Chapter 1

A Guide through Latent Structure Models for Categorical Data

Jürgen Rost and Rolf Langeheine

IPN-Institute for Science Education, Kiel

Contents:

The world of latent structure models for categorical data is complex enough to need some guideline for those researchers who look for a model that fits to their substantive theory and to their data. The basic dichotomy for systematizing these models still is the distinction between latent traits and latent classes. Latent trait refers to a latent continuum or a dimension which all individuals are mapped on, based on their pattern of responses on a set of categorical variables. Latent classes refer to the categories of a latent variable that is discrete, the categories of which may but need not be ordered along a continuum.

This distinction still gives some orientation despite the fact that the borderline between both types of models is not very sharp. Latent trait models, e.g., the Rasch model or the Mokken analysis, may be formalized as special latent class models and latent class models can be specified in terms of discretized and sometimes restricted latent trait models (see e.g., Heinen, 1993) The distinction between traits and classes is used here to define the two sections of the present chapter. The criterion for allocating the models properly is the question of whether a model is aimed at measuring a dimension, be it discretized or treated as a continuum, or at differentiating between classes or types of individuals. In this sense, e.g., the model of Mokken analysis is referred to as a latent trait model (see section 1.6) although it can be formalized as a special latent class model (see Croon, 1991).

1. Measuring latent dimensions

This section presents latent trait models that are applied in one or more of the following chapters on applications. The structure of this section is the result of several distinctions.

First, an ordinary latent trait model refers to (and helps measure) only *one* latent dimension. Such models are treated in sections 1.1 to 1.7. Sections 1.8 to 1.10 refer to models assuming more than one continuous variable, be it that two latent traits are measured simultaneously (1.8), be it that covariates of the trait or the manifest indicators are taken into account (1.9), or that a distinction is made between a trait and a state variable (1.10).

Second, the type of data a model has been tailored for is a criterion for differentiating between models. The simpliest and most common situation is given by *dichotomous* data, where the manifest variables take on two values only, 0 and 1. Sections 1.1 to 1.3, 1.8 and 1.9 refer to this type of data, although some of the models presented can be or have already been generalized to other types of data. *Polytomous* data are dealt with in sections 1.4, 1.6, 1.7 and 1.10. What is called 'count data' in section 1.5 are polytomous data without an upper limit as obtained from, e.g., counting reading errors when students are reading essays.

Third, the complexity of parameterization discriminates between latent trait models. Whereas section 1.1 deals with the Rasch model, which is the most simple but a very attractive model within the family of logistic item response models, section 1.2 introduces complexity to the Rasch model by allowing for different kinds of structural parameters (multifacets). Section 1.3 describes the Birnbaum model which is, in a different sense, more complex than the Rasch model, i.e., it has a second parameter, a discrimination parameter for each item. Section 1.6, finally, does not provide metric parameters for the items, i.e., it deals with nonparametric item response models.

1.1 The dichotomous Rasch model

In the Rasch model (Rasch 1960, 1980) it is assumed that the probability of a positive or correct response of individual v on item i, $p(X_{vi} = 1)$, depends only on the ability or trait value of person v, θ_v, and the difficulty of item i, σ_i. The most simple model that takes these two factors into account is an additive main-effects model for the logits of the response probabilities:

$$\log \frac{p(X_{vi} = 1)}{p(X_{vi} = 0)} = \theta_v - \sigma_i \ . \tag{1}$$

The item parameter σ_i is written here with a negative sign, because it is a difficulty and not an easiness parameter and, hence, related reciprocally to the probability of a positive response.

Usually, the Rasch model is written in the algebraically equivalent form

$$p(X_{vi} = x) = \frac{\exp(x(\theta_v - \sigma_i))}{1 + \exp(\theta_v - \sigma_i)} \quad , x \in \{0,1\} \quad , \tag{2}$$

or as the probability of a response vector of k locally independent item responses $\mathbf{x_v} = (x_1, x_2, \ldots x_i \ldots x_k)$,

$$p(\mathbf{x}_v) = \prod_{i=1}^{k} \frac{\exp\left(x_{vi}(\theta_v - \sigma_i)\right)}{1 + \exp(\theta_v - \sigma_i)} = \frac{\exp\left(r_v\,\theta_v - \sum_{i=1}^{k} x_{vi}\,\sigma_i\right)}{\prod_{i=1}^{k}\left(1 + \exp(\theta_v - \sigma_i)\right)} \quad . \tag{3}$$

This equation shows the most relevant statistical property of Rasch models, i.e., the score $r_v = \sum_{i=1}^{k} x_{vi}$ of an individual v is sufficient for estimating his or her trait parameter θ_v. A consequence of this property is the possibility to estimate the item parameters conditionally upon the scores of the observed response patterns by means of conditional pattern probabilities

$$p(\mathbf{x}_v | r_v) = \frac{\exp\left(-\sum_{i=1}^{k} x_{vi}\,\sigma_i\right)}{\gamma_r(\exp(-\boldsymbol{\sigma}))} \quad , \tag{4}$$

where γ_r denotes the so-called symmetric functions of the de-logarithmized item parameters (see Rasch, 1980). These pattern probabilities do not contain the trait parameters θ_v and parameter estimation based upon these probabilities avoids problems of simultaneous estimation of incidental and structural parameters.

Another consequence of this property is the possibility of formalizing the Rasch model as a log-linear model (Kelderman, 1984; Tjur, 1982). Using equation (4), the probability of an arbitrary pattern **x** with score $r_{\mathbf{x}}$ can be rewritten as

$$p(\mathbf{x}) = p(\mathbf{x} | r_{\mathbf{x}}) \cdot p(r_{\mathbf{x}}) \quad , \tag{5}$$

the logarithm of which is

$$\log(p(\mathbf{x})) = -\sum_{i=1}^{k} x_i\,\sigma_i - \log\left(\gamma_r(\exp(-\boldsymbol{\sigma}))\right) + \log\left(p(r_{\mathbf{x}})\right) \quad . \tag{6}$$

In this equation, x_i are the components of the response vector **x** and the second and third additive term only depend on the score r of vector **x**, but not on the particular pattern of the r positive responses in vector **x**. In a log-linear notation of equation (6), let $m(\mathbf{x}) = N\,p(\mathbf{x})$ denote the expected frequency of pattern **x**, and N the sample size, then the log-linear structure

$$\log(m(\mathbf{x})) = -\sum_{i=1}^{k} x_i\,\sigma_i + \lambda_r \tag{7}$$

is an equivalent formalization of the Rasch model with score parameters (Kelderman, 1984)

$$\lambda_r = \log\left(\frac{N \cdot p(r_{\mathbf{x}})}{\gamma_r(\exp(-\boldsymbol{\sigma}))}\right) \quad . \tag{7a}$$

Summarizing the statistical properties, the dichotomous Rasch model can be characterized by the following features:

- All item response functions have the same slope, i.e., they are intersection free, and all items have the same discrimination with regard to the latent trait.
- There is no interaction between individuals and items,

i.e., person ability and item difficulty are additively connected on a logarithmic scale.

- The score of a response vector is a sufficient statistic for estimating the trait parameter associated with this vector, i.e., the pattern of a given number of positive responses provides no additional information about the ability of the person.
- Parameter estimation is unproblematic since the item parameters can be estimated without estimating the ability parameters.

In this volume the dichotomous Rasch model is applied to a scale of delinquent behavior (chapter 17 by Sorenson, Brownfield and Jensen) and to cognitive tasks in a study aimed at testing a Piagetian hypothesis on cognitive development (chapter 9 by Spiel, Gittler, Sirsch and Glück).

1.2 The multi-facet extension of the Rasch model

In the previous section the dichotomous Rasch model was introduced as a kind of main-effects model for two factors, i.e., persons and items. In many applications more than two factors are involved, e.g., when persons are tested at different time points or when the performances or free responses to test items are rated by several judges. In these cases, three factors or facets (persons × items × time-points, persons × items × judges) or even more facets have to be considered.

Assuming that there are no interactions of any order between the facets involved, the *multi-factorial* or *multi-facet* generalization of the Rasch model is (for the sake of simplicity the description is restricted to three factors only):

$$p\left(X_{vij} = x\right) = \frac{\exp\left(x\left(\theta_v - \sigma_i - \delta_j\right)\right)}{1 + \exp\left(\theta_v - \sigma_i - \delta_j\right)} \quad , \qquad (8)$$

where X_{vij} denotes the response of person v on item i at time point j (or rated by judge j) and δ_j is a parameter for the third factor, e.g., the severity of judge j. This extension of the Rasch model goes back to the early ideas of Rasch (unpublished) and has been elaborated on by Micko (1970), Fischer (1974), and by Linacre (1989) who presents a general version for more than three factors.

Although the model seems to be quite symmetric with respect to its three (or more) factors, the asymmetry caused by the distinction of incidental and structural parameters (Neyman and Scott, 1948) is still a feature of this model extension: The trait parameters are incidental and the parameters of all other facets are structural parameters. This differentiation is not only relevant for estimating the parameters of the multi-facet model (e.g., by conditioning out the incidental (trait) parameters) but also allows to consider model (8) as a special kind of a more general model structure. The general *linear logistic Rasch model* allows for decomposing a set of structural parameters into a weighted sum of component parameters:

$$p\left(X_{vi}=x\right)=\frac{\exp\left(x\left(\theta_v-\sum_{j=1}^{h} q_{ij}\,\eta_j\right)\right)}{1+\exp\left(\theta_v-\sum_{j=1}^{h} q_{ij}\,\eta_j\right)} \quad , \qquad (9)$$

where η_j denotes the difficulty parameter of component j and the q_{ij} are a priori fixed constants defining the weight of component j for item i. This model, called LLTM (linear logistic test model) has been considered extensively by Fischer (1972, 1973, 1983,1987).

In order to see that the multi-facet Rasch model (8) is a special case of the linear logistic model (9), let $\boldsymbol{\eta}$ be a vector of $k+m$ components, where the k components are the item parameters and the second m components are the parameters of judges (or time points). The response variables X_{vi} are defined for a set of $k \cdot m$ virtual items, i.e., the responses on k items rated by m judges (or observed at m time points). The q_{ij}-vector for each virtual item then simply is a vector of 0's and 1's so that each vector i has exactly two 1's, one for the real item involved and one for the judge.

The advantage of the linear logistic extension of the multi-facet model is that hypotheses about interactions between items and judges or items and time points can be modelled and tested empirically by specifying additional component parameters η_j.

Summarizing the features of these generalizations of the Rasch model, multi-facet and linear logistic models

- allow taking additional factors into account that influence the response behavior,
- allow testing the hypothesis that only main effects of these factors and no interaction effects influence the observed item responses.

The paper by Lunz and Wright (chapter 6) describes an application of the multi-facet Rasch modelto performance data of medical examinations.

1.3 The two-parameter logistic model

The dichotomous Rasch model is called the one-parameter logistic model, because only one parameter, i.e., a difficulty parameter is considered for each item. The *Birnbaum model* (Lord and Novick, 1968) provides a second parameter for each item, β_i, which takes into account the slope of the item response function:

$$p\left(X_{vi}=x\right)=\frac{\exp\left(x\,\beta_i\left(\theta_v-\sigma_i\right)\right)}{1+\exp\left(\beta_i\left(\theta_v-\sigma_i\right)\right)} \quad , x\in\{0,1\} \quad . \qquad (10)$$

Although this generalization of the Rasch model has the appealing feature of allowing for different item discriminations among the items of a test, this feature is not without problems, both from a psychological and a statistical point of view. From a psychological perspective it is not very plausible to assume that an item A has a higher probability of a correct solution than an item B for one part of the person population while it has a lower probability than B for the other part. The intersection point of two item response functions splits the latent continuum exactly into those two parts where a different order of response probabilities is given.

From a statistical point of view this multiplicative parameter β_i destroys the nice properties of the Rasch model, i.e., the score r_v is no longer a sufficient statistic for the ability parameter θ_v. Instead, the weighted sum

$$\sum_{i=1}^{k} \beta_i x_{vi} \tag{11}$$

is sufficient for θ_v, where β_i are the discrimination parameters. As a consequence, conditional parameter estimation as in the Rasch model is not possible for the structural parameters of the two-parameter logistic model.

Another consequence of the multiplicative connection between the model parameters is the fact that the Birnbaum model cannot be formalized as a log-linear model because the logarithmic pattern probabilities are no linear functions of the model parameters.

From an applied perspective, the Birnbaum model does what many practitioners regard as useful, i.e., to take into account more information from a response pattern than simply the number of positive responses. However, the intuitive belief of practitioners that the item difficulty should serve as a weight of a correct item response ('solving a difficult task counts more than solving an easy task') turns out to be wrong: It is not the difficulty but the discrimination parameters of the items that provide the best weights for correct responses in order to obtain an improved ability estimate.

Allerup (chapter 3) presents an application of the Birnbaum model to data from a large scale assessment study and discusses a simple estimation method for the model parameters. Croon and Heinen (chapter 12) take the Birnbaum model as a basis for their causal relationship model (see also section 1.9). In the description of optimal test design methods by Berger (chapter 5) reference is made to the two-parameter logistic model as the underlying IRT-model.

1.4 The polytomous Rasch model

With dichotomous item responses, the respondent has to pass only one threshold, i.e. the threshold between 0 and 1, between false and right, or between no and yes. When there are more than two response categories, the number of thresholds increases accordingly, i.e., a response format with $m + 1$ categories has m thresholds. Just as the dichotomous Rasch model defines the probability of passing the only given threshold by the simple logistic function (2), the polytomous extension of the Rasch model is obtained by defining *each* threshold probability q_{vix} by the simple logistic function with an own difficulty parameter τ_{ix} for each threshold:

$$q_{vix} = \frac{\exp\left(\theta_v - \tau_{ix}\right)}{1 + \exp\left(\theta_v - \tau_{ix}\right)} \quad , x \in \{1,2,\ldots m\} . \tag{12}$$

A threshold probability q_{vix} is formally defined as the probability of responding in the upper category x, if only two adjacent categories are considered

$$q_{vix} = \frac{p\left(X_{vi} = x\right)}{p\left(X_{vi} = x-1\right) + p\left(X_{vi} = x\right)} \quad . \tag{13}$$

From these equations the following response probability can be derived to serve as the model equation of the polytomous Rasch model (Andrich, 1978; Masters and Wright, 1984)

$$p\left(X_{vi} = x\right) = \frac{\exp\left(x\,\theta_v - \sum_{s=1}^{x} \tau_{is}\right)}{\sum_{y=0}^{m} \exp\left(y\,\theta_v - \sum_{s=1}^{y} \tau_{is}\right)} \quad , x \in \{0,1,\ldots m\} \quad . \tag{14}$$

This equation covers the definition of the probability of category 0, so that the definition $\sum_{s=1}^{0} \tau_{is} = 0$ must be taken into account (there is no threshold for category 0).

Figure 1 provides an illustration of the interpretation of the model parameters.

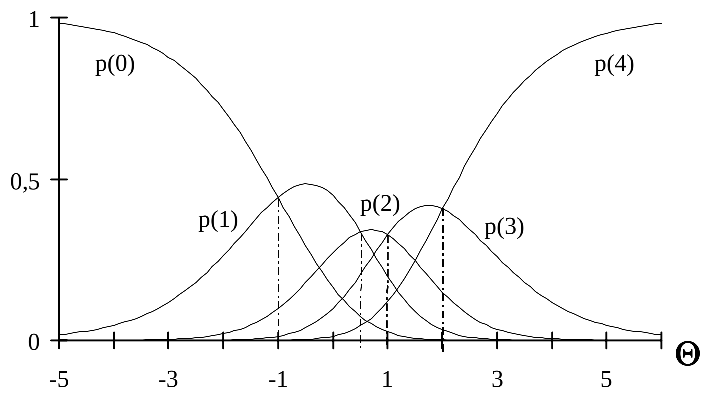

Figure 1: Category response functions of an item with 5 response categories and threshold parameters $\tau_{i1} = -1.0$, $\tau_{i2} = +0.5$, $\tau_{i3} = +1.0$, $\tau_{i4} = +2.0$

The threshold parameters τ_{ix} define the intersection points of each of two adjacent categories and, hence, provide a partition of the latent continuum into m + 1 sections, where one of the response categories has highest probability. However, this only holds if the threshold parameters are ordered, i.e., if the threshold difficulties increase with category number increasing. The order of the threshold parameters is *not* a formal characteristic of the model, i.e., it is possible to obtain unordered parameter estimates for empirical data without getting a bad model fit. However, it can be argued that ordered thresholds are an implication of the idea the model is conceptualized on and, hence, unordered threshold parameters would contradict the model assumption of ordered categories (see Andrich, de Jong and Sheridan, chapter 4).

Another feature of the model obvious from equation (14) is that the probability of each category depends on the difficulty of *all* thresholds. Although the numerator of (14) is only a function of the threshold difficulties up to category x, the denominator clearly depends on all thresholds, including those that *follow* category x. This is important for understanding the response process postulated by this model: it is *not a sequential* process climbing the steps from category 0 to the chosen category x, regardless of which (more difficult) steps follow after this category. Obviously it is a *simultaneous* process taking into account the difficulty of all thresholds and deciding for the most appropriate category (see Andrich et al., chapter 4).

The polytomous extension of the Rasch model which has been called *partial credit model* by Masters (1982) preserves all attractive statistical features of the dichotomous Rasch model, the most fundamental of which are simple sum scores as sufficient statistics of the model parameters. The pattern probability

$$p(\mathbf{x}_v) = \frac{\exp\left(r_v\,\theta_v - \sum_{i=1}^{k}\sum_{s=1}^{x_{vi}}\tau_{is}\right)}{\prod_{i=1}^{k}\left(\sum_{y=0}^{m}\exp\left(y\,\theta_v - \sum_{s=1}^{y}\tau_{is}\right)\right)} \qquad (15)$$

reveals that the sum score $r_v = \sum_{i=1}^{k} x_{vi}$ is a sufficient statistic for θ_v, but further derivations would also show that the number of responses in category x with item i is sufficient for the estimation of τ_{ix}.

As a consequence, the ability parameters can be conditioned out (see section 1.1) and the model can be formalized as a log-linear model

$$\log(m(\mathbf{x})) = -\sum_{i=1}^{k}\sigma_{ix} + \lambda_r \quad , \qquad (16)$$

where $\sigma_{ix} = \sum_{s=1}^{x}\tau_{is}$ is a cumulative parameter and λ_r a score parameter that depends on N, the score r of pattern **x**, and the symmetric functions of the cumulative parameters (cf. equation (7a)). Agresti (chapter 20) discusses several log-linear models that are equivalent to generalized Rasch models for ordinal data.

The polytomous Rasch model (14) is the super-model of some restricted versions aimed at modelling specific assumptions about the threshold distances. For example, when the model is applied to data gained with rating scales as a constant response format for all items, it is reasonable to assume that the threshold distances are constant for *all* items. The *rating scale model* based on the restriction

$$\tau_{ix} = \sigma_i + \psi_x \quad \text{and} \quad \sum_{x=1}^{m}\psi_x = 0 \qquad (17)$$

is a parameterization of this assumption, where σ_i is a difficulty parameter of item i which is, due to the norming condition, defined as the midpoint of all threshold locations (Andrich, 1978).

Another restriction follows from the assumption of equidistant thresholds for *each* item, introducing a second item parameter δ_i that defines this constant distance for item i:

$$\tau_{ix} = \sigma_i + \left(x - \frac{m+1}{2}\right)\delta_i. \qquad (18)$$

Again, σ_i is a difficulty parameter and δ_i a kind of dispersion parameter of item i (Andrich, 1982). The model based on this restriction is called the *dispersion model* or *equidistance model*.

The polytomous Rasch models (14), (17) and (18) can be used to test

- whether the response categories are ordered in the sense that they can be mapped onto successive intervals of the latent trait,
- whether the threshold distances are a characteristic of the response format (model (17)) or of the particular items (model (14)), and
- whether an equidistance assumption holds for the response categories or not.

All polytomous Rasch models provide information about the dispersion of the item responses as the threshold distances are inversely related to the dispersion of the expected responses.

Applications of polytomous Rasch models can be found in chapter 20 by Agresti on social survey data, in chapter 4 by Andrich et al. on an achievement test and an attitude questionnaire, in chapter 38 by Meiser and Rudinger on developmental data, in chapter 31 by Rost, Carstensen and v. Davier on personality questionnaire data, and in chapter 11 by Willmes on data of an aphasia test.

1.5 The Poisson model for count data

In some achievement tests, e.g., in reading skills tests, the response variable does not have a limited number of categories but rather an unlimited number, e.g., counts of special events like reading or writing errors. If the events occur independently of each other with constant probability π_v for a person v and if there *is* a clear upper limit, m, of the number of events, then the number of observed events would be binomially distributed:

$$p\left(X_v = x\right) = \binom{m}{x} \pi_v^x \left(1 - \pi_v\right)^{m-x} \quad , x \in \{0,1,2\ldots m\} . \tag{19}$$

When the event has *innumerable* opportunities to occur but little probability of doing so at any particular opportunity, the number of events follows the Poisson distribution

$$p\left(X_v = x\right) = \frac{\lambda_v^x}{x!\exp\left(\lambda_v\right)} \quad , \tag{20}$$

where λ_v is the expectation of the number of events for person v (see Rasch, 1960; Wright and Masters, 1982).

This model can be extended to a factorial design, where the same individuals solve various tasks, but the same ability θ_v is involved in all these tasks (items):

$$p\left(X_{vi} = x\right) = \frac{\lambda_{vi}^x}{x!\exp\left(\lambda_{vi}\right)} \tag{21}$$

and

$$\lambda_{vi} = \exp\left(\theta_v - \sigma_i\right).$$

In this notation, σ_i is a difficulty parameter for task or item i which is additively connected to the person's ability on a logarithmic scale. This model is equivalent to the multiplicative Poisson model, where

$$\lambda_{vi} = \varepsilon_v \cdot \tau_i \tag{22}$$

and

$$\varepsilon_v = \exp(\theta_v) \quad , \quad \tau_i = \exp(-\sigma_i).$$

The Poisson count model shares the most significant features with the logistic Rasch models: The score r_v of person v, which is the total number of events in all tasks, is a sufficient statistic for his/her ability parameter θ_v. Moreover, the ability parameter can be eliminated by using a conditional likelihood function (cf. section 1.1).

The Poisson count model can be applied whenever the observed item responses are counts of a particular event (failures or successes), which have no clear upper limit. In chapter 37 by Jansen the model is applied to a reading test, where the counts are the number of words that a subject can read within one minute. The chapter also presents further extensions of the model.

1.6 Nonparametric latent trait models

In all previous sections the item response probability was described by a particular function containing one or more parameters for each item. Nonparametric item response models are based on the idea that it might be too restrictive assuming a particular type of item response function (IRF) for all items of a test. As an alternative, these models formulate only minimal assumptions about the shape of the IRFs, e.g., that they do not intersect or that they are single peaked, i.e., unimodal. Data analysis in this case is not aimed at estimating parameters that describe the shape of the IRF but at checking whether the data fulfil specific requirements that follow from these minimal assumptions. Of course, there are also parameters to be estimated in those models, e.g., the person parameters, and model tests follow the same scheme as with 'parametric' IRT models, i.e., investigating the discrepancies of observed and (under model assumptions) expected pattern frequencies.

The most prominent nonparametric IRT model is the Mokken approach to IRT (Mokken, 1971) which follows from the assumption that all IRFs are monotonously increasing and intersection free. (The less restrictive model omitting the latter assumption can be ignored in the present context.) This assumption, known as *double-monotonicity assumption,* can be formalized as follows

$$p(X_{vi}=1) \geq p(X_{wi}=1) \Rightarrow p(X_{vj}=1) \geq p(X_{wj}=1) \tag{23}$$

for all persons v and w and all items i and j, and

$$p(X_{vi}=1) \geq p(X_{vj}=1) \Rightarrow p(X_{wi}=1) \geq p(X_{wj}=1).$$

One of the testable consequences of this assumption is, e.g., the following implication for the joint probabilities of solving two items

$$p\left(X_i = 1 \wedge X_j = 1\right) \geq p\left(X_i = 1 \wedge X_k = 1\right) \Rightarrow p\left(X_m = 1 \wedge X_j = 1\right) \geq p\left(X_m = 1 \wedge X_k = 1\right), \tag{24}$$

for all items i, j, k and m, where $p(X_i = 1)$ means the probability that item i is solved in the given sample of persons. Since these probabilities can be estimated by the relative frequencies of 1-responses, the condition (24) for Mokken-scalability can be tested by inspecting the matrix of pairwise item frequencies (Mokken, 1971, p. 132 ff). Meijer, Sijtsma and Smid (1990) describe the relations to the Rasch model and Sijtsma, Debets and Molenaar (1989) presented the generalization of Mokken analysis for polytomous items. The chapter by Sijtsma and Junker (chapter 8) presents an extension of this method and an application of the model to reasoning tasks.

Ramsay (1991) extended the nonparametric IRT-approach by estimating the IRFs of the single items by means of the kernel smoothing technique. Using this method, no special kind of IRF has to be assumed for the items on a priori grounds, but the functions are identified from the data under consideration. In this sense, it is also a *parametric* method which is less restrictive than an ordinary IRT-model because the type of function (and not only its parameters) is a result of the data analysis and need not be the same for all items.

This method has also been generalized to polytomous item responses and is applied in chapter 24 by Kutylowski to a questionnaire on occupational prestige.

1.7 Models with single-peaked response functions

In the previous sections it has been assumed that the IRFs are monotone, i.e., the probability of a positive response, $p(X_{vi} = 1)$, increases with an increasing trait value. This is the typical assumption for achievement tests where the trait respresents an aptitude or ability. In attitude measurement, there is an alternative to monotonous IRFs, i.e., single-peaked response functions. The rationale of single-peaked IRFs goes back on Thurstones scaling methods, where the probability of agreeing with an attitude statement is highest when the trait value of a person is near to the location of the item (attitude statement) on the latent continuum. This probability decreases with the persons distance to the item in *both* directions.

IRT-models based on those single-peaked functions are sometimes called *unfolding-models* as a reference to the unfolding model for preference data by Coombs (1950).

Those unfolding models for item response data have been developed by Andrich (1988), Hoijtink (1990) and Post and Snijders (1993). Two groups (Andrich and Luo, 1993; Verhelst and Verstralen, 1993) have developed (independently from each other) an IRT-unfolding model, that is derived from the polytomous Rasch model (14), (see section 1.4). Starting from this model for three response categories, x = 0, 1 or 2, and collapsing the categories 0 and 2, i.e., adding their pobabilities defined by the IRF, gives the *hyperbolic cosine model* for dichotomous data

$$p\left(X_{vi} = 1\right) = \frac{\exp\left(\delta_i\right)}{\exp\left(\delta_i\right) + 2\cosh\left(\theta_v - \sigma_i\right)}, \tag{25}$$

where δ_i is a second item parameter representing (half of) the threshold distance in the original Rasch model (14) and cosh is the hyperbolic cosine function cosh(x) = (exp(x) + exp(–x))/2.

Andrich (1996) generalized the model for polytomous item responses, following the same idea of collapsing the response catgories symmetric to the midpoint of a rating scale with an odd number of categories. For example the response probabilities of a 4-point rating scale are assumed to be defined by the sum of each two response probabilities of a hypothetical ('unfolded') 7-point scale in the following way:

$$\begin{aligned} p(X_{vi}=0) &= p(Y_{vi}=0)+p(Y_{vi}=6) \\ p(X_{vi}=1) &= p(Y_{vi}=1)+p(Y_{vi}=5) \\ p(X_{vi}=2) &= p(Y_{vi}=2)+p(Y_{vi}=4), \quad \text{and} \\ p(X_{vi}=3) &= p(Y_{vi}=3). \end{aligned}$$

Whereas $p(X_{vi} = x)$ denote the probabilities of the observed item responses, $p(Y_{vi} = y)$ are the probabilities of an unobserved response variable Y_{vi}, that are defined by the polytomous Rasch model (14). From these assumptions the following equation of the *generalized hyperbolic cosine model* can be derived (Andrich 1996):

$$p(X_{vi}=x|x<m)=\frac{1}{d_{vi}}\exp(\kappa_{ix})\,2\cosh\left[(m-x)(\theta_v-\sigma_i)\right], \qquad (26)$$

where κ_{ix} is the partial sum of threshold parameters τ_{ix},

$$\kappa_{ix}=-\sum_{s=1}^{x}\tau_{is} \quad , \quad \kappa_{i0}=0$$

and d_{vi} is a normalizing factor. In chapter 26 (Rost and Luo) an application of a restricted version of model (26) is decribed.

Nonparametric IRT models have also been developed for non-monotonous single peaked IRFs (van Schuur, 1988; van Schuur and Post, 1991; Post and Snijders, 1993). The same kind of methods as for Mokken analysis are applied here in order to check whether the items can be ordered along a latent continuum. The difference is that the highest probability of giving a 1-response is not at one side of the continuum, but at some point on the continuum where the item is located. As a consequence (and different from Mokken scales) an ordering of individuals according to their sum score does not represent the ordering of individuals along the continuum. Thus, a first task of data analysis with nonparametric unfolding models is finding the order of items and persons.

The order of items can be found by inspecting all triples of items and analyzing the frequencies of the 8 possible response patterns for these items. The best order of three items is the one where the pattern 101 is observed least frequently. Starting with the best triple of items, further items are added subsequently in order to establish the complete order of items.

Chapter 15 by van Schuur describes this method as well as methods for testing the fit of the model and applies it to the measurement of work satisfaction, see also van Schuur (1993).

1.8 Two-dimensional latent trait models

All models treated in the previous sections were aimed at measuring only one latent trait, i.e., they are *unidimensional* models. The generalization to more than one dimension starts with the two-parameter logistic model (see section 1.3), which is reparameterized in the following way:

$$p\left(X_{vi} = x\right) = \frac{\exp\left(x\left(\sigma_i^* + \beta_i\,\theta_v\right)\right)}{1 + \exp\left(\sigma_i^* + \beta_i\,\theta_v\right)} \quad , x \in \{0,1\}, \tag{27}$$

where the item easiness parameter $\sigma_i^* = -\beta_i\,\sigma_i$ is a function of both item parameters in equation (10). The parameter β_i is a discrimination parameter that indicates how strong the response on item i is related to the latent dimension θ. In this sense, the β-parameters are comparable with factor loadings in factor analysis.

Introducing two independent latent variables, θ_1 and θ_2, the structure (27) generalizes to

$$p\left(X_{vi} = x\right) = \frac{\exp\left(x\left(\sigma_i + \beta_{i1}\,\theta_{v1} + \beta_{i2}\,\theta_{v2}\right)\right)}{1 + \exp\left(\sigma_i + \beta_{i1}\,\theta_{v1} + \beta_{i2}\,\theta_{v2}\right)} \quad , \tag{28}$$

where each item has three parameters, an item easiness parameter σ_i and two 'factor loadings' β_{i1} and β_{i2} (see Bartholomew, 1987).

In the contributions by Knott and Tzamourani (chapter 23) and Bartholomew, de Menezes and Tzamourani (chapter 21) both latent variables θ_1 and θ_2 are assumed to be independent standard normal variables, i.e., they have unit variance, zero mean, and are uncorrelated. The indeterminacy problem of factor analysis with all solutions obtained by rotation being equivalent also exists here. It can be solved as in factor analysis, i.e., by first fixing the parameters in some arbitrary way and then rotating the axes to an appropriate orientation.

The paper by Knott and Tzamourani (chapter 23) describes a procedure that enables fitting this model to empirical data when there are missing item responses. Bartholomew et al. (chapter 21) apply the model to two data sets, one on social life feelings and one on social attitudes, e.g., towards party politics, healthcare and crime.

The paper by Meiser and Rudinger (chapter 38) also deals with latent trait models where more than one dimension is involved (see Meiser, 1996). However, their models are quite different from the two-factor approach described above. They make use of the log-linear formulation of the polytomous Rasch model (see section 1.4) and specify a model where the probability of passing the first threshold depends on a different trait than passing the second threshold (see also Kelderman and Rijkes, 1994).

1.9 Latent trait models with covariates

In structural equation models like LISREL (Jöreskog and Sörbom, 1993) or EQS (Bentler and Wu, 1993) latent variables are embedded in a network of other, observed or latent variables. Causal relationships are modelled by a system of linear regression equations among endogenous and exogenous variables. Croon and Heinen (chapter 12) present a straightforward generalization of latent trait models in this direction. In contrast to the two-factor model described above, they consider a single latent variable θ with three different kinds of observed variables related to this trait.

First of all, the item responses X_{vi} aimed at measuring the latent trait are related to θ by means of the two-parameter logistic model (see section 1.3, equation (10)). Second, the trait variable θ itself depends on another set of observed variables, Y_j. It is assumed that this relationship can be described by a linear regression equation with weights α_j:

$$\theta_v^* = \sum_{j=1}^{p} \alpha_j Y_{vj}, \tag{29}$$

where θ_v^* denotes the expected value of θ_v. The third type of observed variables is a set of variables Z_k which are related to both the latent trait θ and the Y_j-variables. Again, a linear relationship is assumed so that for each Z_k the regression holds

$$Z_{vk}^* = \sum_{j=1}^{p} \delta_{jk} Y_{vj} + \zeta_k \theta_v, \tag{30}$$

where Z_{vk}^* is the expected value of Z_{vk}, δ_{jk} are regression weights of Z_k with respect to predictor Y_j, and ζ_k is the weight for the trait θ as a predictor of Z_k (see Figure 2).

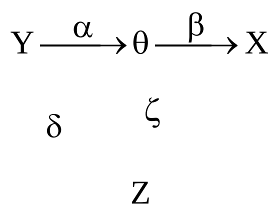

Figure 2: The structure of the Croon-Heinen model

Croon and Heinen (chapter 12) describe this model in more detail and provide an application to life satisfaction data.

1.10 Models with latent trait and latent state variables

The distinction between traits and states refers to the property of a latent variable to be more (traits) or less (states) stable over different occasions of measurement. The typical data structure *latent state trait (LST-)models* refer to, therefore, are responses of a set of persons on a set of items at a set of occasions, X_{vik}, where k refers to the k-th occasion. From this perspective the data structure is similar to that of the multi-facet Rasch model described in section 1.2. However, LST-models differ from the multi-facet Rasch model by introducing more than one individual variable reflecting the position of each individual on a number of trait and occasion-specific dimensions.

In chapter 13 Eid presents an application of a LST-model for polytomous, ordinal item responses. This model is based on Samejima's *graded response model* (Samejima, 1969), i.e., it makes use of the normal ogive for defining the response probabilities:

$$p\left(X_{vik} \geq x\right) = \int_{-\infty}^{\alpha_{ik} - \kappa_{ikx}} \frac{1}{\sqrt{2\pi}} \exp\left(-t^2 / 2\right) dt . \tag{31}$$

In this equation α_{ik} denotes a latent state variable that is additively decomposed into a trait variable θ_i for each item i and a situation-interaction variable ζ_k for each occasion:

$$\alpha_{ik} = \lambda_{ik}\,\theta_i + \delta_{ik}\,\zeta_k \quad . \tag{32}$$

The λ- and δ-parameters are the loadings of items-at-occasions on the two latent variables. The κ-parameters are category parameters that define the location of the category boundaries on the latent continuum. Like the threshold parameters in the polytomous Rasch model, but defined in a different way, they reflect the order of the response categories and provide information about the category sizes.

In the LST-model (Eid 1995, 1996) two item coefficients can be defined as variance components of the latent state variable. The *consistency coefficient*

$$\mathrm{Con}\left(\alpha_{ik}\right) = \frac{\lambda_{ik}^2 \mathrm{Var}\left(\theta_i\right)}{\mathrm{Var}\left(\alpha_{ik}\right)} = \frac{\lambda_{ik}^2 \mathrm{Var}\left(\theta_i\right)}{\lambda_{ik}^2 \mathrm{Var}\left(\theta_i\right) + \delta_{ik}^2 \mathrm{Var}\left(\zeta_k\right)} \tag{33}$$

indicates the degree to which true interindividual state differences are influenced by interindividual trait differences.

The *occasion specificity coefficient*

$$\mathrm{Spe}\left(\alpha_{ik}\right) = \frac{\delta_{ik}^2 \mathrm{Var}\left(\zeta_k\right)}{\mathrm{Var}\left(\alpha_{ik}\right)} = \frac{\delta_{ik}^2 \mathrm{Var}\left(\zeta_k\right)}{\lambda_{ik}^2 \mathrm{Var}\left(\theta_i\right) + \delta_{ik}^2 \mathrm{Var}\left(\zeta_k\right)} \tag{34}$$

on the other hand represents the degree to which true interindividual state differences depend on differences in the situations (and/or) interactions that have been realized on an occasion of measurement.

Items of a questionnaire for the assessment of enduring traits should have high consistency and low specificity coefficients. Items of a state mood questionnaire on the other hand should have high specificity and low consistency coefficients. Thus, this model can be used for the selection of state-like as well as trait-like items for psychometric questionnaires.

The model parameters can be estimated by means of structural equation models for ordinal variables (Jöreskog and Sörbom, 1993; Muthen, 1988). In chapter 13 Eid presents an application of the model to a questionnaire for the assessment of the momentary mood.

2. Identifying latent classes

All models described in this section have their origin in Lazarsfeld's (1950) latent class analysis (LCA, see also Lazarsfeld and Henry, 1968) and are in one way or the other extensions of the basic LC-model. While some of these models are strongly tied to the basic ideas of latent class analysis others try to extend the limitations of the basic model. Those staying within the limits of LCA are described in section 2.1, including models for ordinal data and the simultaneous analysis of different groups of individuals.

One basic principle of these models is *local independence,* i.e., the joint probability of several responses is the product of the marginal response probabilities given a latent class. Mixed Markov models drop the restriction of local independence as a Markov process is formalized in each latent class in order to describe the pattern of responses (which is typically a pattern of repeated responses at a series of time points). This approach is described in section 2.2.

Another common feature of latent class models is the assumption that all individuals within a particular class have identical response probabilities for the items included, i.e., individuals do not differ within latent classes. Models that take into account quantitative variation among individuals within each latent class are described in section 2.3.

Latent class membership may also be related to other observed variables that can be taken into account as covariates of the latent variable. This approach is described in section 2.4.

Usually it is assumed that the same type of model holds in each latent class, with the only difference of parameter values varying among classes. Section 2.5 refers to latent class models with different kinds of models assumed to hold within classes.

2.1 Classes with independent response behavior

The basic assumption of latent class (LC) analysis is local independence, which means that the probability of a response vector given that individual v belongs to class C_g is the product of class specific response probabilities $\pi_{ixg} := p(X_{vi} = x | v \in C_g)$:

$$p\left(\mathbf{x}_v \middle| v \in C_g\right) = \prod_{i=1}^{k} \pi_{ixg} \quad , x \in \{0,1,\ldots m\} \quad \text{with} \quad \sum_{x=0}^{m} \pi_{ixg} = 1. \tag{35}$$

The additional assumption that each individual belongs to class g of G classes with probability π_g gives the general structure of LC-models:

$$p\left(\mathbf{x}_v\right) = \sum_{g=1}^{G} \pi_g \prod_{i=1}^{k} \pi_{ixg} \quad \text{with} \quad \sum_{g=1}^{G} \pi_g = 1. \tag{36}$$

As can be derived from these equations, the probability of an individual v to belong to class g, given his/her response vector $\mathbf{x}_v$ is

$$p\left(g \middle| \mathbf{x}_v\right) = \frac{\pi_g \prod_{i=1}^{k} \pi_{ixg}}{\sum_{h=1}^{G} \pi_h \prod_{i=1}^{k} \pi_{ixh}}. \tag{37}$$

In order to obtain a manifest classification of all individuals similar to cluster analytic methods, each person is assigned to that class having the *highest* conditional probability (37) for his or her response vector.

Compared with the Rasch model as the basic model of latent trait analysis (see section 1.1), LC-models are *less restrictive* with respect to the assumption of person-homogeneity, i.e., one group (class) of persons can be characterized by a completely different set of item parameters (even with a different order) than another class of individuals. However, they are *more restrictive* in as much as all differences between persons have to be explained by a relative small number of classes of individuals. Whereas each individual may (in principle) have its own position on the latent continuum in latent trait models , all individuals of a latent class are treated as identical with respect to their response behavior.

Applications of LCA are presented by Dayton and Scheers (chapter 16, applied to survey data dealing with academic dishonesty), Keller and Kempf (chapter 30, applied to the Beck-Depression-Inventory), Kempf (chapter 22, applied to an analysis of German newspaper coverage of the Gulf war), Kohlmann and Formann (chapter 33, an analysis of epidemiological data on self-reported musculo-skeletal symptoms), Matschinger and Angermeyer (chapter 34, investigating artifacts in a panel study on symptoms of depression), and Uebersax (chapter 18, analyzing student problem behavior). Von Eye, Rovine and Spiel (chapter 10) discuss the relations of prediction analysis (PA) to LCA and apply PA to the prediction of school performance by intelligence.

The parameters of LC-models are usually estimated by means of the EM-algorithm which makes LCA a very flexible basic structure for deriving a variety of more specific models. The EM-algorithm has the attractive feature that many kinds of parameter restrictions, e.g., equality constraints or additive decompositions, can be imposed as long as maximum likelihood estimates of the restricted parameters are computed in each M-step.

This property enables, e.g., the *simultaneous analysis* of the latent class structure in different groups of persons. It can be done by introducing a manifest group variable as an additional item and fixing one and only one category probability π_{ixg} of this item at 1.0 in different classes. Members of a manifest group are thus forced into certain latent classes. Specific hypotheses about the equivalence of latent classes from different groups of persons can be tested by constraining the unconditional class probabilities and/or the response probabilities to be equal across groups (Clogg and Goodman, 1986).

Sorenson, Brownfield and Jensen (chapter 17) present an application of simultaneous LCA for constructing a measure of delinquency, McCutcheon and Hagenaars (chapter 25) apply this method in comparative social survey studies, v.d. Heijden, 't Hart and Dessens (chapter 19) to a questionnaire on anti-social behaviour, and v.d. Pol (chapter 40) uses multiple group LCA to cluster educational mobility tables defined by birth cohorts and gender.

Another way of making use of this flexibility as a result of parameter restrictions is the specification of *latent distance* and response error models. V.d. Wittenboer, Hox, and de

Leeuw (chapter 14) specify such models in order to analyze the scalability of elderly respondents with respect to a questionnaire on living arrangements and social networks.

Linear constraints can hardly be imposed on the parameters π_{ixg} of the model given by equation (36) because these parameters are probabilities and thus restricted to the [0,1]-interval. However, it is easy to introduce linear constraints in the logistic version of LCA

$$p(\mathbf{x}_v) = \sum_{g=1}^{G} \pi_g \prod_{i=1}^{k} \frac{\exp(\alpha_{ixg})}{\sum_{s=0}^{m} \exp(\alpha_{isg})}, \tag{38}$$

which is an equivalent reparameterization of (36).

Formann (1992) discusses the general case of an additive decomposition of the logistic parameters α_{ixg}, whereas Rost (1988 a, b) presents decompositions that take into account the order of the response categories. These models are analogs of polytomous Rasch models for ordinal data (see section 1.4) and specify, e.g.,

$$\alpha_{ixg} = x\,\theta_{ig} + \sum_{s=1}^{x} \tau_{is} \tag{39}$$

as an analog of the partial credit model (14). Again, the τ_{is}-parameters define the threshold location on a hypothetical continuum and should be ordered if the response categories are ordered. In contast to the corresponding latent trait model single individuals are not located on the continuum but only the ability of class g with respect to item i, represented by θ_{ig}.

Specific hypotheses about the threshold distances in particular classes or items can be tested by means of other linear constraints, e.g.

$$\alpha_{ixg} = x\,\theta_{ig} + \sum_{s=1}^{x} \psi_{sg} \tag{40}$$

which assume equal threshold distances for all items (like the rating scale model (17)) that are, however, different from class to class.

These LC-models for ordinal data are applied by Frick, Rehm and Thien (chapter 29) to the Beck Depression Inventory, and by Tarnai and Wuggenig (chapter 27) to a comparative study on aesthetic attitudes and lifestyles.

2.2 Classes of Markov models

Markov models are used to describe categorical panel data, e.g., repeated responses to one or more items at two or more time points. In order to introduce the typical notation for these models, only the case of repeated observations of one item at 4 time points is considered. An ordinary Markov chain describes the frequencies of the response patterns $\mathbf{x} = (x_i,x_j,x_k,x_l)$ in the following way

$$p(x_i, x_j, x_k, x_l) = \delta_i^1\, \tau_{j|i}^{21}\, \tau_{k|j}^{32}\, \tau_{l|k}^{43} \quad , \tag{41}$$

where the δ's are unconditional probabilities for time point t = 1 and the τ's are conditional or *transition probabilities* at time t + 1, given a particular category at time t (with subscripts referring to response categories and superscripts referring to time points).

The relationship to LC-models becomes obvious when the model is generalized to a *mixed Markov model*, where it is assumed that several, i.e., G Markov chains with different sets of parameters describe the data in different unknown subpopulations. Hence, the pattern probability is

$$p\left(x_i, x_j, x_k, x_l\right) = \sum_{g=1}^{G} \pi_g \, \delta^{1}_{i|g} \, \tau^{21}_{j|ig} \, \tau^{32}_{k|jg} \, \tau^{43}_{l|kg} \tag{42}$$

where all parameters, i.e., the initial probabilities $\delta^1_{i|g}$ and the transition probabilities are class or chain specific parameters. The relation to LCA becomes obvious by writing the LC-model (36) in the notation used in this section:

$$p\left(x_i, x_j, x_k, x_l\right) = \sum_{g=1}^{G} \pi_g \, \delta^{1}_{i|g} \, \tau^{2}_{j|g} \, \tau^{3}_{k|g} \, \tau^{4}_{l|g}. \tag{43}$$

Model (42) reduces to model (43) if the probability of being in some category at time point t+1 is independent of which category a person did belong to at the most recent point in time t.

The latent class model (43) may thus be considered as a special case of the mixed Markov model (42) with all transition probabilities assumed to be independent of the responses at the preceding time point. In other words, the mixed Markov model substitutes the assumption of local independence by the less restrictive assumption of first order serial dependency. While mixed Markov models operate on the latent level they may still be characterized as mixtures of G manifest Markov chains, that is, they assume the data to be free of measurement error, an assumption that is rather unrealistic in the social sciences. The latent Markov model (Wiggins, 1955).

$$p\left(x_i, x_j, x_k, x_l\right) = \sum_{abcd} \delta^{1}_{a} \, \rho^{1}_{i|a} \, \tau^{21}_{b|a} \, \rho^{2}_{j|b} \, \tau^{32}_{c|b} \, \rho^{3}_{k|c} \, \tau^{43}_{d|c} \, \rho^{4}_{l|d} \tag{44}$$

therefore assumes a latent distribution at each point in time that is measured by a fallible indicator, the manifest response at time t. Contrary to the ordinary Markov model (41), model (44) thus starts with latent unconditional probabilities (δ^1_a) at t = 1, with the manifest variable mapped onto the latent one by conditional response probabilites $\rho^1_{i|a}$. As a consequence, transition probabilities, the τ's, operate at the latent level now as well. As an aside, note that the LC-model (36) is a special case of model (44) as well if model (44) is constrained to be a no-change model (i.e., by setting all matrices of transition probabilities equal to the identity matrix).

The model has been further generalized to the latent mixed Markov (Langeheine and v.d. Pol, 1990), that contains both models (42) and (44) as special cases. Finally, extensions to multiple group analysis (v.d. Pol and Langeheine, 1990) and to multiple indicator models (Langeheine and v.d. Pol 1993, 1994) have been presented.

Multiple indicator latent Markov models are applied by Meiser and Rudinger (chapter 38) to data on the development of deductive reasoning and by Reinecke (chapter 39) to a study on AIDS-prevention behavior. The focus of both Böckenholt's (chapter 35) and Humphreys' chapter (36) is on latent change as well, that, in addition, is assumed to depend on external, concomitant variables. While Böckenholt presents an analysis of a consumer panel study and change in choice behavior, Humphreys shows how estimation of female labor force participation flows may be improved by postulating regressions of both latent transition probabilities and classification (response) error probabilities on observed covariates and unobserved random effects.

2.3 Classes of Rasch-homogeneous persons

One drawback of ordinary latent class models is the very restrictive assumption that all individuals of a given class are identical with respect to their response probabilities. In comparison with latent trait models, where each person may be 'slightly different' from every other person, LCA allows for dramatic differences *between* classes but for no inter-individual difference at all *within* classes. As a consequence, a large number of classes is often needed in order to adequately represent the quantitative variation when analyzing, e.g., achievement or attitude items.

The feature of *mixed Rasch models* is to extend the Rasch model to a mixture of latent classes, with the Rasch model assumed to hold within each of these classes but with different item parameters among classes. This model structure allows to represent quantitative differences among subjects *within* a class by means of the ability parameter of the Rasch model. Differences *between* classes, on the other hand, such as different item difficulties or even a different order of item difficulties, are qualitative (Rost, 1990).

The mixed Rasch model for polytomous ordinal data (Rost 1991) is defined by

$$p(\mathbf{x}_v) = \sum_{g=1}^{G} \pi_g \prod_{i=1}^{k} \frac{\exp\left(x\theta_{vg} - \sum_{s=1}^{x} \tau_{isg}\right)}{\sum_{y=0}^{m} \exp\left(y\theta_{vg} - \sum_{s=1}^{y} \tau_{isg}\right)} \quad , x \in \{0,1\ldots m\}, \tag{45}$$

i.e., the class specific response probabilities π_{ixg} of the LC-model (36) are modelled by the polytomous Rasch model (14) with class specific ability and threshold parameters.

The class specific ability parameters θ_{vg} can be conditioned out (cf. equation (4)) resulting in a model equation where the probabilities of a particular score r conditional on class g, π_{rg}, have to be considered as new model parameters:

$$p(\mathbf{x}_v) = \sum_{g=1}^{G} \pi_g \, \pi_{rg} \frac{\exp\left(-\sum_{i=1}^{k} \sigma_{ixg}\right)}{\gamma_r\left(\exp\left(-\boldsymbol{\sigma}_g\right)\right)}. \tag{46}$$

The σ-parameters are the cumulated threshold parameters $\sigma_{ixg} = \sum_{s=1}^{x} \tau_{isg}$ (cf. also equation (16)) and γ_r denotes the symmetric functions of order r of the de-logarithmized item parameters (v. Davier and Rost, 1995). After the estimation of the model parameters via an extended EM-algorithm, the ability parameters θ_{vg} can be estimated using unconditional maximum likelihood methods (Rost, 1996a; v. Davier 1994).

The mixed Rasch model is applied by Meiser and Rudinger (chapter 38) to the analysis of changes in a personality dimension of activity and by Rost et al.(chapter 31) to the analysis of a questionnaire on the „big five“ personality traits. v. Davier and Rost (chapter 28) present results of an analysis of the self-monitoring construct. In chapter 9 (Spiel et al.) the model is applied to test the unidimensionality of cognitive tasks.

2.4 Classes and their covariates

Within the framework of latent trait theory, Croon and Heinen (section 1.9) have proposed a model, where the latent trait itself depends on one or more manifest variables, i.e., θ is some linear function of a set of observed variables Y_j (cf. (29)). In LC-models, the latent variable is discrete and, hence, cannot be a continuous function of a real-valued covariate. Dayton and Macready (1988) discussed a model, where the unconditional class probabilities π_g (see (36)) are a logistic function of the *covariates* Y_j:

$$\pi^*_{g|y} = \frac{\exp\left(\sum_{j=1}^{p} \alpha_{jg} Y_{vj}\right)}{1 + \exp\left(\sum_{j=1}^{p} \alpha_{jg} Y_{vj}\right)} \tag{47}$$

and α_{jg} are the weights of the j-th covariate with respect to the size of class g. By appropriate restrictions it has to be assured, however, that the predicted class sizes $\pi^*_{g|y}$ add up to unity. The unconditional pattern probability then is the integral:

$$p(\mathbf{x}) = \int_a^b \sum_{g=1}^{G} \pi^*_{g|y} \prod_{i=1}^{k} \pi_{ixg} \, dY, \tag{48}$$

when Y is a covariate defined in the interval [a, b].

In this extended LC-model, only the class size parameters π_g, but not the class specific response probabilities π_{ixg}, are a function of the covariate. Böckenholt (chapter 35) generalizes this approach by introducing dependencies of latent change probabilities on the covariates and Humphreys (chapter 36) does so by postulating dependencies of latent transition probabilities and response error probabilities on the covariates (see also section 2.1).

2.5 Mixtures of different models

LC-models usually assume that the same type of model holds for all latent classes. In some applications, however, different response mechanisms are assumed to operate in different classes.

Simple examples are models where one special observation, i.e., the (0,0,0,..,0)-pattern of item responses, or the 'zero' of count data, is treated differently from the rest of the observations. Such observations may be allocated to one class exclusively in order to not disturb the distribution of the entire set of observations.

Böhning, Dietz and Schlattmann (chapter 32) describe such a model for count data and apply it to data from epidemiological research, where zero counts are over-represented (inflated).

Another example is the so-called *hybrid-model* of Gitomer and Yamamoto (1991) where it is assumed that a latent trait model fits the data in one latent class, whereas the multinomial independence model fits the data in one or more other classes. The general structure of these hybrid models is:

$$p(\mathbf{x}_v) = \pi_1 \prod_{i=1}^{k} \frac{\exp\left(x\,\beta_i(\theta_v - \sigma_i)\right)}{1 + \exp\left(\beta_i(\theta_v - \sigma_i)\right)} + \sum_{g=2}^{G} \pi_g\, \pi_{ixg} \quad , x \in \{0,1\} \tag{49}$$

if the two parameter logistic model (10) is assumed to hold within the first latent class.

Yamamoto and Everson (chapter 7) apply a variant of this model to data from a speeded test in order to take into account the different lengths of the response vectors produced in a given time limit.

V. Davier and Rost (chapter 28) describe an application of a hybrid Rasch-LC model to data from a self-monitoring questionnaire in order to investigate whether self-monitoring is a trait or a type construct. The hybrid Rasch-LC model is equivalent to model (49) where all $\beta_i = 1$, but it can also be generalized to polytomous ordinal data and any number of Rasch-classes:

$$p(\mathbf{x}_v) = \sum_{g=1}^{G} \pi_g\, \pi_{rg} \frac{\exp\left(-\sum_{i=1}^{k} \sigma_{ixg}\right)}{\gamma_r\left(\exp(\sigma_g)\right)} + \sum_{h=1}^{H} \pi_h\, \pi_{ixh} \quad , x \in \{0,1\ldots m\}. \tag{50}$$

This quick trip through the world of latent structure models does not claim to be a complete overview of all trait and class models that are worth mentioning in a volume like this one. However, it is a guide through the spectrum of models that are applied and discussed in the following chapters. For overviews of latent trait, latent class or both types of models see also Clogg (1995), Fischer and Molenaar (1995), Formann (1984), Kubinger (1989), Langeheine and Rost (1988), Rost (1996b) and v.d. Linden and Hambleton (1996).

References

Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika* , *43,* 4, 561-573.

Andrich, D. (1982). An extension of the Rasch model for ratings providing both location and dispersion parameters. *Psychometrika*, *47,* 1, 105-113.

Andrich, D. (1988). The application of an unfolding model of the PIRT type to the measurement of attitude. *Applied Psychological Measurement*, *12*, 33-51.

Andrich, D. (1996). A hyperbolic cosine latent trait model for unfolding polytomous responses: Reconciling Thurstone and Likert methodologies. *British Journal of Mathematical and Statistical Psychology* (in press).

Andrich, D. & Luo, G. (1993). A hyperbolic cosine latent trait model for unfolding dichotomous single-stimulus responses. *Applied Psychological Measurement*, *17*, 253-276.

Bartholomew, D.J. (1987). *Latent variable models and factor analysis*. London: Charles Griffin & Company LTD.

Bentler, P.M. & Wu, E.J.C. (1993). *EQS/Windows user's guide*. Los Angeles: BMDP Statistical Software.

Clogg, C.C. (1995). Latent class models: Recent developments and prospects for the future. In G. Arminger, C.C. Clogg & M.E. Sobel (eds.), *Handbook of statistical modeling in the social sciences* (pp. 311-359). New York: Plenum.

Clogg, C.C. & Goodman, L.A. (1986). On scaling models applied to data from several groups. *Psychometrika*, *51*, 123-135.

Coombs, C.H. (1950). Psychological scaling without a unit of measurement. *Psychological Review*, *57*, 145-158.

Croon, M. (1991). Investigating Mokken scalability of dichotomous items by means of ordinal latent class analysis. *British Journal of Mathematical and Statistical Psychology*, *44*, 315-331.

Davier, von, M. (1994). *WINMIRA - A Program System for Analyses with the Rasch Model, with the Latent Class Analysis and with the Mixed Rasch Model . User Manual*. Kiel: IPN.

Davier, von, & Rost, J. (1995). Polytomous mixed Rasch models. In G. Fischer & I. Molenaar (eds.), *Rasch models: Foundations, recent developments, and applications* (pp. 371-379). Berlin: Springer.

Dayton, M. & Macready, G.B. (1988). Concomitant-variable latent class models. *Journal of the American Statistical Association*, *83*, 173-178.

Eid, M. (1995). *Modelle der Messung von Personen in Situationen (models for the measurement of persons in situations)*. Weinheim: Beltz - Psychologie Verlags Union.

Eid, M. (1996). Longitudinal confirmatory factor analysis for polytomous item responses: Model definition and model selection on the basis of stochastic measurement theory. *Methods of Psycholocical Research - online (http://www.hsp.de/MPR/)*, *1*,1.

Fischer, G.H. (1972). A measurement model for the effect of mass-media. *Acta Psychologica*, *36*, 207-220.

Fischer, G.H. (1973). The linear logistic test model as an instrument in educational research. *Acta Psychologica*, *37,* 359-374.

Fischer, G.H. (1974). *Einführung in die Theorie psychologischer Tests*. Bern: Huber.

Fischer, G.H. (1983). Logistic latent trait models with linear constraints. *Psychometrika*, *48,* 3-26.

Fischer, G.H. (1987). Applying the principles of specific objectivity and generalizability to the measurement of change. *Psychometrika*, *52*, 565-587.

Fischer, G.H. & Molenaar, I.W. (1995). *Rasch models - Foundations, recent developments, and applications*. New York: Springer.

Formann, A.K. (1984). *Die Latent-Class-Analyse*. Weinheim: Beltz.

Formann, A.K. (1992). Linear logistic latent class analysis for polytomous data. *Journal of the American Statistical Association*, *87*, 476-486.

Gitomer, D.H. & Yamamoto, K. (1991). Performance modeling that integrates latent trait and class theory. *Journal of Educational Measurement*, *28,* 2, 173-189.

Heinen, T. (1993). *Discrete Latent Variable Models*. Tilburg: University Press.

Hoijtink, H. (1990). A latent trait model for dichotomous choice data. *Psychometrika*, *55*, 641-656.

Jöreskog, K.G. & Sörbom, D. (1993). *LISREL 8. A guide to the program and its applications*. Chicago: SPSS.

Kelderman, H. (1984). Log linear Rasch model tests. *Psychometrika*, *49,* 223-245.

Kelderman, H. & Rijkes, C.P.M. (1994). Loglinear multidimensional IRT models for polytomously scored items. *Psychometrika*, *59,* 149-176.

Kubinger, K.D. (1988). *Moderne Testtheorie*. Weinheim: PVU.

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

Langeheine, R. & Rost, J. (1993). Latent Class Analyse. *Psychologische Beiträge*, *35*, 177-198.

Langeheine, R. & van de Pol, F. (1990). A unifying framework for Markov modeling in discrete space and discrete time. *Sociological Methods & Research, 18,* 4, 416-441.

Langeheine, R. & Van de Pol, F. (1993). Multiple indicator Markov models. In R. Steyer, K.F. Wender & K.F. Widaman (eds.), *Psychometric methodology. Proceedings of the 7th European Meeting of the Psychometric Society in Trier.* (pp. 248-252). Stuttgart: Fischer.

Langeheine, R. & Van de Pol, F. (1994). Discrete time mixed Markov latent class models. In A. Dale & R. Davies (eds.), *Analyzing social and political change: A casebook of methods* (pp. 170-197). Newbury Park, CA: Sage.

Lazarsfeld, P.F. (1950). The logical and mathematical foundation of latent structure analysis. In S.A. Stouffer, L. Guttman, E.A. Suchman, P.F. Lazarsfeld, S.A. Star, J.A. Clausen (eds.), *Studies in social psychology in world war II, Vol. IV* (pp. 362-412). Princeton/N.J.: Princeton Univ. Press.

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

Linacre, J.M. (1989). *Many-faceted Rasch measurement*. Chicago: MESA Press.

Lord, F.M. & Novick, M.R. (1968). *Statistical theories of mental test scores*. Reading, Mass.: Addison-Wesley.

Masters, G.N. (1982). A Rasch model for partial credit scoring. *Psychometrika, 47,* 149-174.

Masters, G.N. & Wright, B.D. (1984). The essential process in a family of measurement models. *Psychometrika, 49,* 4, 529-544.

Meijer, R.R., Sijtsma, K., Smid, N.G. (1990). Theoretical and empirical comparison of the Mokken and the Rasch approach to IRT. *Applied Psychological Measurement, 14,* 3, 283-298.

Meiser, T. Loglinear Rasch models for the analysis of stability and change. *Psychometrika* (in press).

Micko, H.-C. (1970). Eine Verallgemeinerung des Meßmodells von Rasch mit einer Anwendung auf die Psychophysik der Reaktionen. *Psychologische Beiträge, 12*, 4-22.

Mokken, R.J. (1971). *A theory and procedure of scale analysis*. The Hague: Mouton.

Muthen, B. (1988). *LISCOMP: Analysis of linear sturctural equations with a comprehensive measurement model (2nd ed.).*

Neymann, J. & Scott, E.L. (1948). Consistent estimates based on partially consistent obervations. *Econometrica, 16,* 1, 1-32.

Post, W.J. & Snijders, T.A.B. (1993). Nonparametric unfolding models for dichotomous data. *Methodika, 7,* 130-156.

Ramsay, J.O. (1991). Kernel smoothing approaches to nonparametric item characteristic curve estimation. *Psychometrika, 56,* 611-630.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Nielsen & Lydiche (2nd Edition, Chicago, University of Chicago Press, 1980).

Rost, J. (1988a). Rating scale analysis with latent class models. *Psychometrika, 53,* 3, 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*. New York: Plenum.

Rost, J. (1990). Rasch models in latent classes: An integration of two approaches to item analysis. *Applied Psychological Measurement, 14,* 3, 271-282.

Rost, J. (1991). A logistic mixture distribution model for polychotomous item responses. *The British Journal of Mathematical and Statistical Psychology, 44,* 75-92.

Rost, J. (1996a). Logistic mixture models. In W. van der Linden & R. Hambleton, *Handbook of modern item response theory*. Berlin: Springer (in press).

Rost, J. (1996b). *Lehrbuch Testtheorie, Testkonstruktion*. Bern: Huber.

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Special Monograph, Monograph Supplement No.17.*

Sijtsma, K., Debets, P. & Molenaar, I.W. (1989). Mokken scale analysis for polychotomous items: Theory, a computer program and an empirical application. *Quality and Quantity, 24,* 173-188.

Tjur, T. (1982). A connection between Rasch's item analysis model and a multiplicative Poisson model *Scandinavian Journal of Statistics*, *9*, 23-30.

Van de Pol, F. & Langeheine, R. (1990). Mixed Markov latent class models. In C.C. Clogg (ed.), *Sociological Methodology, 1990* (pp. 213-247). Oxford: Blackwell.

Van der Linden, W. & Hambleton, R. (1996). *Handbook of modern item response theory*. Berlin: Springer.

Van Schuur, W.H. (1988). Stochastic unfolding. In W.E. Saris & I.N. Gallhofer (eds.), *Sociometric Research, vol.1: Data collection and scaling*. London: MacMillan.

Van Schuur, W.H. (1993). Nonparametric unidimensional unfolding for multicategory data. *Political Analysis*, *4,* 41-74.

Verhelst, N.D. & Verstralen, H.H.F.M. (1993). A stochastic unfolding model derived from the partial credit model *Kwantitatieve Methoden*, *42*, 73-92.

Wiggins, L.M. (1955). Mathematical Models for the Analysis of Multi-wave Panels (PhD Dissertation. Columbia University). Ann Arbour: University Micorfilms.

Wright, B.D. & Masters, G.N. (1982). *Rating scale analysis: Rasch measurement*. Chicago: Mesa Press.