Technical Appendix for Latent GOLD 3.0

Jeroen K. Vermunt

Jay Magidson

This appendix describes the various types of latent class (LC) models implemented in Latent GOLD 3.0. First, we present the general Latent GOLD model. Then, attention is paid to the three special cases "LC cluster", "LC factor", and "LC regression" that have been implemented as separate modules in the program. Subsequently, we describe estimation procedures and the corresponding technical options. The output provided by the various modules is described in the last section.

1 General Latent GOLD Model

The specification of a LC model typically consists of three steps:

1. setting up a probability structure using certain assumptions regarding conditional independence,
2. specifying distributional forms for the dependent variables,
3. specifying regression-type constraints to gain parsimony.

Latent GOLD automatically sets up the correct probability structure for each of the three special cases. Before discussing these special cases, we describe the general probability structure, the available distribution functions for the latent and dependent variables, and the corresponding regression-type constraints.

1.1 Probability Structure

Latent GOLD implements three types of LC models: LC cluster, LC factor, and LC regression. Each of these models is based on the same general probability structure which defines the relationships between three types of variables:

1. J latent (unobserved) variables denoted by $x_1, x_2, \ldots, x_J$. When referring to the entire set of x variables, we will use the symbol $\mathbf{x}$. The latent variables may be treated as nominal or ordinal (discrete interval). Latent GOLD models must contain at least one latent variable.

2. K indicators or dependent variables[1] denoted by $y_1, y_2, \ldots, y_K$. The symbol $\mathbf{y}$ is used to denote the entire set of dependent variables. We will also use the symbol $\mathbf{y}_m$ to denote one of the M subsets of y variables, and K_m to denote the number of variables in set m. Dependent variables may be nominal, ordinal, continuous, or counts. Users must assign at least one variable to be included in the model as a y variable.

3. L covariates or independent variables[2] denoted by $z_{1,} z_{2,} \ldots, z_L$. The bold face symbol $\mathbf{z}$ is used to refer to the entire set of independent variables. Covariates may be nominal or numeric variables. Numeric covariates may be ordinal, discrete interval, or continuous; that is, the scale type numeric refers to situations in which we treat observed scores as quantities rather than only as category labels. Latent GOLD models can be specified without any covariates. In addition, it is possible to have covariates which are inactive. These are not part of the model, but are used to obtain descriptive information about the latent classes.

The simple but very general probability structure that is assumed for these variables is of the form:

$$f(\mathbf{y}|\mathbf{z}) = \sum_{\mathbf{x}} \pi(\mathbf{x}|\mathbf{z})\, f(\mathbf{y}|\mathbf{x}, \mathbf{z}) = \sum_{\mathbf{x}} \pi(\mathbf{x}|\mathbf{z}) \prod_{m=1}^{M} f(\mathbf{y}_m|\mathbf{x}, \mathbf{z}) \tag{1}$$

As can be seen, we are interested in modeling the probability density of observing a particular set of $\mathbf{y}$ values given a particular set of $\mathbf{z}$ values, that is, $f(\mathbf{y}|\mathbf{z})$. The second part of equation (1) shows that the unobserved variables $\mathbf{x}$ intervene between the $\mathbf{z}$ and the $\mathbf{y}$ variables. Here, $\pi(\mathbf{x}|\mathbf{z})$ is the probability of having a certain set of values on the latent variables given an individual's realized covariate values, and $f(\mathbf{y}|\mathbf{x}, \mathbf{z})$ is the probability density

[1]Other terms that may be used to describe the y's are response variables, outcome variables, outputs, endogenous variables, and items.

[2]Alternative terms for the z's are concomitant variables, grouping variables, external variables, exogenous variables, predictors, and inputs.

of $\mathbf{y}$ given $\mathbf{x}$ and $\mathbf{z}$. Thus, x variables may be influenced by z variables, and y variables may be influenced by x and z variables.

The last part of the model formulation described in equation (1) implies that y variables belonging to different sets are assumed to be mutually independent given the latent variables and the covariates:

$$f(\mathbf{y}|\mathbf{x},\mathbf{z}) = \prod_{m=1}^{M} f(\mathbf{y}_m|\mathbf{x},\mathbf{z}).$$

On the other hand, it is also important to note that the y's belonging to the same set m may be dependent of one another given the x's and the z's.

1.2 Distributions and Regression-type Constraints

Depending on the scale types of the variables in a set, a particular distributional form is assumed for $\mathbf{y}_m$. A set may consist of one or more categorical (nominal or ordinal) variables, one or more continuous variables, or a single variable containing counts. When the variables are categorical, a multinomial distribution is assumed for $\mathbf{y}_m$. For continuous variables, we use (multivariate) normal distributions. Counts can be modeled via a Poisson or a binomial distribution.

As explained in more detail below, in most cases, restrictions will be imposed on the distributions $\pi(\mathbf{x}|\mathbf{z})$ and $f(\mathbf{y}_m|\mathbf{x},\mathbf{z})$ in the form of regression-type constraints. For each distribution, we define a linear predictor that can be restricted by a simple regression model. Within the context of generalized linear modeling, the assumed distribution function is denoted as the error function and the transformation yielding the linear predictor as the link function (McCullagh and Nelder, 1983). Since these regression-type of constraints are somewhat different for LC cluster, LC factor, and LC regression models, they are discussed in more detail in the next sections.

1.2.1 Latent Variables

The scores on the latent variables given the covariate values are assumed to come from a (joint) multinomial distribution. This was already reflected by the fact that in equation (1) we used the symbol π when referring to the probability of belonging to a certain latent class. The multinomial probability

$\pi(\mathbf{x}|\mathbf{z})$ is parameterized as follows:

$$\pi(\mathbf{x}|\mathbf{z}) = \frac{\exp(\eta_{\mathbf{x}|\mathbf{z}})}{\sum_{\mathbf{x}} \exp(\eta_{\mathbf{x}|\mathbf{z}})};$$

that is, by means of a logit or logistic regression model. The linear term $\eta_{\mathbf{x}|\mathbf{z}}$ equals

$$\eta_{\mathbf{x}|\mathbf{z}} = \sum_{j=1}^{J} \gamma^0_{x_j} + \sum_{j=1}^{J} \sum_{j'=j+1}^{J} \gamma^0_{x_j x_{j'}} + \sum_{\ell=1}^{L} \sum_{j=1}^{J} \gamma^1_{z_\ell x_j}.$$

As can be seen, the linear term contains the main effects for all latent variables ($\gamma^0_{x_j}$), their two-way interactions ($\gamma^0_{x_j x_{j'}}$), and the covariate effects on the latent variables ($\gamma^1_{z_\ell x_j}$). In the default setting, the bivariate associations between the latent variables are excluded from the model: $\gamma^0_{x_j x_{j'}} = 0$. Note that higher-order interactions are not included in the model.

When the latent variables are ordinal (or discrete interval) variables, as in a LC factor model, the $\gamma^0_{x_j x_{j'}}$ and $\gamma^1_{z_\ell x_j}$ terms are constrained by using fixed equal-interval category scores for the latent variables. For example, $\gamma^0_{x_j x_{j'}} = \gamma^0_{jj'} \cdot v_{x_j} \cdot v_{x_{j'}}$, where v_{x_j} denotes the fixed category for x_j. Such a restricted interaction term is sometimes referred to as a uniform or linear-by-linear association (Goodman, 1979; Agresti, 2002).

When a covariate is numeric (ordinal, discrete interval, or continuous), the corresponding $\gamma^1_{z_\ell x_j}$ term is restricted by using the observed scores z_ℓ in the two-way association term. This yields $\gamma^1_{z_\ell x_j} = \gamma^1_{\ell x_j} \cdot z_\ell$; that is, a standard effect term of a logistic regression model with continuous covariates.

For identification purposes, effect or ANOVA-like coding is used for the nominal variables, which means that the parameters are restricted to sum to zero over the categories of any nominal variable.

1.2.2 Nominal and Ordinal Dependent Variables

Categorical (nominal or ordinal) dependent variables are assumed to come from a multinomial distribution. Let us first have a look at the situation in which each set m consist of a single dependent or indicator variable: $M = K$ and $m = k$. In other words, we postulate that the local independence assumption holds. In that case, the distribution for each of the K indicators

y_k is of the form

$$f(y_k|\mathbf{xz}) = \pi(y_k|\mathbf{xz}) = \frac{\exp(\eta_{y_k|\mathbf{xz}})}{\sum_{y_k} \exp(\eta_{y_k|\mathbf{xz}})}.$$

To show the difference between the treatment of nominal and ordinal variables, we will consider a more concrete case. Suppose that we have a LC cluster model, which is a LC model with a single nominal latent variable x_1, with two nominal covariates z_1 and z_2. If the dependent variable y_k is nominal, the linear predictor will be

$$\eta_{y_k|x_1 z_1 z_2} = \beta^0_{y_k} + \beta^1_{x_1 y_k} + \beta^2_{z_1 y_k} + \beta^2_{z_2 y_k},$$

where the parameters are restricted to sum to zero over all their indices. What can be seen is that all three- or more-variable terms are excluded from the model. Note that in most applications of LC cluster models, we will assume that covariates do not have direct effects on the indicators; that is, $\beta^2_{z_1 y_k} = \beta^2_{z_2 y_k} = 0$.

When a dependent or indicator variable y_k is ordinal, the regression terms are further restricted using fixed scores v_{y_k} for the categories of y_k. This yields

$$\eta_{y_k|x_1 z_1 z_2} = \beta^0_{y_k} + \beta^1_{x_1 k} \cdot v_{y_k} + \beta^2_{z_1 k} \cdot v_{y_k} + \beta^2_{z_2 k} \cdot v_{y_k}.$$

The two-way interaction terms now have the structure of row associations (Goodman, 1979; Agresti, 2002). Situations in which the latent variables are ordinal are discussed in more detail in the section on LC factor analysis.

So far, we assumed that each set m consists of a single nominal or ordinal categorical variable. However, in models with local dependencies, some sets will contain more than one indicator. For sets containing more than one categorical indicator, we also have a multinomial distribution; that is,

$$f(\mathbf{y}_m|\mathbf{xz}) = \pi(\mathbf{y}_m|\mathbf{xz}) = \frac{\exp(\eta_{\mathbf{y}_m|\mathbf{xz}})}{\sum_{\mathbf{y}_m} \exp(\eta_{\mathbf{y}_m|\mathbf{xz}})}.$$

The effects of the x's and the z's on the separate y's belonging to set m are the same as in the local independence case described above. What differs is that we now have to specify the direct relationships between the y variables of set m. These local dependencies are included in the model via two-variable terms appearing in the linear predictor.

Suppose that we have a set of three nominal variables y_1, y_2, and y_3 in a model with a single nominal latent variable x_1 and without covariates. In that case, the linear predictor will be

$$\begin{aligned}\eta_{y_1 y_2 y_3|x_1} &= \sum_{k=1}^{3} \beta^0_{y_k} + \sum_{k=1}^{3} \sum_{k'=k+1}^{3} \beta^0_{y_k y_{k'}} + \sum_{k=1}^{3} \beta^1_{x_1 y_k} \\ &= \beta^0_{y_1} + \beta^0_{y_2} + \beta^0_{y_3} + \beta^0_{y_1 y_2} + \beta^0_{y_1 y_3} + \beta^0_{y_2 y_3} + \beta^1_{x_1 y_1} + \beta^1_{x_1 y_2} + \beta^1_{x_1 y_3}.\end{aligned}$$

The linear predictor contains, besides the three main effects and the three effects of the latent variable on the indicators, three direct effects between indicators. Note that it is not necessary to include all $\beta^0_{y_k y_{k'}}$ terms that can be defined in a particular set: From the above model we could, for instance, exclude $\beta^0_{y_2 y_3}$.

1.2.3 Continuous Dependent Variables

Continuous dependent variables are modeled by means of multivariate normal distributions; that is,

$$\begin{aligned}f(\mathbf{y}_m|\mathbf{xz}) &= (2\pi)^{-K_m/2} \left|\boldsymbol{\Sigma}_{m|\mathbf{x}}\right|^{-1/2} \\ &\quad \exp\left[-\frac{1}{2}\left(\mathbf{y}_m - \boldsymbol{\mu}_{m|\mathbf{xz}}\right)' \boldsymbol{\Sigma}^{-1}_{m|\mathbf{x}} \left(\mathbf{y}_m - \boldsymbol{\mu}_{m|\mathbf{xz}}\right)\right]. \qquad (2)\end{aligned}$$

Here, the vector $\boldsymbol{\mu}_{m|\mathbf{xz}}$ contains the conditional expectations of the y variables belonging to set m, and $\boldsymbol{\Sigma}_{m|\mathbf{x}}$ is their variance-covariance matrix. As indicated by their indices, expectations may depend on the latent class to which one belongs and on an individual's covariate values, while variances and covariances may only be class dependent.

A special case occurs when a set consists of a single continuous y variable. In that case, we get a univariate normal distribution, i.e.,

$$f(y_k|\mathbf{xz}) = \frac{1}{\sqrt{2\pi\sigma_{k|\mathbf{x}}}} \exp\left[-\frac{1}{2}\left(\frac{y_k - \mu_{k|\mathbf{xz}}}{\sigma_{k|\mathbf{x}}}\right)^2\right].$$

Here, $\mu_{k|\mathbf{xz}}$ denotes the mean and $\sigma_{k|\mathbf{x}}$ the standard deviation of the distribution.

In normal models, the linear predictor is the expectation of the distribution: $\mu_{k|\mathbf{xz}} = \eta_{k|\mathbf{xz}}$, where

$$\eta_{k|\mathbf{xz}} = \beta^0_k + \sum_{j=1}^{J} \beta^1_{x_j k} + \sum_{\ell=1}^{L} \beta^2_{z_\ell k}.$$

This applies to both the univariate and the multivariate case. The difference between the two models is that in the multivariate case the errors of the K_m regression equations in set m are correlated.

1.2.4 Counts

Counts or numbers of events can be modeled by Poisson distributions. The form of this distribution is

$$f(y_k|\mathbf{xz}) = \frac{1}{y_k!}(\theta_{k|\mathbf{xz}} E_k)^{y_k} \exp(-\theta_{k|\mathbf{xz}} E_k).$$

Here, $\theta_{k|\mathbf{xz}}$ denotes the Poisson rate and E_k the exposure to the event concerned. The rate can also be written as

$$\theta_{k|\mathbf{xz}} = \exp(\eta_{k|\mathbf{xz}}),$$

where $\eta_{k|\mathbf{xz}}$ is again the linear predictor. In LC cluster and factor models, the exposures are assumed to be equal to 1 for all cases. On the other hand, in LC regression models, the exposure is a variable that can be specified by the user.

In LC regression models, counts can also be modeled by a binomial model. The binomial distribution for counts equals

$$f(y_k|\mathbf{xz}) = \left(\frac{T_k!}{y_{k!}(T_k - y_k)!}\right)(\pi_{k|\mathbf{xz}})^{y_k}(1 - \pi_{k|\mathbf{xz}})^{T_k - y_k}.$$

Here, T_k denotes a total, or the maximum number of events that an individual can experience. The parameterization of the binomial probability is similar to the one for dichotomous categorical y variables; i.e.,

$$\pi_{k|\mathbf{xz}} = \frac{\exp(\eta_{k|\mathbf{xz}})}{1 + \exp(\eta_{k|\mathbf{xz}})}.$$

As can be seen, we have a binary logit model with linear predictor $\eta_{k|\mathbf{xz}}$. Note that there is a difference between this logit model and the logit models we used so far: here, we use the category "no event" as reference category, which means that we use dummy coding for the dependent variable instead of effect coding.

1.3 Effect coding for categorical variables

In Latent GOLD, we use effect coding for nominal variables, as well as for the intercept of ordinal variables. This means that the parameters are restricted to sum to zero over the categories of the variable concerned.

Suppose z_1 is a nominal predictor with 4 levels in a model for a continuous dependent variable y_1. The effect coding constraint implies that $\sum_{z_1=1}^{4} \beta_{1z_1}^2 = 0$. An equivalent way to formulate this constraint is by $\beta_{14}^2 = -\sum_{z_1=1}^{3} \beta_{1z_1}^2$; that is, the parameter of the last category equals minus the sum of the other ones. The design matrix that is set up for the 3 non-redundant terms (β_{11}^2, β_{12}^2, β_{13}^2) is as follows:

category 1	1	0	0
category 2	0	1	0
category 3	0	0	1
category 4	−1	−1	−1

where each row corresponds to a category and each column to a parameter. Although the parameter for the last category is omitted from model, you do not notice that because it is computed by the program after the model is estimated.

2 LC Cluster Models

The LC cluster model implemented in Latent GOLD is a model with:

1. a single nominal latent variable,
2. indicator (dependent) variables which can be nominal, ordinal, continuous, and/or counts,
3. nominal and/or numeric covariates,

4. direct relationships between indicators and/or direct effects of covariates on indicators.

This section introduces the most important types of LC cluster models. We start with the description of the standard LC model for (nominal or ordinal) categorical y variables without covariates. Then, we show how nominal and numeric covariates are included in the model. Next, we indicate how the assumption of local independence is relaxed or, in other words, how z-y and y-y relationships are included in a model. The last two parts of this section deal with LC cluster models for continuous y variables and LC cluster models for mixed mode data, respectively.

2.1 LC Cluster Models for Categorical Variables

The standard LC model as described by Lazarsfeld and Henry (1968), Goodman (1974a, 1974b), and Haberman (1979) is a LC cluster model containing only categorical indicators.[3] An example of such a model with four categorical indicators ($K = 4$) is

$$f(y_1 y_2 y_3 y_4) = \sum_{x_1} \pi(x_1) \prod_{k=1}^{4} f(y_k|x_1). \tag{3}$$

By using the symbol π rather that f for a probability, which is more in agreement with the notation used in standard LC models, we can also write the model as

$$\begin{aligned} \pi(y_1 y_2 y_3 y_4) &= \sum_{x_1} \pi(x_1) \prod_{k=1}^{4} \pi(y_k|x_1) \\ &= \sum_{x_1} \pi(x_1)\pi(y_1|x_1)\pi(y_2|x_1)\pi(y_3|x_1)\pi(y_4|x_1). \end{aligned} \tag{4}$$

As can be seen from this probability structure, the indicators y_1, y_2, y_3, and y_4 are assumed to be mutually independent given that one belongs to a certain

[3] Textbooks and introductory papers on the standard LC model include McCutcheon (1987), Hagenaars (1990, 1993), Shockey (1988), Vermunt (1997), Dillon and Kumar (1994), Bartholomew and Knott (1999), Heinen (1996), Clogg (1995), Dayton (1998), Hagenaars and McCutcheon (2002), Magidson and Vermunt (2003), Vermunt and Magidson (2003).

latent class. This conditional independence constraint is sometimes referred to as the local independence assumption. Note that – using the terminology introduced above – each of the M sets consists of a single y variable.

As described in section 1, the conditional (response) probabilities $\pi(y_k|x)$ are parameterized as follows:

$$\pi(y_k|x_1) = \frac{\exp(\eta_{y_k|x_1})}{\sum_{y_k} \exp(\eta_{y_k|x_1})},$$

with

$$\eta_{y_k|x_1} = \beta^0_{y_k} + \beta^1_{x_1 y_k}.$$

If the corresponding indicator is a nominal variable, we do not need to impose further restrictions on the β parameter. On the other hand, if y_k is ordinal, the two-variable interaction term appearing in the logistic form of $\pi(y_k|x)$ is restricted by using fixed category scores for y_k:

$$\beta^1_{x_1 y_k} = \beta^1_{x_1 k} \cdot v_{y_k}.$$

Such a restricted association term is sometimes referred to as a row- or column-association model, depending on whether we see the latent variable as the row or the column variable (Goodman, 1979; Agresti, 2002).

2.2 Covariates

An important extension of the standard LC model described above is the inclusion of covariates in the model (Clogg, 1981; Dayton and McReady, 1988; Hagenaars, 1990, 1993).[4] Inclusion of covariates to predict class membership is straightforward within the framework defined by the general model of equation (1).

Suppose we have a model with four categorical indicators and one nominal and one numeric covariate (z_1 and z_2). The LC cluster model for this situation is

$$\pi(y_1 y_2 y_3 y_4|z_1 z_2) = \sum_{x_1} \pi(x_1|z_1, z_2) \prod_{k=1}^{4} \pi(y_k|x_1).$$

[4]Other references on this topic are Van der Heijden, Dessens, and Böckenholt (1996), Vermunt (1997), Dayton (1998), and McCutcheon (1988, 1994).

Note that compared to the model without covariates described in equation (4), we replaced $\pi(x_1)$ by $\pi(x_1|z_1, z_2)$, which makes the distribution of x_1 dependent on z_1 and z_2. It is important to be aware of the fact that we are making an additional set of conditional independence assumptions: The indicators are assumed to be independent of the covariates given the latent variable x_1.

The probability $\pi(x_1|z_1, z_2)$ is restricted by means of a logit or logistic regression model to exclude higher-order interaction terms, as well as to restrict the two-variable interaction terms for numeric (ordinal, discrete interval, or continuous) covariates. If z_1 is nominal and z_2 numeric, this yields

$$\pi(x_1|z_1 z_2) = \frac{\exp(\eta_{x_1|z_1 z_2})}{\sum_{x_1} \exp(\eta_{x_1|z_1 z_2})}, \tag{5}$$

with

$$\eta_{x_1|z_1 z_2} = \gamma^0_{x_1} + \gamma^1_{z_1 x_1} + \gamma^1_{2x_1} z_2.$$

As can be seen, the three-variable interaction is not included in the logit model. To simplify the model, Latent GOLD always excludes higher-order interactions between covariates from a model.[5] Because z_2 is numeric, the corresponding two-variable term is constrained in the standard manner: $\gamma^1_{z_2 x_1} = \gamma^1_{2x_1} z_2$.

We call this procedure for including covariates in a model the "active covariates method": Covariates are active in the sense that the LC cluster solution with covariates can be somewhat different from the solution without covariates. An alternative method, labeled "inactive covariates method", involves computing descriptive measures for the association between covariates and the latent variable after estimating a model without covariate effects. More detail on the latter method is given in the subsection explaining the ProbMeans output.

2.3 Local Dependencies

As mentioned above, the local independence assumption is the basic assumption of the standard LC model. Lack of fit of a latent model is caused by violation of this assumption. The usual way to proceed is to increase the

[5]Note that higher-order interactions can always be included in a model by adding the corresponding product terms to the data base and using these as additional covariates.

number of classes until a model with an acceptable fit is obtained. An alternative model fitting strategy that we would like to propagate is to relax the local independence assumption by allowing for associations between indicators, as well as direct effects of covariates on the indicators (Hagenaars, 1988; Vermunt, 1997). Latent GOLD calculates bivariate z-y and y-y residuals which can be used to detect which pairs of observed variables are more strongly related than can be explained by the formulated model.

As in the previous subsection, we will use an example of a LC model with four indicators and one nominal (z_1) and one numeric (z_2) covariate. Suppose that we would like to relax two local independence assumptions by assuming that y_1 and y_2 are directly related and that y_3 is influenced by z_2. This yields the following probability structure:

$$\pi(y_1 y_2 y_3 y_4 | z_1 z_2) = \sum_{x_1} \pi(x_1 | z_1 z_2) \pi(y_1 y_2 | x_1) \pi(y_3 | x_1 z_2) \, \pi(y_4 | x_1). \qquad (6)$$

The dependent variables y_1 and y_2 now serve as a joint dependent variable and y_3 is allowed to depend on z_2. The logit models for $\pi(x_1|z_1, z_2)$ and $\pi(y_4|x_1)$ are the same as above. The linear term in the logit model for $\pi(y_1, y_2|x_1)$ equals

$$\eta_{y_1 y_2 | x_1} = \beta^0_{y_1} + \beta^0_{y_2} + \beta^0_{y_1 y_2} + \beta^1_{x_1 y_1} + \beta^1_{x_1 y_2},$$

and the term for $\pi(y_3|z_2)$ equals

$$\eta_{y_3 | z_2} = \beta^0_{y_3} + \beta^1_{x_1 y_3} + \beta^2_{2 y_3} z_2.$$

The linear terms are used again to exclude higher-order interactions from the model, as well as to use the information on the scale type of the variables.

Latent Gold starts setting up the probability structures corresponding to local independence models like the ones described in equations (4) and (6). When users include local dependencies using information on bivariate residuals, the program automatically sets up the correct and most parsimonious probability structure for the situation concerned.

2.4 LC Cluster Models for Continuous Variables

The LC cluster module can not only be used to specify cluster-type models for categorical indicators, but also to estimate models with continuous indicators

(Vermunt and Magidson, 2002). The basic structure of a LC cluster for continuous y variables is:

$$f(\mathbf{y}) = \sum_{x_1} \pi(x_1)\, f(\mathbf{y}|x_1).$$

The different variants are obtained via the specification of $f(\mathbf{y}|x_1)$. The least restrictive model is obtained by assuming that the y's come from class-specific multivariate normal distributions

$$f(\mathbf{y}|x_1) = (2\pi)^{-K_m/2} \left|\boldsymbol{\Sigma}_{\mathbf{x}_1}\right|^{-1/2} \exp\left[-\frac{1}{2}\left(\mathbf{y}-\boldsymbol{\mu}_{x_1}\right)' \boldsymbol{\Sigma}_{x_1}^{-1}\left(\mathbf{y}-\boldsymbol{\mu}_{x_1}\right)\right]$$

This model is also known as a finite mixture of multivariate normals (Wolfe, 1970; McLachlan and Basford, 1988; Banfield and Raftery, 1993; McLachlan and Peel, 2000). As can easily be seen, each class has its own set of means $\boldsymbol{\mu}_{x_1}$ and its own variance-covariance matrix $\boldsymbol{\Sigma}_{x_1}$.

Specification of more restricted models for continuous y variables typically involves fixing some elements of the covariance matrix to zero. The most restrictive model assumes that all covariances equal zero, which is equivalent to the local independence assumption. This model can also be written as

$$f(\mathbf{y}) = \sum_{x_1} \pi(x_1) \prod_{k=1}^{K} f(y_k|x_1), \tag{7}$$

with

$$f(y_k|x_1) = \frac{1}{\sqrt{2\pi\sigma_{k|x_1}}} \exp\left[-\frac{1}{2}\left(\frac{y_k-\mu_{k|x_1}}{\sigma_{k|x_1}}\right)^2\right].$$

Note that the structure in equation (7) is equivalent to the standard LC model for categorical variables given in equation (3).

Many intervening models can be obtained by setting some but not all off-diagonal elements of the matrices $\boldsymbol{\Sigma}_{x_1}$ to zero. This yields a specification in which some indicators are independent of one another within classes while others are not. As was explained in section 1, in such situations, we will work with sets of y variables, where indicators belonging to different sets are assumed to be locally independent. Indicators belonging to the same set may be dependent or independent of one another. As with categorical y variables, the default is the local independence model described in equation (7). Users can include local dependencies one by one, and the program subsequently sets up the correct and most parsimonious probability structure.

An important issue in the specification of mixtures of normal distributions is whether to work with class-dependent or class-independent error variances and covariances. So far, we assumed the error variances and covariances to be class dependent. However, when the number of y variables and/or number of latent classes is large, this may yield models that have many parameters to be estimated. By replacing $\sigma_{k|x_1}$ by σ_k or $\mathbf{\Sigma}_{x_1}$ by $\mathbf{\Sigma}$ in the above formulas, one obtains more parsimonious structures with class-independent variances and covariances.

Finally, as in LC cluster models for categorical variables, it is possible to include covariates as predictors of class membership in mixtures of normals. Covariates can not only influence class membership, but can also have direct effects on the class specific means. The most general LC cluster model with covariates is similar to the general Latent GOLD model described in equation (1), with class-specific densities of the form described in equation (2). The only difference is that, contrary to the general model described in equation (1), LC cluster models contain only one latent variable.

2.5 LC Cluster Models for Mixed Mode Data

The most general LC cluster model is the model for mixed mode data (Everitt, 1988; Lawrence and Krzanowski, 1996; Moustaki, 1996; Hunt and Jorgensen, 1999; Vermunt and Magidson, 2002). This model is used when one has y variables of different scale types. The basic structure is again the local independence structure that we also used for categorical and continuous variables. For each indicator, the user has to specify whether it is nominal, ordinal, continuous, or a count. As in the above models for categorical and continuous indicators, it is possible to include covariates in LC cluster models for mixed mode data. These covariates can also have direct effects on the various types of indicators.

Local dependencies between pairs of categorical variables and between pairs of continuous variables are dealt with in the same way as discussed above; that is, via joint multinomial and multivariate distributions. However, currently, there is no direct way to specify other kinds of y-y interactions. There is, however, an indirect method (a trick) to specify a local dependency between, for instance, a categorical, say y_1, and a continuous, say y_2, indicator (Vermunt and Magidson, 2002). This can be accomplished by duplicating the categorical indicator and using it both as a covariate and as an indicator in the model. The local dependency is obtained by specifying that the "covariate y_1" has direct effect on y_2, but does not affect the latent variable x_1. This yields the conditional Gaussian distribution for y_2 proposed by Hunt and Jorgensen (1999). In a similar way, one can specify direct effects of a continuous y on a categorical y, of a categorical y on a Poisson count, and of a continuous y on a Poisson count.

3 LC Factor Models

The LC factor model implemented in Latent GOLD is a model with:

1. one or more dichotomous/ordinal latent variables, which may be assumed to be mutually dependent or independent,
2. indicator (dependent) variables which are nominal, ordinal, continuous, and/or counts,
3. certain loadings restricted to zero,
4. nominal and/or numeric covariates,
5. direct relationships between indicators and/or direct effects of covariates on indicators.

The main difference between the LC factor and the LC cluster model is that the former may contain more than one latent variable. Another difference is that in the factor model the categories of the latent variables are assumed to be ordered. Thus, rather than working with a single nominal latent variable, here we work with one or more dichotomous or ordered polytomous latent variables (Magidson and Vermunt, 2001). The advantage of this approach is that it guarantees that each of the factors is one-dimensional.

The primary difference between our LC factor and the traditional factor analysis model is that the latent variables (factors) are assumed to be dichotomous or ordinal as opposed to continuous and normally distributed. Because of the strong similarity with traditional factor analysis, we call this approach LC factor analysis. There is also a strong connection between LC factor models and item response or latent trait models. Actually, LC factor models are discretized variants of well-known latent trait models for dichotomous and polytomous items (Heinen, 1996; Vermunt, 2001).[6]

As in maximum likelihood factor analysis, modeling under the LC factor approach can proceed by increasing the number of factors until a good fitting model is achieved. This approach to LC modeling provides a general alternative to the traditional method of obtaining a good fitting model by increasing the number of latent classes. In particular, when working with dichotomous uncorrelated factors, there is an exact equivalence in the number of parameters of the two models. A LC factor model with 1 factor has the same number of parameters as a 2-class LC cluster model, a model with 2 factors as a 3-class model, a model with 3 factors as a 4-class model, etc. Thus, in an exploratory analysis, rather than increasing the number of classes one may instead increase the number of factors until an acceptable fit is obtained.

3.1 A Two-factor Model for Nominal Indicators

To illustrate the LC factor model, let us assume that we have a two-factor model for four nominal categorical indicators. The corresponding probability structure is of the form

$$\pi(y_1 y_2 y_3 y_4) = \sum_{x_1} \sum_{x_2} \pi(x_1 x_2) \prod_{k=1}^{4} \pi(y_k | x_1 x_2).$$

The conditional response probabilities $\pi(y_k | x_1, x_2)$ are restricted by means of logit models with linear terms

$$\eta_{y_k | x_1 x_2} = \beta^0_{y_k} + \beta^1_{1 y_k} \cdot v_{x_1} + \beta^1_{2 y_k} \cdot v_{x_2}.$$

Because the factors are assumed to be ordinal (or discrete interval) variables, the two-variable terms are restricted by using fixed category scores for the levels of the factors, which yields so-called row-association structures. The scores v_{x_j} for factor j are equidistant scores ranging from 0 to 1. The first level of a factor gets the score 0 and the last level the score 1.

Note that above logit model does not include the three-variable interaction term of the two factors and the indicator. The parameters describing the strength of relationships between the factors and the indicators – here, $\beta^1_{1 y_k}$ and $\beta^1_{2 y_k}$ – can be interpreted as factor loadings.

[6] Two edited volumes paying special attention to the relationship between latent trait and latent class models are Langeheine and Rost (1988) and Rost and Langeheine (1997).

In the standard LC factor model, the factors are specified to be dichotomous, which means that the row-association structures are not real constraints. An important extension of this standard model is, however, increasing the number of levels of a factor, which makes it possible to describe more precisely the distribution of the factor concerned. Note that the levels of the factors remain ordered by the use of fixed equal-interval category scores in their relationships with the indicators.[7] Therefore, each additional level costs only one degree of freedom; that is, there is one additional class size to be estimated.

In the default setting, the factors are assumed to be independent of one another. This is specified by the appropriate logit constraints on the latent probabilities. In the two-factor case, this involves restricting the linear term in the logit model for $\pi(x_1 x_2)$ by

$$\eta_{x_1 x_2} = \gamma^0_{x_1} + \gamma^0_{x_2}.$$

Working with correlated factors is comparable to performing an oblique rotation. The association between each pair of factors is described by a single uniform association parameter:

$$\eta_{x_1 x_2} = \gamma^0_{x_1} + \gamma^0_{x_2} + \gamma^0_{12} \cdot v_{x_1} \cdot v_{x_2}.$$

It should be noted that contrary to traditional factor analysis, the LC factor model is identified without additional constraints, such as setting certain factor loadings equal to zero.[8] Nevertheless, it is possible to specify models in which factor loadings are fixed to zero. Together with the possibility to include factor correlation in the model, this option can be used for a confirmatory factor analysis.

3.2 Other Possibilities

Above, we presented the two-factor LC model for nominal categorical variables. We discussed several important extensions of the standard model,

[7] Several authors have described such a scoring of latent classes. See, for instance, Uebersax (1993, 1997), Heinen (1996), Clogg (1988), Formann (1992), Formann and Kolhmann (1998), Vermunt (2001).

[8] Of course, this is only true if there are sufficient indicators compared to the number of factors.

such as increasing the number of categories of the latent variables, assuming factors to be correlated, and setting factor loadings equal to zero.

Other extended possibilities are similar to what we discussed in the context of LC clustering. Indicators can not only be nominal but also ordinal, continuous, or counts. In addition, as in the LC cluster models described in the previous section, one can specify models with local dependencies, as well as models with nominal and numeric covariates.

4 LC Regression Models

The LC or mixture regression model implemented in Latent GOLD is a model with:

1. a single nominal latent variable,
2. a single dependent variable which may be nominal, ordinal, continuous, or a binomial or Poisson count,
3. replication or repeated observations,
4. zero, equality, and order restrictions on regression coefficients
5. nominal and/or numeric covariates.

The main difference between LC regression analysis and the other forms of LC analysis implemented in Latent GOLD is that it contains a single dependent variable. This dependent variable, which may be observed more than once for each case, can be nominal, ordinal, a count, or continuous. As with LC cluster models, LC regression models contain a single nominal latent variable.[9]

4.1 The General LC Regression Model

An important feature of LC regression models is that for each case we may have more than one observation. These multiple observations may be experimental replications, repeated measurements at different time points or

[9]References to LC or finite mixture regression analysis are Wedel and DeSarbo (1994), Wedel and Kamakura (1998), Vermunt and Van Dijk (2001), Agresti (2002, section 13.2).

occasions, clustered observations, or other types of dependent observations. Here, we will use the term replications, where the replication number will be denoted by t. The number of replications is not necessarily the same for all cases. The total number of replications for case i is denoted by T_i. The value of the dependent variable at replication t is denoted by y_t.

In the context of LC regression analysis, it makes sense to make a distinction between two types of z variables:

1. variables influencing the latent variable,

2. variables influencing the dependent variable.

Because the first type is similar to the z variables used so far, we will call them "covariates". For the second type, we introduce the name "predictors" since they serve as predictors in the regression equation for the dependent variable y. Covariates will be denoted by z^c and predictors by z_t^p. The index t in z_t^p reflects that the value of a predictor may change across replications. A covariate, on the other hand, has the same value across the replications belonging to the same case.

Note that we are describing a two-level data structure: covariates are the higher-level independent variables while predictors are lower-level independent variables. Here, t indexes the (dependent) lower-level observations within a certain higher-level observation. This illustrates that Latent GOLD can be used to estimate (non-parametric) two-level or random-coefficient models. Using t as an index for time points or time intervals, we obtain models for longitudinal data, like growth or event history models with non-parametric random coefficients (Wedel et al., 1995; Vermunt, 1997, 2002, Vermunt and Van Dijk, 2002).

Using the slightly modified notation described above, the most general LC regression model can be defined as

$$f(\mathbf{y}|\mathbf{z}^c\mathbf{z}^p) = \sum_{x_1} \pi(x_1|\mathbf{z}^c) \prod_{t=1}^{T_i} f(y_t|x_1\mathbf{z}_t^p) \,.$$

For nominal or ordinal dependent variables, the probability density $f(y_t|x_1\mathbf{z}_t^p)$ will be assumed to be multinomial, for continuous variables, univariate normal, and for counts, Poisson or binomial.

In Latent GOLD 3.0, it is possible to specify a replication weight rw_{it} for each record belonging to case i. These modify the definition of $f(\mathbf{y}|\mathbf{z}^c\mathbf{z}^p)$ as follows:

$$f(\mathbf{y}|\mathbf{z}^c\mathbf{z}^p) = \sum_{x_1} \pi(x_1|\mathbf{z}^c) \prod_{t=1}^{T_i} [f(y_t|x_1\mathbf{z}_t^p)]^{rw_{it}} . \tag{8}$$

The linear predictor in $\pi(x_1|\mathbf{z}^c)$ has the same form as in LC cluster models (see equation 5). In the case of a nominal dependent variable, the linear predictor in $f(y_t|x_1\mathbf{z}_t^p)$ equals

$$\eta_{y|x_1\mathbf{z}_t^p} = \beta^1_{x_1 y} + \sum_{\ell=1}^{L^p} \beta^2_{x_1 z^p_{t\ell} y}, \tag{9}$$

where L^p denotes the number of predictors. With counts or continuous dependent variables, we would omit the index y – which reflects that there is a separate parameter for each level of the dependent variable – from the β parameters. For ordinal variables, $\beta^2_{x_1 z^p_{t\ell} y} = \beta^2_{x_1 z^p_{t\ell}} \cdot v_y$.

The part of the LC regression model described in (9) contains two types of parameters: a set of class-specific intercepts and a set of class-specific regression coefficients. Note that the parameters are assumed to be independent of the replication number t.[10] In the case of numeric predictors, the β's are restricted in the usual manner: $\beta^2_{x_1 z^p_{t\ell} y} = \beta^2_{x_1 y} \cdot z_{t\ell}$.

While in the standard LC regression model all regression coefficients are assumed to be class dependent, it is also possible to specify models in which one or more regression coefficients are class independent; that is, $\beta^2_{x_1 z^p_{t\ell} y} = \beta^2_{z^p_{t\ell} y}$. Wald statistics are reported that help users to determine for which coefficients such restrictions might be valid.

4.2 The Most Important Special Cases

The simplest probability structure for a LC "regression model" is

$$f(y_1) = \sum_{x_1} \pi(x_1) f(y_1|x_1) .$$

This is a finite mixture model in which the mean and possibly also the variance of the distribution of y_1 is assumed to be class-dependent. Such a model

[10] By defining an independent variable "replication number" or "time", as well as the necessary product terms, one can make parameters replication or time dependent.

without predictors makes it possible to describe the unobserved heterogeneity with respect to the distribution of y_1 in the population under study.[11]

A more useful LC regression model is obtained by including independent variables in the model, such as,

$$f(y_1|z_1^p z_2^p) = \sum_{x_1} \pi(x_1) f(y_1|x_1 z_1^p z_2^p) \, .$$

Here, $f(y_1|x_1 z_1^p z_2^p)$ denotes the distribution of the dependent variable y_1 given a person's scores on the predictors z_1^p, and z_2^p, and latent position x_1. Depending on the type of dependent variable, the mean of the appropriate distribution is restricted by means of a logit, log-linear, or linear regression model.[12]

An important extension of the above LC regression model is obtained by making class membership dependent on covariates (Kamakura, Wedel, and Agrawal, 1994; Vermunt, 1997). An example of such a model is:

$$f(y_1|z_1^c z_2^c z_1^p z_2^p) = \sum_{x_1} \pi(x_1|z_1^c z_2^c) f(y_1|x_1 z_1^p z_2^p) \, .$$

In this model, it is assumed that the probability of belonging to latent class x_1 depends on the values of z_1^c and z_2^c. This is equivalent to the way covariates can be used in the other types of LC models.

As already mentioned above, there may be more than one observation per case; that is, there may be more than one replication of the same dependent and independent variables for each observational unit. The model above with replications yields the following probability structure:

$$f(\mathbf{y}|z_1^c z_2^c \mathbf{z}_1^p \mathbf{z}_2^p) = \sum_{x_1} \pi(x_1|z_1^c z_2^c) \prod_{t=1}^{T_i} f(y_{1t}|x_1 z_{t1}^p z_{t2}^p) \, .$$

Such a LC regression model for repeated measures is very similar to multilevel (two-level), mixed, or random-coefficients models, in which random effects are included to deal with the dependent observations problem. The LC regression model is a non-parametric random-effects model (Vermunt and Van Dijk, 2001; Agresti, 2002, section 13.2).

[11] Relevant references on this topic are Laird (1978), Everitt and Hand (1981), Titterington, Smith, and Makov, (1985), McLachlan and Basford (1988), McLachlan and Peel (2000), Dillon and Kumar (1994), Böhning (2000), and Magidson and Vermunt (2003).

[12] For some applications, see Wedel and DeSarbo (1994), Böckenholt (1993), Land, McCall and Nagin (1996), and Mare (1994).

4.3 Restrictions on Class-specific Regression Coefficients

Various types of restrictions can be imposed on the class-specific regression coefficients: predictor effects can be fixed to zero, restricted to be equal across certain or all classes, and constrained to be ordered.

Equality and zero restrictions do not require further explanation. The order restrictions (ascending/descending) are somewhat more complex. For a numeric predictor, the ascending restriction implies that its coefficient should be at least zero ($\beta \geq 0$) and descending that it should be at most zero ($\beta \leq 0$). For a nominal predictor, ascending implies that the coefficient of category $s+1$ is larger or equal to the one of category s ($\beta_s \leq \beta_{s+1}$, for each s) and descending that the coefficient of category $s+1$ is smaller or equal to the one of category s ($\beta_s \geq \beta_{s+1}$, for each s).

5 Estimation

5.1 Log-likelihood and Log-posterior Function

The parameters of the various types of LC models are estimated by means of Maximum Likelihood (ML) or Posterior Mode (PM) methods. The likelihood function is derived from the probability density function defined in equation (1). Let $\boldsymbol{\vartheta}$ denote the vector containing the various types of model parameters, let $\mathbf{y}_i$ and $\mathbf{z}_i$ denote the vector of dependent and independent variables – or indicators and covariates – for response pattern i, and let n_i be the number of cases with pattern i.

Identical response patterns are grouped by the program. It should be noted that it is possible to specify a *Case Weight* (cw). The cell counts n_i are obtained by summing the case weights of identical cases. When cw are sampling weights in a complex survey, Latent GOLD will provide pseudo ML estimates (Patterson, Dayton, and Graubard, 2002). Another type of weight is a *Replication Weight* (rw). As explained in the context of the LC regression model, these modify the definition of the relevant probability density (see equation 8).

ML estimation involves finding the estimates for $\boldsymbol{\vartheta}$ that maximize the log-likelihood function

$$\log \mathcal{L} = \sum_i n_i \log f(\mathbf{y}_i|\mathbf{z}_i, \boldsymbol{\vartheta}).$$

Here, $f(\mathbf{y}_i|\mathbf{z}_i, \boldsymbol{\vartheta})$ is the probability density for pattern i given parameter values $\boldsymbol{\vartheta}$.

In order to prevent boundary solutions or, equivalently, to circumvent the problem of non-existence of ML estimates, we implemented some ideas from Bayesian statistics in Latent GOLD. The boundary problems are that multinomial probabilities and Poisson rates may become zero, and that error variances in normal models may converge to zero. The first problem is circumvented by using Dirichlet priors for the latent and the conditional response probabilities and gamma priors for the Poisson rates, and the second by using inverse-Wishart priors for the error variance-covariance matrices (Gelman et. al., 1996; Schafer, 1997; Clogg et al., 1991). These are so-called conjugate priors since they have the same form as the corresponding multinomial, Poisson, and multivariate normal probability densities. The implication of using priors is that the estimation method is no longer ML but PM (Posterior Mode).

Denoting the assumed priors for $\boldsymbol{\vartheta}$ by $h(\boldsymbol{\vartheta})$ and the posterior by $\mathcal{P}$, PM estimation involves finding the estimates for $\boldsymbol{\vartheta}$ that maximize the log-posterior function

$$\begin{aligned}\log \mathcal{P} &= \log \mathcal{L} + \log h(\boldsymbol{\vartheta}) \\ &= \sum_i n_i \log f(\mathbf{y}_i|\mathbf{z}_i, \boldsymbol{\vartheta}) + \log h(\boldsymbol{\vartheta}),\end{aligned}$$

or, in other words, finding the point where $\frac{\partial \log \mathcal{P}}{\partial \boldsymbol{\vartheta}} = 0$. Algorithms that are used to solve this problem are described in the next subsection.

The user-defined parameters in the priors $h(\boldsymbol{\vartheta})$ can be chosen in such a way that $\log h(\boldsymbol{\vartheta}) = 0$, which makes PM estimation turn into ML estimation. PM estimation can also be seen as a form of penalized ML estimation, in which $h(\boldsymbol{\vartheta})$ serves as a function penalizing solutions that are too near to the boundary of the parameter space and, therefore, smoothing the estimates away from the boundary.

We will not present again the different kinds of forms that the density $f(\mathbf{y}_i|\mathbf{z}_i, \boldsymbol{\vartheta})$ can take on since these were described in detail in section 1.

5.2 Missing Data

5.2.1 Indicators and dependent variable

Latent GOLD has a provision for dealing with missing data in the y variables. As mentioned above, the likelihood function is derived from the probability density function defined in equation (1). However, to be able to deal with missing data, we need to make a minor modification in the definition of the density function of interest; that is,

$$f(\mathbf{y}_i|\mathbf{z}_i, \boldsymbol{\vartheta}) = \sum_{\mathbf{x}} \pi(\mathbf{x}|\mathbf{z}_i, \boldsymbol{\vartheta}) \prod_{m=1}^{M} f(\mathbf{y}_{im}|\mathbf{x}, \mathbf{z}_i, \boldsymbol{\vartheta})^{\delta_{im}}.$$

As can be seen, we have defined an additional variable δ_{im}. This variable equals one if all y variables in set m are observed for data pattern i, and zero otherwise. What actually happens is that terms for which $\delta_{im} = 0$ cancel from the likelihood equation. This implies that we estimate the model parameters using the available information for each of the response patterns, which amounts to assuming a missing at random (MAR) nonresponse mechanism (Little and Rubin, 1987; Schafer, 1998, Vermunt, 1997; Skrondal, 1996).

Two additional remarks have to be made with respect to the missing data procedure implemented in Latent GOLD. First, it should be noted that we use a single δ_{im} to indicate whether all y's in set m are observed. Thus, δ_{im} is also equal to zero if some y's in set m are observed while others are missing. This implies that when local dependencies are included in a model, it may happen that some available information is disregarded.

The second remark concerns the chi-squared goodness-of-fit statistics. Although parameter estimation with missing data is based the MAR assumption, the chi-squared statistics not only tests whether the model of interest holds, but also the much more restrictive MCAR (missing completely at random) assumption (see Vermunt, 1997). Thus, caution should be used when interpreting the overall goodness-of-fit tests in situations in which there is missing data.

5.2.2 Covariates and predictors

It is also possible to use cases with missing values on the predictors and covariates. If the predictor or covariate is numeric, Latent GOLD will impute the sample mean for the missing values. This is the mean over all cases without a missing value for covariates and the mean of all replications without a missing value for predictors.

Missing values on nominal predictors and covariates is dealt with via the design matrix. In fact, the effect is set equal to zero for the missing value category. Recall the effect coding scheme illustrated in subsection 1.3 for the case of a nominal predictor with 4 categories. Suppose there is also a missing category. The design matrix that is set up for the 3 non-redundant terms is then

category 1	1	0	0
category 2	0	1	0
category 3	0	0	1
category 4	−1	−1	−1
missing	0	0	0

As can be seen, the entries corresponding to the missing category are all equal to 0, which amounts to setting its coefficient equal to zero. Note that in effect coding the mean of the coefficients equals to zero. This imputation method for nominal variables is therefore similar to mean imputation for numeric variables.

5.3 Prior Distributions

The different types of priors have in common the fact that their user-defined parameters (*Bayes Constants*) denoted by α can be interpreted as adding α observations – for instance, the program default of one – generated from a conservative null model (as decribed below) to the data. All priors are defined in such a way that if the corresponding α's are set equal to zero, log $h(\boldsymbol{\vartheta}) = 0$, in which case we will obtain ML estimates. We could label such priors as "non-informative". Below we present the log $h(\boldsymbol{\vartheta})$ terms for the various types of distributions without their normalizing constants.

The Dirichlet prior for the latent probabilities equals

$$\log h\left(\pi(\mathbf{x})\right) = \sum_{\mathbf{x}} \frac{\alpha_1}{C_{\mathbf{x}}} \log \pi(\mathbf{x}).$$

Here, $C_{\mathbf{x}}$ denotes the number of classes and α_1 the *Bayes Constant* to be specified by the user. As can be seen, the influence of the prior is equivalent to adding $\frac{\alpha_1}{C_{\mathbf{x}}}$ cases to each latent class. This prior makes the sizes of the latent classes slightly more equal and the covariate effects somewhat smaller.

For categorical dependent variables and binomial counts, we use the following Dirichlet prior for $\pi(y_k|\mathbf{x})$:

$$\log h(\pi(y_k|\mathbf{x})) = \sum_{y_k} \frac{p(y_k)\alpha_2}{C_{\mathbf{x}}} \log \pi(y_k|\mathbf{x}),$$

where $p(y_k)$ is the observed distribution of y_k. This prior can be interpreted as adding $\frac{\alpha_2}{C_{\mathbf{x}}}$ observations to each class with preservation of the observed distribution of y_k, where α_2 is a parameter to be specified by the user. This prior makes the class-specific distributions more similar or, in other words, the β parameters will be slightly closer to zero.

For Poisson counts, we implemented a gamma prior. Let $r(y_k)$ be the observed Poisson rate. The prior we use has the form

$$\log h(y_k|\mathbf{x}) = \frac{\alpha_3}{C_{\mathbf{x}}} \log \left[\frac{\theta_{k|\mathbf{x}}\alpha_3}{C_{\mathbf{x}} r(y_k)}\right] - \frac{\theta_{k|\mathbf{x}}\alpha_3}{C_{\mathbf{x}} r(y_k)}.$$

This prior can be interpreted as adding $\frac{\alpha_3}{C_{\mathbf{x}}}$ events to each class with preservation of the overall Poisson rate $r(y_k)$.

The inverse-Wishart priors[13] for the error variance-covariance matrices are of the form

$$\log h(\mathbf{\Sigma}_{m|\mathbf{x}}) = -0.5\frac{\alpha_4}{C_{\mathbf{x}}} \log \left|\mathbf{\Sigma}_{m|\mathbf{x}}\right| - 0.5\frac{\alpha_4}{C_{\mathbf{x}}} \mathrm{tr}\left(\mathbf{D}_{\mathbf{s}^2_m} \mathbf{\Sigma}^{-1}_{m|\mathbf{x}}\right).$$

Here, $\mathbf{D}_{\mathbf{s}^2_m}$ is a diagonal matrix containing the observed variances of the K_m dependent variables belonging to set m. This prior can be interpreted as incrementing each class with $\frac{\alpha_4}{C_{\mathbf{x}}}$ observations which are at a distance of one standard deviation of the class-specific mean and which have covariances of zero. Here, α_3 is the parameter to be specified by the user. This prior slightly increases the class-specific error variances and slightly decreases the class-specific covariances.

The influence of the priors on the final parameter estimates depends on the values chosen for the α's, as well as on the sample size. The default settings are $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 1.0$. This means that with moderate sample sizes the influence of the priors on the parameter estimates is negligible. Setting $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0$ yields ML estimates.

[13] Actually, our prior differs somewhat from a real inverse-Wishart distribution since we omit the term $-\frac{K_m+1}{2}\log\left|\mathbf{\Sigma}_{m|\mathbf{x}}\right|$, where K_m is the number of variables in set m. The reason for doing this is that this term does not become zero if $\alpha_4 = 0$, which is something we want to happen in order to be able to switch from PM to ML estimation.

5.4 Algorithms

To find the ML or PM estimates for the model parameters $\boldsymbol{\vartheta}$, Latent GOLD uses both the EM and the Newton-Raphson algorithm. In practice, the estimation process starts with a number of EM iterations. When close enough to the final solution, the program switches to Newton-Raphson. This is a way to exploit the advantages of both algorithms; that is, the stability of EM when it is far away from the optimum and the speed of Newton-Raphson when it is close to the optimum.

The tasks to be performed for obtaining PM estimates for $\boldsymbol{\vartheta}$ is finding the parameter values for which

$$\frac{\partial \log \mathcal{P}}{\partial \boldsymbol{\vartheta}} = \frac{\partial \log \mathcal{L}}{\partial \boldsymbol{\vartheta}} + \frac{\partial \log h(\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}} = 0. \tag{10}$$

Here

$$\begin{aligned} \frac{\partial \log \mathcal{L}}{\partial \boldsymbol{\vartheta}} &= \sum_i n_i \frac{\partial \log f(\mathbf{y}_i|\mathbf{z}_i, \boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}} \\ &= \sum_i n_i \frac{\partial \log \sum_{\mathbf{x}} \pi(\mathbf{x}|\mathbf{z}_i, \boldsymbol{\vartheta}) f(\mathbf{y}_i|\mathbf{x}\mathbf{z}_i, \boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}} \\ &= \sum_i \sum_{\mathbf{x}} n_{i\mathbf{x}} \frac{\partial \log \pi(\mathbf{x}|\mathbf{z}_i, \boldsymbol{\vartheta}) f(\mathbf{y}_i|\mathbf{x}\mathbf{z}_i, \boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}}, \end{aligned} \tag{11}$$

where

$$n_{i\mathbf{x}} = n_i \pi(\mathbf{x}|\mathbf{z}_i \mathbf{y}_i, \boldsymbol{\vartheta}) = n_i \frac{\pi(\mathbf{x}|\mathbf{z}_i, \boldsymbol{\vartheta}) f(\mathbf{y}_i|\mathbf{x}\mathbf{z}_i, \boldsymbol{\vartheta})}{f(\mathbf{y}_i|\mathbf{z}_i, \boldsymbol{\vartheta})}. \tag{12}$$

The *EM algorithm* is a general method for dealing with ML estimation with missing data (Dempster, Laird, and Rubin, 1977; McLachlan and Krishnan, 1997). This method exploits the fact that the first derivatives of the incomplete data log-likelihood ($\log \mathcal{L}$) equal the first derivatives of the complete data log-likelihood ($\log \mathcal{L}^c$). In our case, the complete data log-likelihood is the log-likelihood that we would have if we knew to which class cases with response pattern i belong:

$$\log \mathcal{L}^c = \sum_i \sum_{\mathbf{x}} n_{i\mathbf{x}} \log \pi(\mathbf{x}|\mathbf{z}_i, \boldsymbol{\vartheta}) f(\mathbf{y}_i|\mathbf{x}\mathbf{z}_i, \boldsymbol{\vartheta}). \tag{13}$$

Each νth cycle of the EM algorithm consists of two steps. In the Expectation (E) step, estimates $\widehat{n}_{i\mathbf{x}}^{\nu}$ are obtained for $n_{i\mathbf{x}}$ via equation (12) filling in $\widehat{\boldsymbol{\vartheta}}^{\nu-1}$ as parameter values. The Maximization (M) step, involves finding new $\widehat{\boldsymbol{\vartheta}}^{\nu}$ improving $\log \mathcal{L}^c$. Note that, actually, we use PM rather than ML estimation, which means that in the M step we update the parameters in such a way that

$$\log \mathcal{P}^c = \log \mathcal{L}^c + \log h(\boldsymbol{\vartheta}) \tag{14}$$

increases rather than (13). Sometimes closed-form solutions are available in the M step. In other cases, standard iterative methods can be used to improve the complete data log-posterior defined in equation (14). Latent GOLD uses iterative proportional fitting (IPF) and unidimensional Newton in the M step (see Vermunt 1997, Appendices).

Besides the EM algorithm, Latent GOLD also implements a *Newton-Raphson* (NR) method.[14] This general optimization algorithm works as follows:

$$\widehat{\boldsymbol{\vartheta}}^{\nu} = \widehat{\boldsymbol{\vartheta}}^{\nu-1} - \varepsilon\, \mathbf{H}^{-1}\mathbf{g}.$$

The gradient vector $\mathbf{g}$ contains the first-order derivatives of the log-posterior to all parameters evaluated at $\widehat{\boldsymbol{\vartheta}}^{\nu-1}$, $\mathbf{H}$ is the Hessian or observed information matrix containing the second-order derivatives, and ε is a scalar denoting the step size. Element g_r of $\mathbf{g}$ equals

$$g_r = \sum_i n_i \frac{\partial \log f(\mathbf{y}_i|\mathbf{z}_i, \boldsymbol{\vartheta})}{\partial \vartheta_r} + \frac{\partial \log h(\boldsymbol{\vartheta})}{\partial \vartheta_r},$$

and element H_{rs} of $\mathbf{H}$ equals

$$H_{rs} = \sum_i n_i \frac{\partial^2 \log f(\mathbf{y}_i|\mathbf{z}_i, \boldsymbol{\vartheta})}{\partial \vartheta_r \partial \vartheta_s} + \frac{\partial^2 \log h(\boldsymbol{\vartheta})}{\partial \vartheta_r \partial \vartheta_s}.$$

Latent GOLD computes these derivatives analytically. The step size ε prevents decreases of the log-posterior.

The matrix $\mathbf{H}^{-1}$ evaluated at the final $\widehat{\boldsymbol{\vartheta}}$, $\widehat{\mathbf{H}}^{-1}$, yields the estimated asymptotic variance-covariance matrix of the model parameters. Note that $\widehat{\mathbf{H}}^{-1}$ can be used to obtain the standard error for any function $f(\widehat{\boldsymbol{\vartheta}})$ of $\widehat{\boldsymbol{\vartheta}}$ by

[14] Haberman (1988) proposed estimation standard LC models by Newton Raphson.

the delta method:

$$\widehat{se}\left(f(\widehat{\boldsymbol{\vartheta}})\right) = \sqrt{\left(\frac{f(\widehat{\boldsymbol{\vartheta}})}{\partial\widehat{\boldsymbol{\vartheta}}}\right)' \widehat{\mathbf{H}}^{-1} \left(\frac{f(\widehat{\boldsymbol{\vartheta}})}{\partial\widehat{\boldsymbol{\vartheta}}}\right)}. \tag{15}$$

The order restrictions in regression are dealt with using an active-set method (Gill, Murray, and Wright, 1981). For that purpose, we first transform the effects involved in the order constraints into nonnegativity constraints of the form $\beta \geq 0$. In an active-set method, the equality constraint associated with an inequality constraint is activated if it is violated (here, effect set equal to 0), and inactivated if its update yields an admissible value (here, a positive update).

5.5 Convergence

The exact algorithm implemented in Latent GOLD works as follows. The program starts with EM until either the maximum number of EM iterations (*Iteration Limits EM*) or the EM convergence criterion (*EM Tolerance*) is reached. Then, the program switches to NR iterations which stop when the maximum number of NR iterations (*Iteration Limits Newton-Raphson*) or the overall converge criterium (*Tolerance*) is reached. The convergence criterion that is used is

$$\sum_r \left| \frac{\widehat{\vartheta}_r^{\nu} - \widehat{\vartheta}_r^{\nu-1}}{\widehat{\vartheta}_r^{\nu-1}} \right|,$$

which is the sum of the absolute relative changes in the parameters. The program also stops its iterations when the change in the log-posterior is negligible, i.e., smaller than 10^{-12}.

The program reports the iteration process in Iteration Detail. Thus, it can easily be checked whether the maximum number of iterations is reached without convergence. In addition, a warning is given if one of the elements of the gradient is larger than 10^{-3}.

It should be noted that sometimes it is more efficient to use only the EM algorithm, which is accomplished by setting *Iteration Limits Newton-Raphson* = 0 in the Technical Tab. This is, for instance, the case in models with many parameters and in models for continuous y variables with complicated variance-covariance matrix structures. For very large models one might also consider suppressing the computation of standard errors and Wald statistics.

5.6 Start Values

Latent GOLD generates random start values. These differ every time that a model is estimated because the seed of the random number generator is obtained from the system time, as long as the technical option *Seed* equals 0. The seed used by the program is reported in the output. A run can be replicated by specifying the reported best start seed as *Seed* in the Technical Tab and setting the number of *Random Sets* to zero.

Since the EM algorithm is extremely stable, the use of random starting values is generally good enough to obtain a converged solution. However, there is no guarantee that such a solution is also the global PM or ML solution. A well-known problem in LC analysis is the occurrence of one or more local maxima, which also fulfill the conditions defined by likelihood equations given in (10).

The best way to avoid ending up with a local solution is to use multiple sets of starting values since different sets of starting values may yield solutions with different log-posterior values. In Latent GOLD, the use of such multiple sets of random starting values is automated. The user can specify how many sets of starting values the program should use by changing the *Random Sets* option in the Technical Tab. Another relevant parameter is *Iterations* specifying the number of iterations to be performed per start set. More precisely, within each of the random sets, Latent GOLD performs the specified number of EM iterations. Subsequently, within the best 10 percent in terms of log-posterior, the program performs an extra 2 times *Iterations* EM iterations. Finally, it continues with the best solution until convergence. It should be noted that such a procedure considerably increases the probability of finding the global PM or ML solution, especially if both parameters are set large enough, but in general does not guarantee that it will be found in a single.

5.7 Bootstrapping P-Value of L^2 Statistic

Rather than relying on the asymptotic p-value, it also possible to estimate the p-value associated with the L^2 statistic by means of a parametric bootstrap. This option is especially useful with sparse tables (Langeheine, Pannekoek, and Van de Pol, 1996). The model of interest is then not only estimated for the sample under investigation, but also for R replication samples. These are generated from the probability distribution defined by the ML estimates. The estimated bootstrap p-value, $\widehat{p}_{boot}$, is defined as the proportion of bootstrap samples with a larger L^2 than the original sample. The standard error of $\widehat{p}_{boot}$ equals $\sqrt{\frac{\widehat{p}_{boot}(1-\widehat{p}_{boot})}{R}}$. The precision of $\widehat{p}_{boot}$ can be increased by increasing the number of replications R. The number of replications is specified by the parameter *Replications*.

The other parameter is *Seed*, which can be used to replicate a bootstrap. The seed used by the bootstrap to generate the data sets is reported in the output.

Two technical details about the implementation of the bootstrap should be mentioned. For each bootstrap replication, the maximum likelihood estimates serve as start values. Thus, no random sets are used for the replications. To gain efficiency in terms of computation time, the iterations within a bootstrap replication are stopped when the L^2 is smaller that the original one, even if the convergence criterion or the maximum number of iterations is not reached.

5.8 Identification Issues

Sometimes LC models are not identified; that is, it may not be possible to obtain unique estimates for some parameters. Non-identification implies that different parameter estimates yield the same log-posterior or log-likelihood value. When a model is not identified, the observed information matrix $\mathbf{H}$ is not full rank, which is reported by the program. Another method to check whether a model is identified is to run the model again with different starting values. Certain model parameters are not identified if two sets of starting values yield the same $\log \mathcal{P}$ or $\log \mathcal{L}$, but different parameter estimates.

With respect to possible non-identification it should be noted that the use of priors may make identified models that would otherwise not be identified. In such situations, the prior information is just enough to uniquely determine the parameter values. Examples of such situations are LC cluster models for two polytomous nominal indicators and the LC two-factor models for four dichotomous items.

A related problem is "weak identification", which means that even though the parameters are uniquely determined, sometimes the data is not informative enough to obtain stable parameter estimates. Weak identification can be detected from the occurrence of large asymptotic standard errors. Local solutions may also result under weak identification.

Other "identification issues" are related to the order of the classes of the latent variables and the uniqueness of parameters for nominal variables. We force the order of the classes to be unique by reordering the classes according to their sizes: the first class is always the largest class. In LC factor analysis, the order between factors is based on the classification R-squared (largest first), and the order of the levels of each factor is such that the first level is larger than the last level. Parameters (γ's and β's) involving nominal variables are identified by using effect or ANOVA-type coding, which means that they sum to zero over all their indices. Note that the Parameters output also contains the redundant γ and β parameters and their standard errors.

6 Latent Gold's Output

Below, we provide the necessary technical details on the quantities presented in the various Latent GOLD output sections: Summary Information for a Model, Parameters, Profile, ProbMeans, Iteration Detail, Frequencies, Bivariate Residuals, and Classification.

6.1 Summary Information for a Model

This first part of the output section reports the number of cases (N), the total number of replications (in regression models with a case ID), the number of estimated parameters ($npar$), the number of activated constraints (in regression models with order restrictions), the seed used by the pseudo random number generator, the seed of the best start set, and the seed used by the bootstrap procedure. The last part (Variable Detail) contains information on the variables that are used in the analysis. The other four parts - Chi-squared Statistics, Log-likelihood Statistics, Classification Statistics, and Prediction Statistics - are described in more detail below.

6.1.1 Chi-squared Statistics

In LC models containing only discrete dependent or indicator variables (nominal, ordinal, count, or binomial count), the program reports chi-squared and related statistics. The three reported chi-squared measures are the likelihood-ratio chi-squared statistic L^2, the Pearson chi-squared statistic X^2, and the

Cressie-Read chi-squared statistic CR^2.[15] Let n_i denote an observed cell entry, $\widehat{m}_i$ an estimated cell entry, I the number of non-zero observed cell entries, and N the total sample size. Using these definitions, the chi-squared statistics are calculated as follows:

$$\begin{aligned} L^2 &= 2\sum_{i=1}^{I} n_i \log \frac{n_i}{\widehat{m}_i}, \\ X^2 &= \sum_{i=1}^{I} \frac{(n_i)^2}{\widehat{m}_i} - N, \\ CR^2 &= 1.8\sum_{i=1}^{I} n_i \left[\left(\frac{n_i}{\widehat{m}_i}\right)^{2/3} - 1\right]. \end{aligned}$$

An expected cell entries $\widehat{m}_i$ is obtained by:

$$\widehat{m}_i = N_{\mathbf{z}_i} f(\mathbf{y}_i|\mathbf{z}_i,\widehat{\boldsymbol{\vartheta}}) = N_{\mathbf{z}_i} \pi(\mathbf{y}_i|\mathbf{z}_i,\widehat{\boldsymbol{\vartheta}}), \tag{16}$$

i.e., the total number of cases with the same covariate values as response pattern i times the estimated multinomial probability.

The observed cell counts n_i are obtained by grouping identical cases. With replications, all records of a case have to be considered. The observed frequency table is set up by grouping cases with identical records; that is, people that have the same covariate, predictor, and exposure values, and give the same responses. In order to obtain the chi-squared statistics, we also need the totals of the relevant multinomials, which are denoted by $N_{\mathbf{z}_i}$. The full predictor, covariate, and exposure pattern defines the multinomial to which a case belongs. What we have to do is group cases with identical predictor, covariate, and exposure values. This amounts to grouping cases without taking into account their responses.

The estimated probabilities under the saturated model are obtained by dividing the observed frequencies by the corresponding totals. L^2 can also be written as

$$L^2 = 2\sum_{i=1}^{I} n_i \log \frac{n_i/N_{\mathbf{z}_i}}{\pi(\mathbf{y}_i|\mathbf{z}_i,\widehat{\boldsymbol{\vartheta}})};$$

[15] For a comparison of the performance of these statistics see Collins et. al. (1993).

that is, as minus twice the difference of the log-likelihood of the data (=the saturated model) and the log-likelihood of the model of interest.

Let $C_{\mathbf{z}}$ be the total number of different covariate patterns in the data set, C_{y_k} the number categories of y_k, and $npar$ the number of parameters in the model.[16] The number of degrees of freedom is defined by

$$df = C_{\mathbf{z}} \left(\prod_k C_{y_k} - 1 \right) - npar.$$

The chi-squared values with the corresponding df yield the asymptotic p-values, which can be used to determine whether the specified model fits the data.[17]

If the bootstrap option is used, the program also provides the estimated bootstrap p-value corresponding to the L^2 statistic, as well as its standard error. This option is especially useful with sparse tables, in which case the asymptotic p-values cannot be trusted.

The program reports the Bayesian Information Criterium (BIC), the Akaike Information Criterium (AIC), and the Consistent Akaike Information Criterium ($CAIC$) based on the L^2 and df, which is the more common formulation in the analysis of frequency tables. They are defined as

$$\begin{aligned} BIC_{L^2} &= L^2 - \log(N)\, df, \\ AIC_{L^2} &= L^2 - 2\; df, \\ CAIC_{L^2} &= L^2 - [\log(N) + 1]\;\; df. \end{aligned}$$

These information criteria weigh the fit and the parsimony of a model: the lower BIC, AIC, or $CAIC$, the better the model.

[16] A binomial count can take on a value between 0 and the total exposure T_k. C_{y_k} is therefore equal to $T_k + 1$. A Poisson count can take on any value, which means that the number of categories is in fact infinity. However, we set $C_{y_k} = \max(y_k) + 2$, which amounts to treating all values that are larger than the largest observed score as a single category.

[17] When there is missing data, computation of the number of degrees of freedom is a bit more complicated. In that case, we have to compute a separate term of the form $C_{\mathbf{z}r}(\prod_k (C_{y_k})^{\delta_{kr}} - 1)$ for each missing data pattern r, where δ_{kr} denotes whether variable y_k is observed or not in the missing data pattern concerned. The sum of these terms minus $npar$ yields df. A similar situation as with missing data occurs if cases differ with respect to their number of replications or their exposures.

6.1.2 Log-likelihood Statistics

For all models, we report the values of the log-likelihood ($\log \mathcal{L}$), the log-prior ($\log h(\boldsymbol{\vartheta})$), and log-posterior ($\log \mathcal{P}$) , where

$$\begin{aligned} \log \mathcal{L} &= \sum_i n_i \log f(\mathbf{y}_i|\mathbf{z}_i,\widehat{\boldsymbol{\vartheta}}), \\ \log \mathcal{P} &= \log \mathcal{L} + \log h(\widehat{\boldsymbol{\vartheta}}). \end{aligned}$$

In addition, the Bayesian Information Criterium (BIC), the Akaike Information Criterium (AIC), and the Consistent Akaike Information Criterium (AIC) based on the log-likelihood are reported. The latter are defined as

$$\begin{aligned} BIC_{\log \mathcal{L}} &= -2\log \mathcal{L} + (\log N)\ npar, \\ AIC_{\log \mathcal{L}} &= -2\log \mathcal{L} + 2\ npar, \\ CAIC_{\log \mathcal{L}} &= -2\log \mathcal{L} + [(\log N)\ + 1]\ \ npar. \end{aligned}$$

6.1.3 Classification Statistics

This set of statistics contains information on how well we can predict to which latent class cases belong given their observed y and z values, or, in other words, how well the latent classes are separated. Classification is based on the latent classification or posterior class membership probabilities. For response pattern i, these are calculated as follows:

$$\pi(\mathbf{x}|\mathbf{z}_i\mathbf{y}_i, \widehat{\boldsymbol{\vartheta}}) = \frac{\pi(\mathbf{x}|\mathbf{z}_i, \widehat{\boldsymbol{\vartheta}}) f(\mathbf{y}_i|\mathbf{x}\mathbf{z}_i, \widehat{\boldsymbol{\vartheta}})}{f(\mathbf{y}_i|\mathbf{z}_i,\widehat{\boldsymbol{\vartheta}})}. \tag{17}$$

In LC factor models with several factors, these are actually joint classification probabilities for all J latent variables at once. However, latent classification probabilities for a single latent variable can be obtained by collapsing these joint probabilities over the other latent variables; that is,

$$\pi(x_j|\mathbf{z}_i\mathbf{y}_i, \widehat{\boldsymbol{\vartheta}}). = \sum_{x_{j'} \neq x_j} \pi(\mathbf{x}|\mathbf{z}_i\mathbf{y}_i, \widehat{\boldsymbol{\vartheta}}). \tag{18}$$

These quantities are used to compute the estimated proportion of classifications errors (E_j) for each latent variable, as well as three R^2-type measures: the reduction of classification error ($R^2_{j,error}$), an entropy based R^2-type measure for nominal variables ($R^2_{j,entropy}$), and a "standard" $R^2_{j,variance}$. Each of these R^2 measures is based on the same kind of reduction of error structure; namely,

$$R_j^2 = \frac{\text{error}(x_j) - \text{error}(x_j|\mathbf{z}_i\mathbf{y}_i)}{\text{error}(x_j)}.$$

where $\text{error}(x_j)$ is the total error without using information on $\mathbf{z}_i$, and $\mathbf{y}_i$, and $\text{error}(x_j|\mathbf{z}_i\mathbf{y}_i)$ the error if we use all available information from case i. In $R^2_{j,error}$,

$$\begin{aligned}\text{error}(x_j) &= 1 - \max \pi(x_j|\widehat{\boldsymbol{\vartheta}}), \\ \text{error}(x_j|\mathbf{z}_i\mathbf{y}_i) &= E_j = \sum_i \frac{n_i}{N}\left[1 - \max \pi(x_j|\mathbf{z}_i\mathbf{y}_i, \widehat{\boldsymbol{\vartheta}})\right],\end{aligned}$$

in $R^2_{j,entropy}$,

$$\begin{aligned}\text{error}(x_j) &= -\sum_{x_j} \pi(x_j|\widehat{\boldsymbol{\vartheta}}) \log \pi(x_j|\widehat{\boldsymbol{\vartheta}}), \\ \text{error}(x_j|\mathbf{z}_i\mathbf{y}_i) &= -\sum_i \frac{n_i}{N} \sum_{x_j} \pi(x_j|\mathbf{z}_i\mathbf{y}_i, \widehat{\boldsymbol{\vartheta}}) \log \pi(x_j|\mathbf{z}_i\mathbf{y}_i, \widehat{\boldsymbol{\vartheta}}),\end{aligned}$$

and in $R^2_{j,variance}$,

$$\begin{aligned}\text{error}(x_j) &= \sum_{x_j} [v_{x_j} - \widehat{E}(x_j)]^2 \pi(x_j|\widehat{\boldsymbol{\vartheta}}), \\ \text{error}(x_j|\mathbf{z}_i\mathbf{y}_i) &= \sum_i \frac{n_i}{N} \sum_{x_j} [v_{x_j} - \widehat{E}(x_j)]^2 \pi(x_j|\mathbf{z}_i\mathbf{y}_i, \widehat{\boldsymbol{\vartheta}}).\end{aligned}$$

Here, v_{x_j} denotes a fixed category score for the jth latent variable and $\widehat{E}(x_j)$ its mean.

In the cluster and regression model, the latent variable is a nominal variable. The definition of $R^2_{j,variance}$ is therefore based on the qualitative variance (Magidson, 1981); that is,

$$\begin{aligned}\text{error}(x_j) &= \sum_{x_j} [1 - \pi(x_j|\widehat{\boldsymbol{\vartheta}})] \pi(x_j|\widehat{\boldsymbol{\vartheta}}), \\ \text{error}(x_j|\mathbf{z}_i\mathbf{y}_i) &= \sum_i \frac{n_i}{N} \sum_{x_j} [1 - \pi(x_j|\mathbf{z}_i\mathbf{y}_i, \widehat{\boldsymbol{\vartheta}})]\, \pi(x_j|\mathbf{z}_i\mathbf{y}_i, \widehat{\boldsymbol{\vartheta}}),\end{aligned}$$

which yields a $R^2_{j,variance}$ of the form of a Goodman and Kruskal tau-b.

The Average Weight of Evidence (AWE) criterion adds a third dimension to the information criteria described above. It weighs fit, parsimony, and the performance of the classification (Banfield and Raftery, 1993). This measure uses the so-called classification log-likelihood, which is equivalent to the complete data log-likelihood $\log \mathcal{L}^c$, i.e.,

$$\log \mathcal{L}^c = \sum_i \sum_{\mathbf{x}} n_{i\mathbf{x}} \log \pi(\mathbf{x}|\mathbf{z}_i, \boldsymbol{\vartheta}) f(\mathbf{y}_i|\mathbf{x}\mathbf{z}_i, \boldsymbol{\vartheta}).$$

AWE can now be defined as

$$AWE = -2 \log \mathcal{L}^c + 2 \left(\frac{3}{2} + \log N \right) npar.$$

The lower AWE, the better a model.

6.1.4 Prediction Statistics

In the regression model, we have a section called prediction statistics. These are based on the comparison between observed and predicted y_{it} values. The predicted value of the dependent variable given the model parameters and the predictor and covariate values of individual i at replication t, $\widehat{y}_{it}$, is defined as the weighted average of the class-specific expected values $E(y_{it}|x, \mathbf{z}_i, \widehat{\boldsymbol{\vartheta}})$. In other words,

$$\widehat{y}_{it} = E(y_{it}|\mathbf{z}_i, \widehat{\boldsymbol{\vartheta}}) = \sum_{x_1} \pi(x|\mathbf{z}_i \mathbf{y}_i, \widehat{\boldsymbol{\vartheta}}) E(y_{it}|x, \mathbf{z}_i, \widehat{\boldsymbol{\vartheta}}).$$

As can be seen, person i's posterior membership probabilities serve as weights. This yields *posterior-mean or expected a posteriori predicted values.*

The most natural predicted value for ordinal, continuous, and count variables is the estimated expected value presented above. For nominal variables, this the mode. As shown below, error measures for categorical variables may, however, also be based on the estimated probabilities instead of a single predicted value. For categorical dependent variables, we report a *prediction table* cross-classifying observed and predicted values based on modal assignment.

The average prediction error can be defined in various ways. We implemented measures based on squared error (MSE), minus the log-likelihood ($-MLL$), absolute error (MAE), and, for categorical variables, also the proportion of predictions errors (PPE). The various types of error measures can be denoted by the generic name Error(model). Except for nominal dependent variables, computation of MSE and MAE is straightforward:

$$\begin{aligned} MSE &= \frac{\sum_i cw_i \sum_t rw_{it}[y_{it} - \widehat{y}_{it}]^2}{\sum_i cw_i \sum_t rw_{it}} \\ MAE &= \frac{\sum_i cw_i \sum_t rw_{it}|y_{it} - \widehat{y}_{it}|}{\sum_i cw_i \sum_t rw_{it}} \end{aligned}$$

Here, cw_i and rw_{it} denote case and replication weights. For nominal variables, $[y_{it} - \widehat{y}_{it}]^2$ and $|y_{it} - \widehat{y}_{it}|$ are replaced by a sum over all categories: $\sum_q [I_q(y_{it}) - \pi(y_{it} = q|\mathbf{z}_i)]^2$ and $\sum_q |I_q(y_{it}) - \pi(y_{it} = q|\mathbf{z}_i)|$, where indicator variable $I_q(y_{it})$ equals 1 if $y_{it} = q$ and otherwise 0. The mean minus log-likelihood ($-MLL$) is obtained using $\widehat{y}_{it}$ as expected value in the appropriate log density function. Again, we take the average over all cases and replications. More precisely,

$$-MLL = -\frac{\sum_i cw_i \sum_t rw_{it} \log f[y_{it}|\widehat{y}_{it}]}{\sum_i cw_i \sum_t rw_{it}}.$$

With categorical variables, we replace $\log f[y_{it}|\widehat{y}_{it}]$ by $\sum_q I_q(y_{it}) \ln[\pi(y_{it} = q|\mathbf{z}_i)]$.

The general definition of the (pseudo) R^2 of an estimated model is the reduction of errors compared to the errors of a baseline model. More precisely,

$$R^2 = \frac{\text{Error(baseline)} - \text{Error(model)}}{\text{Error(baseline)}}.$$

Our baseline prediction, $\widehat{y}_0$, is the average predicted value according to the specified model; that is,

$$\widehat{y}_0 = \frac{\sum_i cw_i \sum_t rw_{it}\widehat{y}_{it}}{\sum_i cn_i \sum_t rw_{it}}.$$

Notice that in most situations, $\widehat{y}_0$ is simply the observed sample mean, or the predicted value in the intercept-only model. This is, however, not necessarily the case when restrictions are imposed on the intercept.

6.2 Parameters

In the output section Parameters, the program reports the estimates for the β and γ parameters appearing in the linear predictors η and the error variances and covariances σ, as well as the corresponding estimated asymptotic standard errors, $\widehat{se}(\beta)$, $\widehat{se}(\gamma)$, and $\widehat{se}(\sigma)$.

The significance of sets of parameters can be tested by means of the reported Wald statistic labeled *Wald*. In LC regression models, we also report a Wald statistic labeled *Wald(=)*. This statistic tests whether regression coefficients can be assumed to be equal between classes (class-independent). The general formula for a Wald statistic (W^2) is

$$W^2 = \left(\mathbf{C}'\boldsymbol{\vartheta}\right)' \left(\mathbf{C}'\,\widehat{\mathbf{H}}^{-1}\mathbf{C}\right)^{-1} \left(\mathbf{C}'\boldsymbol{\vartheta}\right),$$

where the tested set of linear constraints is: $\mathbf{C}'\boldsymbol{\vartheta} = \mathbf{0}$. The Wald test is a chi-squared test. Its number of degrees of freedom equals the number of constraints. Computation of standard errors and Wald statistics can be suppressed. This option may be useful in models with many parameters.

In the LC regression model, the parameters output contains *class-specific* R^2 values based on MSE (see also prediction statistics). In the computation of the class-specific errors, each observed value y_{it} is compared with the class-specific prediction given by the model, $E(y_{it}|x)$. The posterior membership probabilities $\pi(x|\mathbf{z}_i, \mathbf{y}_i, \widehat{\boldsymbol{\vartheta}})$ quantify the contribution case i to the error of class x. More precisely, MSE_x is obtained as

$$MSE_x = \frac{\sum_i cw_i \pi(x|\mathbf{z}_i, \mathbf{y}_i, \widehat{\boldsymbol{\vartheta}}) \sum_t rw_{it} \left[y_{it} - E(y_{it}|x)\right]^2}{\sum_i cw_i \pi(x|\mathbf{z}_i, \mathbf{y}_i, \widehat{\boldsymbol{\vartheta}}) \sum_t rw_{it}}.$$

Similarly to the overall R^2, in the computation of each class-specific R^2, we use the average $E(y_{it}|x)$ to derive a baseline error.

We also report the means and standard deviations of the regression coefficients. These are the typical fixed and random effects in multilevel, mixed, or random-coefficient models. The mean equals $\sum_x \pi(\mathbf{x})\widehat{\beta}_x$, and the standard deviation $\sqrt{\sum_x \pi(\mathbf{x})(\widehat{\beta}_x)^2 - \left[\sum_x \pi(\mathbf{x})\widehat{\beta}_x\right]^2}$.

The parameters output in the cluster and factor models contains a separate R^2 value for each indicator. These are similar to explained variances in analysis of variance and item communalities in traditional factor analysis. For the scale types ordinal, continuous, and count, these are standard explained variances; that is, they are defined as the ratio of the between-class and total variation of the y variable concened. For nominal indicators, the R^2 is based on the qualitative variance, which is the sum of the category-specific variances (see classification statistics), yielding a Goodman and Kruskal tau-b measure.

6.3 Profile

The Profile output contains transformations of the β and γ parameters that make the interpretation of the results somewhat easier. First, we report the estimated *marginal latent probabilities* or *class sizes* for each latent variable, $\widehat{\pi}(x_j)$, which are obtained as follows:

$$\widehat{\pi}(x_j) = \sum_{\mathbf{z}} \sum_{x_{j'} \neq x_j} \frac{N_{\mathbf{z}}}{N} \widehat{\pi}(\mathbf{x}|\mathbf{z}).$$

Here, $N_{\mathbf{z}}$ denotes the number of cases with covariate pattern $\mathbf{z}$. In factor, we also provide the joint marginal distribution of all latent factors.

The second part of the Profile output contains *partial conditional means or probabilities* for each y_k-x_j combination. These show the strength of the effects of the latent variables on the dependent or indicator variables using a more natural scale; that is, depending on the scale type of the y variable: probabilities, Poisson rates, or normal means. In the factor model, the Profile output also contains *joint partial conditional means or probabilities* for each indicator.

For nominal and ordinal y variables, the linear term used to obtain the partial conditional probabilities $\widehat{\pi}(y_k|x_j)^{partial}$ equals

$$\widehat{\eta}^{partial}_{y_k|x_j} = \widehat{\beta}^{0partial}_{y_k} + \widehat{\beta}^{1}_{y_k|x_j}.$$

Here, $\widehat{\beta}^{1}_{y_k|x_j}$ is the effect that we want to transform to a probability scale. The intercept term $\widehat{\beta}^{0partial}_{y_k}$ contains the main effect $\widehat{\beta}^{0}_{y_k}$, as well as all the other β terms for y_k evaluated at the mean of the other variable entering in the term concerned. When the other variable is nominal, the term concerned is simply omitted.

The resulting partial conditional probabilities show the effect of x_j on y_k on a probability scale for a person with average values on all other variables appearing in the model for y_k. Note that in models with a single latent variable, without effects of covariates on y_k, and without local dependencies involving y_k, the resulting partial conditional probabilities equal the conditional probabilities appearing in the model: $\widehat{\pi}(y_k|x_j)^{partial} = \widehat{\pi}(y_k|x_j)$. This also applies to the joint partial conditional probabilities in the factor model.

For continuous indicators, we report the normal means $\widehat{\mu}(k|x_j)^{partial}$, for counts, the Poisson rates $\widehat{\theta}(k|x_j)^{partial}$, and for binomial counts, the success probabilities $\widehat{\pi}(k|x_j)^{partial}$. These are based on linear terms of the form

$$\widehat{\eta}^{partial}_{k|x_j} = \widehat{\beta}^{0partial}_{k} + \widehat{\beta}^{1}_{k|x_j}.$$

Also *(joint) marginal conditional means or probabilities* are provided. These are obtained using the relevant model probabilities and means, and aggregating over the other variables involved in the model for the y variable concerned. To show how these are obtained, let us consider one of the more complicated cases. Suppose we have a LC factor model with covariates having direct effects on indicators and with a local dependency between nominal indicators y_1 and y_2. The marginal probability $\widehat{\pi}(y_1|x_j)$ is obtained as follows:

$$\widehat{\pi}(y_1|x_j) = \frac{\sum_{\mathbf{z}} \sum_{x_{j'} \neq x_j} \sum_{y_2} \frac{N_{\mathbf{z}}}{N} \widehat{\pi}(\mathbf{x}|\mathbf{z}) \widehat{\pi}(y_1 y_2|\mathbf{x}\mathbf{z})}{\widehat{\pi}(x_j)}.$$

As can be seen, first we compute the joint probabilities of y_1, y_2, $\mathbf{x}$, and $\mathbf{z}$, collapse these over all other variables, and divide the result by the marginal probabilities of the latent variable(s) concerned.

Standard errors for the (joint) marginal latent probabilities, and the (joint) partial and marginal means are computed by means of the delta method (see equation 15).

Latent GOLD 3.0 also provides Profile output for covariates. This is not computed using the model probabilities and means as for indicators and dependent variables, but from the posterior membership probabilities. We compute the probability of having a certain covariate value given class membership by aggregating and re-scaling posterior membership probabilities. Profile output for covariates is, in fact, re-scaled ProbMeans output (for more details, see Probmeans output below).

The (joint) partial and marginal conditional means are plotted in the Profile Plot. This makes it easy to identify differences between clusters, classes, or levels of factors.

6.4 ProbMeans

The Profile output gives the conditional distribution of a y variable given an individual's score on a latent variable. However, it may also be of interest to look at the distribution of the latent variable for a certain level of an observed variable. Measures of these kinds appear in the ProbMeans output and the associated Uni-, Bi-, and Tri-Plots.

The probability of being classified in a certain latent class given a person's score on a z variable - $\widehat{\pi}(x_j|z_\ell)$ - or y variable - $\widehat{\pi}(x_j|y_k)$ - can be obtained by aggregating and collapsing the latent classification probabilities defined in equation (18) in the appropriate manner. More precisely,

$$\widehat{\pi}(x_j|y_k) = \sum_{all\ y_{ik}=y_k} \pi(x_j|\mathbf{z}_i\mathbf{y}_i, \widehat{\boldsymbol{\vartheta}})\, n_i/N_{y_k}.$$

Here, N_{y_k} denotes the total number of persons in level y_k of indicator k. A similar formula can be used to obtain the probabilities for the covariates, $\widehat{\pi}(x_j|z_\ell)$.

In LC cluster and regression models, the above class membership probabilities are plotted in Uni- and Tri-plots (Vermunt and Magidson 2000; Magidson and Vermunt, 2001). Similar plots have been proposed by Van der Ark and Van der Heijden (1998) and Van der Heijden, Gilula, and Van der Ark (1999) for standard LC and latent budget models.

In LC factor models, we work with factor means rather than class membership probabilities; that is,

$$\widehat{E}(x_j|y_k) = \sum_{x_j} v_{x_j}\widehat{\pi}(x_j|y_k).$$

Here v_{x_j} is the category score of a level of the jth factor. The same formula can be used to obtain a factor mean for a certain covariate level. Note that in LC factor analysis we use scores ranging from 0 and 1 for the factor levels. This implies that with a dichotomous factor, a factor mean equals the probability of being in level (class) 2.

The factor means for each covariate level and for each indicator level are plotted in the Uni- and Bi-plots of our LC factor models (Vermunt and Magidson 2000; Magidson and Vermunt, 2001).

A nice feature of the ProbMeans output is that it describes the relationships between the latent variable(s) and all variables selected as indicators or covariates. This means that even if a certain covariate effect is fixed to zero, one obtains its ProbMeans information. This feature is exploited in the "inactive covariates method". An advantage of working with inactive covariates is that they do not influence the obtained solution.

6.5 Frequencies

Latent GOLD reports observed and estimated cell frequencies (n_i and $\widehat{m}_i$), as well as standardized residuals ($\widehat{r}_i$), if all the y variables are discrete; that is, in LC cluster and factor models with indicators that are nominal, ordinal, or counts and in LC regression models with a dependent variable that is nominal, ordinal, or a (Poisson/binomial) count. This is also the situation in which chi-squared statistics are provided.

The definition of the estimated cell entries was given in equation (16). The standardized residuals are defined as

$$\widehat{r}_i = \frac{n_i - \widehat{m}_i}{\sqrt{\widehat{m}_i}}.$$

Note that $(\widehat{r}_i)^2$ is cell i's the contribution to the X^2 statistic.

6.6 Bivariate Residuals

As was explained when presenting the various types of LC models implemented in Latent GOLD, one of the main assumptions in LC analysis is the local independence assumption. The reported bivariate residuals for z-y and y-y pairs provide very direct checks for this assumption. They indicate how similar the estimated and observed bivariate associations are. The measures can be interpreted as lower bound estimates for the improvement in fit (L^2 or $-2\ \log \mathcal{L}$) if the corresponding local independence constraints were relaxed.[18]

Actually, these measures, which are sometimes referred to as modification indices, are Lagrange-type chi-squared statistics. *Likelihood-ratio* statistics compare the log-likelihood values of a restricted and an unrestricted model, *Wald* statistics estimate the decrease of the log-likelihood value if constraints are imposed in the unrestricted model, and *Lagrange-multiplier* statistics estimate the increase of the log-likelihood value if constraints are relaxed in

[18]Similar but not completely equivalent approaches for detecting local dependencies have been proposed by Bartholomew and Tzamourani (1999), Reiser (1996), Reiser and Lin (1999), and Glas (1999).

the restricted model or, equivalently, if parameters are added to the restricted model (Buse, 1982). Even though the bivariate residuals reported in Latent GOLD are similar to Lagrange-multiplier tests, they are not exactly the same because they do not take into account the dependencies between the parameters in the restricted model and the new set.

Suppose that a particular local dependency contains R parameters denoted by ϑ_r^{local}. The general formula for our Lagrange-type residuals is:

$$T^2 = \frac{1}{R} \sum_{r=1}^{R} \left(\frac{\partial \log \mathcal{P}}{\partial \vartheta_r^{local}} \right)^2 / \left(\frac{\partial^2 \log \mathcal{P}}{\partial^2 \vartheta_r^{local}} \right).$$

Thus, for each of the R parameters, we compute the first- and second-order derivative of the log-posterior. Dependencies between the parameters in a set, as well as with the main effects, are dealt with by including redundant parameters in a set. To further reduce the computational burden, in some situations, we use an approximation for the second-order derivatives. Note that because of the factor $\frac{1}{R}$, the measure should be interpreted as the estimated model fit improvement per extra parameter.

Latent GOLD reports z-y residuals for all types of y variables. These are based on the (unrestricted) $\beta^2_{y_k z_\ell}$ terms. The y-y residuals for pairs of categorical (nominal or ordinal) variables are based on $\beta^0_{y_k y_{k'}}$ terms, and for pairs of continuous variables on the error covariances $\sigma_{kk'}$. For pairs of Poisson counts, we currently do not have bivariate residuals.

Finally, y-y residuals between variables of different scale types are reported. Residuals for categorical-continuous pairs assume that the categorical variable is used as a covariate in the model for the continuous variable. The same method is used in the case of pairs consisting of a categorical and a Poisson variable. As was explained in the subsection on LC cluster models for mixed mode data, these kinds of local dependencies can only be included in an indirect manner in the model. Continuous-Poisson residuals are not reported.

6.7 Classification and Related Output

The Classification output section contains the classification information for each response pattern i. Classification is based on the latent classification or posterior class membership probabilities described equation (18).

In LC cluster models and LC regression models, we have only one latent variable, which means that we can also denote these probabilities by $\pi(x_1|\mathbf{z}_i\mathbf{y}_i, \widehat{\boldsymbol{\vartheta}})$. These quantities can directly be used to determine which class someone belongs to. More precisely, subjects are assigned to the latent class with the highest latent classification probability. This method of assignment is sometimes referred to as empirical Bayes modal (EBM) or modal a posteriori (MAP) estimation (Skrondal, 1996).

In LC factor models, there may be more that one latent variable. In addition, the factors are ordinal or discrete interval variables that can be used as approximations of continuous latent variables with unknown distributions. Therefore, in LC factor models, Latent GOLD not only reports the classification probabilities and the modal allocation for each factor, but also the factor scores or factor means. The factor scores for response pattern i are obtained by

$$E(x_j|\mathbf{z}_i\mathbf{y}_i, \widehat{\boldsymbol{\vartheta}}) = \sum_{x_j} v_{x_j}\pi(x_j|\mathbf{z}_i\mathbf{y}_i, \widehat{\boldsymbol{\vartheta}}).$$

Here, v_{x_j} denote the category scores (ranging from 0 to 1) for the levels of the latent variable concerned. Factor scores computed in this manner are sometimes referred to as empirical Bayes (EB) or expected a posterior (EAP) estimators (Skrondal, 1996).

Classification can also be based on covariates only. This involves using the model probabilities $\pi(\mathbf{x}|\mathbf{z}_i, \widehat{\boldsymbol{\vartheta}})$, sometimes referred to as prior probabilities, as classification probabilities. The same classification rules can be applied as with the posterior class membership probabilities.

An issue related to classification is the estimation of individual regression coefficients. The posterior-mean or expected a posteriori estimate of a particular regression coefficient for case i is defined as follows:

$$\widehat{\beta}_i = \sum_x \pi(\mathbf{x}|\mathbf{z}_i\mathbf{y}_i, \widehat{\boldsymbol{\vartheta}})\,\widehat{\beta}_x\,;$$

that is, as a weighted average of the class-specific coefficients. These estimates are similar to the individual coefficients obtained in multilevel or hierarchical (Bayes) models.

Bibliography

Agresti, A. (2002). Categorical data analysis. Second Edition, New York: Wiley.

Banfield, J.D., and Raftery, A.E. (1993). Model-based Gaussian and non-Gaussian clustering. Biometrics, 49, 803-821.

Bartholomew, D.J., and Knott, M. (1999). Latent variable models and factor analysis. London: Arnold.

Bartholomew, D.J., and Tzamourani, P. (1999). The goodness-of-fit of latent trait models in attitude measurement. Sociological Methods and Research, 27, 525-546.

Böckenholt, U. (1993). A latent class regression approach for the analysis of recurrent choices. British Journal of Mathematical and Statistical Psychology, 46, 95-118.

Böhning, D. (2000). Computer-assisted analysis of mixtures: meta-analysis, disease mapping and others. London: Chapmann & Hall.

Buse, A. (1982). The likelihood ratio, Wald, and Lagrange multiplier tests: An expository note. The American Statistician, 36, 153-157.

Clogg, C.C. (1981). New developments in latent structure analysis. D.J. Jackson and E.F. Borgotta (eds.), Factor analysis and measurement in sociological research, 215-246. Beverly Hills: Sage Publications.

Clogg, C.C. (1988). Latent class models for measuring. R. Langeheine and J.Rost (eds.), Latent trait and latent class models. New York, London: Plenum Press.

Clogg, C.C. (1995). Latent class models. G.Arminger, C.C.Clogg, and M.E.Sobel (eds.), Handbook of statistical modeling for the social and behavioral sciences, 311-359. New York: Plenum Press.

Clogg, C.C., Rubin, D.R., Schenker, N., Schultz, B., Weidman, L. (1991). Multiple imputation of industry and occupation codes in census public-use samples using Bayesian logit regression. Journal of the American Statistical Association, 86, 68-78.

Collins, L.M., Fidler, P.F., Wugalter, S.E., and Long, L.D. (1993). Goodness-of-fit testing for latent class models. Multivariate Behavioral Research, 28, 375-389.

Dayton, C.M. (1998). Latent class scaling analysis. Thousand Oakes: Sage Publications.

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

Dempster, A.P., Laird, N.M., and Rubin, D.B. (1977). Maximum likelihood estimation from incomplete data via the EM algorithm (with discussion). Journal of the Royal Statistical Society, Ser. B., 39, 1-38.

Dillon, W.R., and Kumar, A. (1994). Latent structure and other mixture models in marketing: A integertaive survey and overview. R.P. Bagozzi (ed.), Advanced methods of Marketing Research, 352-388, Cambridge: Blackwell Publishers.

Everitt, B.S. (1988). A finite mixture model for the clustering of mixed-mode data. Statistics and Probability Letters, 6, 305-309.

Everitt, B.S., and Hand, D.J. (1981). Finite mixture distributions. London: Chapman and Hall.

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

Formann, A.K., and Kohlmann,T. (1998). Structural latent class models. Sociological Methods and Research, 26, 530-565.

Gill, P.E., Murray, W., and Wright, M.H. (1981). Practical optimization. London: Academic Press.

Glas, C.A.W. (1999). Modification indices for the 2PL and the nominal response model. Psychometrika, 64, 273-294.

Gelman, Andrew, Carlin, John B., Stern, Hal .S., and Robin, Donald B. (1995). Bayesian data analysis. London: Chapman & Hall.

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. (1979). Simple models for the analysis of association in cross-classifications having ordered categories. Journal of the American Statistical Association, 74, 537-552.

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 observations. C.Clogg (ed.), Sociological Methodology 1988, 193-211. San Francisco: Jossey-Bass.

Hagenaars, J.A. (1988). Latent structure models with direct effects between indicators: local dependence models. Sociological Methods and Research, 16, 379-405.

Hagenaars, J.A. (1990). Categorical longitudinal data - loglinear analysis of panel, trend and cohort data.. Newbury Park: Sage.

Hagenaars, J.A. (1993). Loglinear models with latent variables. Newbury Park: CA: Sage.

Hagenaars. J.A. and McCutcheon.A.L. (2002). Applied latent class analysis, Cambridge: Cambridge University Press.

Heinen, T. (1996). Latent class and discrete latent trait models: similarities and differences. Thousand Oakes: Sage Publications.

Hunt, L, and Jorgensen, M. (1999). Mixture model clustering using the MULTIMIX program. Australian and New Zealand Journal of Statistics, 41, 153-172.

Kamakura, W.A., Wedel, M., and Agrawal, J. (1994). Concomitant variable latent class models for the external analysis of choice data. International Journal of Marketing Research, 11, 541-464.

Kolhmann T., and Formann A.K. (1997). Using latent class models to analyze response patterns in epidemiologic mail surveys. J.Rost and R.Langeheine (eds.). Applications of latent trait and latent class models in the social sciences. Muenster: Waxmann Verlag.

Laird, N. (1978). Nonparametric maximum likelihood estimation of a mixture distribution. Journal of the American Statistical Association, 73, 805-811.

Land, C.K., McCall, P.L., and Nagin, D.S. (1996). A comparison of Poisson, negative binomial, and semiparametric mixed Poisson regression models. Sociological Methods and Research, 24, 387-442.

Langeheine, R., Pannekoek, J., and Van de Pol, F. (1996). Bootstrapping goodness-of-fit measures in categorical data analysis. Sociological Methods and Research, 24, 492-516.

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

Lawrence C.J., Krzanowski, W.J. (1996). Mixture separation for mixed-mode data. Statistics and Computing., 6, 85-92.

Lazarsfeld, P.F., and Henry, N.W. (1968). Latent structure analysis. Boston: Houghton Mill.

Little, R.J., and Rubin, D.B. (1987). Statistical analysis with missing data. New York: Wiley.

Mare, R.D. (1994). Discrete-time bivariate hazards with unobserved heterogeneity: a partially observed contingency table approach. P.V. Marsden (ed.), Sociological Methodology 1994, 341-385. Oxford: Basil Blackwell.

Magidson, J. (1981). Qualitative variance, entropy, and correlation ratios for nominal dependent variables. Social Sciences Research, 10, 177-194.

Magidson, J. and Vermunt, J.K. (2001). Latent class factor and cluster models, bi-plots and related graphical displays, Sociological Methodology, 31, 223-264.

Magidson, J., and Vermunt, J.K, (2003) Latent class analysis. D. Kaplan (ed.), Handbook of Quantitative Methodology for the Social Sciences, Sage Publications.

McCullagh, P., and Nelder, J.A. (1983). Generalized linear models. London: Chapman & Hall, second edition 1989.

McCutcheon, A.L. (1987). Latent class analysis, Sage University Paper. Newbury Park: Sage Publications.

McCutcheon, A.L. (1988). Sexual morality, pro-life values and attitudes toward abortion. Sociological Methods and Research, 16, 256-275.

McCutcheon, A.L. (1994). Latent logit models with polytomous effect variables. A. von Eye and C.C Clogg (eds.), Latent variable analysis. Newbury park, CA: Sage.

McLachlan, G.J., and Basford, K.E. (1988). Mixture models: inference and application to clustering. New York: Marcel Dekker.

McLachlan, G.J., and Krishnan, T. (1997). The EM algorithm and extensions. New York: John Wiley & Sons, Inc.

McLachlan, G.J., and Peel, D. (2000). Finite Mixture models. New York: John Wiley & Sons, Inc.

Moustaki, I. (1996). A latent trait and a latent class model for mixed observed variables. The British Journal of Mathematical and Statistical Psychology, 49, 313-334.

Patterson, B.H., Dayton,C.M.; Graubard, B.I. (2002) "Latent class analysis of complex sample survey data: application to dietary data". Journal of the American Statistical Association, 97, 721-728.

Reiser, M. (1996). Analysis of residuals for the multinomial item response model. Psychometrika, 61, 509-528.

Reiser, M., and Lin, Y. (1999). Goodness-of-fit test for the latent class model when expected frequencies are small. M.Sobel and M.Becker (eds.), Sociological Methodology 1999, 81-111. Boston: Blackwell Publishers.

Rost, J., and Langeheine, R. (1997). Applications of latent trait and latent class models in the social sciences. Muenster: Waxmann Verlag.

Schafer, J.L. (1997). Analysis of incomplete categorical data. London: Chapman & Hall.

Shockey, J. (1988). Latent class analysis: an introduction to discrete data models with unobserved variables. J.S. Long (red.), Common problems in quantitative research. Newbury Park, CA: Sage.

Skrondal, A. (1996). Latent trait, multilevel and repeated measurement modelling with incomplete data of mixed measurement levels. Phd. Dissertation. Section of Medical Statistics, University of Oslo, Norway.

Titterington, D.M., Smith, A.F., and Makov, U.E. (1985). Statistical analysis of finite mixture dsitributions. Chichester: John Wiley & Sons.

Uebersax, J.S. (1993). Statistical modeling of expert ratings on medical treatment appropriateness. Journal of the American Statistical Association, 88, 421-427.

Uebersax, J.S. (1997). Analysis of student problem behaviors with latent trait, latent class, and related probit mixture models. J.Rost and R.Langeheine (eds.). Applications of latent trait and latent class models in the social sciences. Muenster: Waxmann Verlag.

Van der Ark, L.A., and Van der Heijden, P.G.M. (1998). Graphical display of latent budget and latent class analysis, with special reference to correspondence analysis. J. Blasius and M.Greenacre (eds.) Visualization of categorical data. Boston: Academic Press.

Van der Heijden, P.G.M., Dessens, J., and Böckenholt, U. (1996). Estimating the concomitant-variable latent class model with the EM algorithm. Journal of Educational and Behavioral Statistics, 5, 215-229.

Van der Heijden P.G.M., Gilula, Z., and Van der Ark, L.A. (1999). On a relationship between joint correspondence analysis and latent class analysis. M.Sobel and M.Becker (eds.), Sociological Methodology 1999, 81-111. Boston: Blackwell Publishers.

Vermunt, J.K. and Van Dijk. L. (2001). A nonparametric random-coefficients approach: the latent class regression model. Multilevel Modelling Newsletter, 13, 6-13.

Vermunt, J.K. (1997). Log-linear models for event histories. Thousand Oakes: Series QASS, vol 8. Sage Publications.

Vermunt, J.K. (2001) The use restricted latent class models for defining and testing nonparametric and parametric IRT models. Applied Psychological Measurement, 25, 283-294.

Vermunt, J.K.. (2002). A general latent class approach for dealing with unobserved heterogeneity in the analysis of event history data. J.A. Hagenaars and A.L. McCutcheon (eds.), Applied latent class analysis, 383-407. Cambridge: Cambridge University Press.

Vermunt, J.K., and J. Magidson (2000) Graphical displays for latent class cluster and latent class factor models. W. Jansen and J.G. Bethlehem (eds.), Proceedings in Computational Statistics 2000, 121-122. Statistics Netherlands. ISSN 0253-018X.

Vermunt, J.K., and Magidson, J. (2002). Latent class cluster analysis. J.A. Hagenaars and A.L. McCutcheon (eds.), Applied latent class analysis, 89-106. Cambridge: Cambridge University Press.

Vermunt, J.K. and Magidson, J. (2003). Latent class analysis. M. Lewis-Beck, A. Bryman, and T.F. Liao (eds.), Encyclopedia of Research Methods for the Social Sciences. NewBury Park: Sage Publications, Inc.

Wedel, M., and DeSarbo, W.S (1994). A review of recent developments in latent class regression models. R.P. Bagozzi (ed.), Advanced methods of Marketing Research, 352-388, Cambridge: Blackwell Publishers.

Wedel, M., and Kamakura, W.A. (1998). Market segmentation: Concepts and methodological foundations. Boston: Kluwer Academic Publishers.

Wedel, M., Kamakura, W.A., DeSarbo, W.S., and Ter Hofstede, F. (1995). Implications of asymmetry, nonproportionality, and heterogeneity in brand switching from piece-wise exponential mixture hazard models. Journal of Marketing Research, 32, 457-462.

Wolfe, J.H. (1970). Pattern clustering by multivariate cluster analysis. Multivariate Behavioral Research, 5, 329-350.