D. (1985). The difficulty of test items that measure more than one ability. *Applied Psychological Measurement, 9*, 401-412.

Reiser, M. (1981). Latent trait modeling of attitude items. In C.W. Bohrnstedt & E.F. Borgatta (Eds.), *Social Measurement* (pp. 117-144). Beverly Hills: Sage.

Rigdon, S.E. & Tsutakawa, R.K. (1983). Parameter estimation in latent trait models. *Psychometrika, 48*, 567-574.

Roskam, E.E. & Jansen, P.G.W. (1984). A new derivation of the Rasch model. In E. Degreef & J. van Buggenhaut (Eds.), *Trends in Mathematical Psychology.* Amsterdam, The Netherlands: Elsevier.

Rost, J. (1985). A latent class model rating data. *Psychometrika, 50*, 37-49.

Rost, J. (1988a). Rating scale analysis with latent class models. *Psychometrika, 53*, 327-348.

Rost, J. (1988b). Test theory with qualitative and quantitative latent variables. In R. Langeheine & J. Rost (Eds.), *Latent trait and latent class models* (pp. 147-171). New York: Plenum Press.

Rost, J. (1990). Rasch models in latent classes: an integration of two approaches to item analysis. *Applied Psychological Measurement, 14*, 271-282.

Rost, J. (1991). A logistic mixture distribution model for polychotomous item responses. *British Journal of Mathematical and Statistical Psychology, 44*, 75-92.

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph No. 17, 34* (4. Pt. 2).

Schwarz, G. (1978). Estimating the dimensions of a model. *Annals of Statistics, 6*, 461-464.

Sclove, S.L. (1987). Application of model selection criteria to some problems in multivariate analysis. *Psychometrika, 52*, 333-343.

Sobel, M.E. & Bohrnstedt, G.W. (1985). Use of null models in evaluating the fit of covariance structure models. In N.B. Tuma (Ed.), *Sociological Methodology 1985* (pp. 152-178). San Francisco: Jossey-Bass.

Spray, J.A., Davey, T.C., Reckase, M. D., Ackerman, T.A. & Carlson, J.E. (1990). Comparison of two logistic multidimensional item response theory models. ACT Research Report Series, research report ONR90-8. Iowa City: American College Testing Program.

Steiger, J.H. (1979). Factor indeterminacy in the 1930's and the 1970's: some interesting parallels. *Psychometrika, 44*, 157-167.

Stouthard, Ph.C. (1965). *Data modellen. Enkele toepassingen en een methodologische kritiek.* Unpublished doctoral dissertation. Tilburg, The Netherlands.

Stroud, A.H. & Secrest, D. (1966). *Gaussian quadrature formulas.* Englewood Cliffs NJ: Prentice-Hall.

Swaminathan, H. & Gifford, J.A. (1982). Baysian estimation in the Rasch model. *Journal of Educational Statistics, 7*, 175-191.

Sympson, J.B. (1978). A model for testing with multidimensional items. In D.J. Weiss (Ed.), *Proceedings of the 1977 Computerized Adaptive Testing Conference* (pp. 82-98). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program.

Thissen, D. (1982). Marginal maximum likelihood estimation for the one-parameter logistic model. *Psychometrika, 47*, 175-186.

Thissen, D. (1988). *Multilog; multiple, categorical item analysis and test scoring using item response theory.* Version 5.1. Mooresville IN: Scientific Software Inc.

Thissen, D. & Steinberg, L. (1984). A response model for multiple choice items. *Psychometrika, 49*, 501-519.

Thissen, D. & Steinberg, L. (1986). A taxonomy of item response models. *Psychometrika, 51*, 567-577.

Thissen, D. & Mooney, J.A. (1989). Log-linear item response models with applications to data from social surveys. In C.C. Clogg (Ed.), *Sociological Methodology 1989* (pp. 299-330). Oxford: Basil Blackwell.

Tjur, T. (1982). A connection between Rasch's item analysis model and a multiplicative Poisson model. *Scandinavian Journal of Statistics, 9*, 23-30.

Tucker, L.R. (1946). Maximum validity of a test with equivalent items. *Psychometrika, 11*, 1-14.

Torgerson, W.S. (1962). *Theory and methods of scaling.* New York: John Wiley.

Van de Pol, F. & de Leeuw, J. (1986). A latent Markov model to correct for measurement error. *Sociological Methods & Research, 15*, 118-141.

Van den Wollenberg, A.L. (1979). *The Rasch model and time-limit tests.* Unpublished doctoral dissertation. Nijmegen, The Netherlands.

Van den Wollenberg, A.L. (1982a). A simple and effective method to test the dimensionality axiom of the Rasch model. *Applied Psychological Measurement, 6*, 83-91.

Van den Wollenberg, A.L. (1982b). Two new test statistics for the Rasch model. *Psychometrika, 47*, 123-139.

Van den Wollenberg, A.L. (1988). Testing a latent trait model. In R. Langeheine & J. Rost (Eds.), *Latent trait and latent class models* (pp. 31-50). New York: Plenum Press.

Verhelst, N.D., Glas, C.A.W. & van der Sluis, A. (1984). Estimation problems in the Rasch model: the basic symmetric functions. *Computational Statistics Quarterly, 1*, 245-262.

Verhelst, N.D. (1992). Het eenparameter logistisch model (OPLM); een theoretische inleiding en een handleiding bij het computerprogramma. OPD Memorandum 92-3. Arnhem, The Netherlands: CITO.

Vermunt, J. (1992). LEM: log-linear and event history analysis with missing data. User's manual. Tilburg University, Department of Social Sciences.

Wheaton, B. (1987). Assessment of fit in overidentified models with latent variables. *Sociological Methods & Research, 16*, 118-154.

Whittaker, J. (1990). *Graphical models in applied multivariate statistics.* Chichester, England: John Wiley.

Wiggins, L.M. (1955). *Mathematical models for the analysis of multi-wave panels.* Doctoral dissertation series No. 12.481. Ann Arbor, MI.

Wiggins, L.M. (1973). *Panel analysis: latent probability models for attitude and behavior processes.* Amsterdam, The Netherlands: Elsevier.

Wright, B.D. & Panchapakesan, N.A. (1969). A procedure for sample-free item analysis. *Educational and Psychological Measurement, 6*, 83-91.

Wright, B.D. & Douglas, G.A. (1977a). Best procedures for sample-free item analysis. *Applied Psychological Measurement, 1*, 281-295.

Wright, B.D. & Douglas, G.A. (1977b). Conditional versus unconditional procedures for sample-free analysis. *Educational and Psychological Measurement, 37*, 573-586.

Wright, B.D. & Masters, G.N. (1982). *Rating scale analysis.* Chicago: MESA Press.

Zwinderman, A.H. (1991a). *Studies of estimating Rasch models.* NICI Technical Report 91-02. Nijmegen, The Netherlands.

Zwinderman, A.H. (1991b). A generalized Rasch model for manifest predictors. *Psychometrika, 56*, 589-600.

Latent trait and latent class models are both measurement models for discrete manifest indicators. Many authors have stressed the differences between the two kinds of models. In latent trait models it is assumed that the latent variable is metrical and continuous, while latent class models are developed to handle discrete latent variables that can be measured on every measurement level (so, nominal, ordinal or metrical). The fact that for a long time the two models have been considered as related but essentially different can also be explained by their different origins. Latent trait models were developed by psychometricians, latent class models by sociologists. Only quite recently latent trait models have been applied to the analysis of attitudinal data.

In *Discrete Latent Variable Models*, the author shows how to bridge the gap between both models. When the parameters in latent trait models are estimated using semi-parametric MML, the continuous latent variable can be discretized. This results in a latent trait model that is identical to a latent class model with certain restrictions on the log-linear parameters.

The benefit of linking latent trait and latent class models in this way is that results obtained in one tradition can be used in the other tradition. The definition of multidimensional latent variables and the incorporation of external variables in the measurement model, which can straightforwardly be done in latent class models, are easily transferred to discretized latent trait models.

*Discrete Latent Variable Models* offers a review of latent class and latent trait models. Methods for estimating parameters and testing the model fit are described in detail. Moreover, this book helps to understand how the two types of models are related to each other by using a common log-linear parameterization.

TILBURG UNIVERSITY PRESS
P.O. BOX 90153
5000 LE TILBURG
THE NETHERLANDS

ISBN 90-361-9753-8