Journal of Classification 19:277-302 (2002)
DOI: 10.1007/s00357-001-0046-6

Constrained Latent Class Analysis of Three-Way Three-Mode Data

Michel Meulders
Katholieke Universiteit Leuven, Belgium

Paul De Boeck
Katholieke Universiteit Leuven, Belgium

Peter Kuppens
Katholieke Universiteit Leuven, Belgium

Iven Van Mechelen
Katholieke Universiteit Leuven, Belgium

Abstract: The latent class model for two-way two-mode data can easily be extended to the case of three-way three-mode data as a tool to cluster the elements of one mode on the basis of two other modes simultaneously. However, as the number of manifest variables is typically very large in this type of analysis, the number of conditional probabilities rapidly increases with the number of latent classes, which may lead to an overparameterized model. To solve this problem, we introduce a class of constrained latent class models in which the conditional probabilities are a nonlinear function of basic parameters that pertain to each of the two modes. The models can be regarded as a probabilistic extension of related deterministic models or as a generalization of related probabilistic models. For parameter estimation, an EM algorithm can be used to locate the posterior mode, and a Gibbs sampling algorithm can be used to compute a sample of the posterior distribution. Furthermore, model selection criteria and measures to check the fit of the model are discussed. Finally the models are applied to study the types of reactions that occur when one is angry at a person in a certain situation.

Keywords: Three-way data, Bayesian analysis, Constrained latent class analysis.

Note: The research reported in this paper was supported by GOA/2000/02 awarded to Paul De Boeck and Iven Van Mechelen.

Correspondence concerning this paper should be addressed to Michel Meulders, Department of Psychology, Tiensestraat 102, B-3000 Leuven, Belgium; email: Michel.Meulders@psy.kuleuven.ac.be.

1. Introduction

Latent class analysis (LCA) is a well-known technique that is often used for analyzing two-way two-mode data. In particular, LCA allows to cluster the elements of one mode (e.g. persons) on the basis of the elements of the other mode (e.g. behaviors) into homogeneous groups. The key assumption underlying this classification is the existence of one or more discrete latent variables which account for the correlations between the manifest variables. In this regard, Magidson and Vermunt (2001) distinguish between latent class (LC) *cluster* analysis, which involves only one discrete latent variable and LC *factor* analysis, which involves several binary latent variables.

For the analysis of three-way three-mode data one could extend the latent class approach in a straightforward manner so that it can serve as a tool to cluster the elements of one mode on the basis of the elements of the other two modes (e.g. behaviors observed in different situations) simultaneously (Hunt & Basford, 1999). As the number of manifest variables is typically very large in this type of analysis the number of conditional response probabilities rapidly increases when more latent classes are assumed. This is even more the case if the manifest variables are polytomous. As a result, models that involve many latent classes can be overparameterized, which may cause a failure to fit the data (Formann, 1992).

A general solution to this problem is to build more parsimonious models by constraining the model's parameters. This route is taken in linear logistic latent class analysis (Formann, 1985, 1989, 1992) by specifying the log odds of conditional response probabilities or of latent class sizes as a linear function of more basic parameters. As with linearly constrained logistic latent trait models (Fischer, 1983) this approach allows one to investigate to what extent the parameters of the unconstrained model can be modeled by predictor variables that are assumed to be relevant indicators of these parameters. However, in the absence of relevant indicators, it may be interesting to consider other types of constraints.

In the present paper we discuss a new class of constrained latent class models that is particularly suited to analyze three-way three-mode data. Several assumptions are made at the basis of our approach:

1. One of the modes defines the entities to be classified whereas the elements of the other two modes define the basis for this classification. For instance, *persons* can be classified on the basis of the *responses* they show in various *situations*. In this case, the data are

three-way three-mode data on persons showing responses in situations. Responses and situations will be called the classification-basis modes.

2. Each of the latent classes is characterized by a conditional probability for each pair of elements of the two modes that serve as a basis for the classification.

3. The conditional probabilities associated with each pair of elements from the classification-basis modes are not the basic parameters, rather they are constrained to be a function of single-mode parameters associated with the elements of the classification-basis modes. The constraints of the models to be introduced follow from the function that relates the conditional probabilities to the basic parameters. Typical of the models we will discuss is that the basic parameters can be considered as probabilities and are therefore in the [0,1]interval.

The outline of the paper is as follows: First, we formally discuss unconstrained latent class models. Second, we present a class of constrained latent class models and discuss the relationships with existing models. Third, we describe algorithms to obtain point estimates and posterior intervals of the parameters included in the model. Fourth, we deal with model selection and model checking. Fifth, we apply the models to study the types of reactions that occur in answer to being angry on a person in a certain situation. Finally, we discuss some topics for future research.

2. Models

1.1 Unconstrained Latent Class Model for Three-way Three-mode Data

Let D_{ijk} equal 1 if person i $(i = 1, \ldots, I)$ displays behavior k $(k = 1, \ldots, K)$ in situation j $(j = 1, \ldots, J)$ and 0 otherwise, and let d_{ijk} denote a specific realization. Furthermore, let the vector $\mathbf{d}_i$ comprise the observations for person i. To explain the data, the unconstrained latent class model (ULCM) makes three assumptions:

1. Each entity of the classification mode belongs to one of Q mutually exclusive latent classes. To formalize this, let Z_{iq} $(i = 1, \ldots, I; q = 1, \ldots, Q)$ equal 1 if person i belongs to class q and 0 otherwise. Furthermore, the parameters $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_Q)$ (with $\sum_q \xi_q = 1$)

represent the probabilities that a randomly selected person belongs to each of the classes.

2. Persons in a certain latent class have a specific conditional probability to display a particular behavior in a particular situation, that is, $P(D_{ijk} = 1|Z_{iq} = 1) = \pi_{jkq}$. The total set of conditional probabilities is further denoted as $\boldsymbol{\pi}$ and the set of conditional probabilities associated with class q is denoted as $\boldsymbol{\pi}_q$.

3. Observations within a class are stochastically independent, which implies that

$$p(\mathbf{d}_i|Z_{iq} = 1, \boldsymbol{\pi}_q) = \prod_j \prod_k \pi_{jkq}^{d_{ijk}} (1 - \pi_{jkq})^{1-d_{ijk}}.$$

From the above assumptions it follows that the probability to observe pattern $\mathbf{d}_i$ equals

$$p(\mathbf{d}_i|\boldsymbol{\xi}, \boldsymbol{\pi}) = \sum_q \xi_q \prod_j \prod_k \pi_{jkq}^{d_{ijk}} (1 - \pi_{jkq})^{d_{ijk}}.$$

1.2 Constrained Latent Class Models for Three-way Three-mode Data

Product Model. In the absence of a set of substantive predictors that could be hypothesized to explain the conditional probabilities, one could use indicator variables for the situation mode and the response mode as predictors, which yields the simple assumption that π_{jkq} can be decomposed as a function $\phi(\cdot)$ of two parameters that refer to each of the modes, that is, $\pi_{jkq} = \phi(\sigma_{jq}, \rho_{kq})$. Depending on the range of the basic parameters σ and ρ one can specify the function $\phi(\cdot)$ to make sure that the conditional probabilities are in the $[0, 1]$ interval. For instance, in the case of continuous basic parameters, one could assume as in linear logistic latent class analysis (Formann, 1985, 1989, 1992) that $\text{logit}(\pi_{jkq}) = \sigma_{jq} + \rho_{kq}$. Alternatively, one could require that also the basic parameters are in the $[0, 1]$ interval so that also these parameters can be interpreted as probabilities. For example, one may assume that $\pi_{jkq} = \rho_{jq}\sigma_{kq}$. This model will further be denoted as the product model (PM). Note that, similar to LC cluster models (Magidson and Vermunt, 2001), the PM only involves one latent variable. In the present paper the PM serves as a building block for other more complex models because it has the following elegant interpretative and mathematical properties:

First, when using the PM to explain whether persons would display a certain hostile behavior in a certain frustrating situation, the parameters have a straightforward *interpretation*. The situation parameter represents the extent to which a situation elicits hostility; whereas, the response parameter represents the a priori activation level of the response. An interpretation that is in line with this concept is that the situation has an activation force, and that behaviors each have an a priori activation level. When a person encounters a situation, the situational interpretation, as captured in the situational parameter, becomes a multiplier of the behavior activation. Hence, the product model implies that the probability to display a behavior in a situation increases as the situation has a higher activation force (i.e. higher value of σ_{jq}) and if the behavior has a higher a priori activation level (i.e. higher value of ρ_{kq}). Latent classes differ from one another in terms of their basic parameters. Each latent class corresponds to one hypothetical *basic person type* with a situation-response profile labeled *basic profile*. These basic profiles are simply obtained by multiplying the parameters attached to a situation and a response. As will be seen, the basic profiles of the PM may serve as a building block for more complex profiles that will be allowed in an extension of the PM.

Second, the PM has the interesting *mathematical* property that it can be derived from a latent process in which the observed data are considered to be a specific mapping of latent data. As a result, the computations involved in algorithms for parameter estimation are simplified.

More specifically, let the binary latent variables X_{ki}^{jq} and Y_{ji}^{kq} be defined as

$$p(x_{ki}^{jq}|z_{iq}, \sigma_{jq}) = [(\sigma_{jq})^{x_{ki}^{jq}}(1-\sigma_{jq})^{1-x_{ki}^{jq}}]^{z_{iq}}[(0)^{x_{ki}^{jq}}(1)^{1-x_{ki}^{jq}}]^{1-z_{iq}} \quad (1)$$

and

$$p(y_{ji}^{kq}|z_{iq}, \rho_{kq}) = [(\rho_{kq})^{y_{ji}^{kq}}(1-\rho_{kq})^{1-y_{ji}^{kq}}]^{z_{iq}}[(0)^{y_{ji}^{kq}}(1)^{1-y_{ji}^{kq}}]^{1-z_{iq}}, \quad (2)$$

which means that X_{ki}^{jq} and Y_{ji}^{kq} are Bernoulli distributed if person i belongs to latent class q, and are 0 with probability 1 otherwise. Note that in (1) and (2) it is assumed that 0^0 equals 1. Furthermore, by defining each observed variable as a specific mapping of latent variables, namely,

$$D_{ijk} = 1 \Leftrightarrow (X_{ki}^{jq} = 1) \wedge (Y_{ji}^{kq} = 1) \wedge (Z_{iq} = 1), \quad (3)$$

it may be derived that $P(D_{ijk} = 1|Z_{iq} = 1, \boldsymbol{\sigma}, \boldsymbol{\rho})$ equals

$$\sum_{\{(x,y)\}} \Big[P(D_{ijk} = 1 | x_{ki}^{jq}, y_{ji}^{kq}, Z_{iq} = 1) \times p(x_{ki}^{jq} | Z_{iq} = 1, \sigma_{jq}) p(y_{ji}^{kq} | Z_{iq} = 1, \rho_{kq}) \Big] = \sigma_{jq}\rho_{kq}. \quad (4)$$

Note that, in (5) $P(D|X, Y, Z)$ is a binary-valued deterministic function of latent variables X, Y and Z, which is completely defined by the mapping rule in (3).

The introduction of latent variables X and Y into the model enhances parameter estimation as the joint distribution of observed and latent variables $p(D, X, Y, Z|\boldsymbol{\sigma}, \boldsymbol{\rho}, \boldsymbol{\xi})$ has a simple mathematical structure. As a result, an EM algorithm with a closed form solution for the parameters in the M-step can be used to locate the mode(s) of the posterior distribution.

Extension of the product model. With the PM we seek Q homogeneous classes of persons that are each characterized by a specific probabilistic situation-response profile (i.e. $\pi_{jkq} = \sigma_{jq}\rho_{kq}$). A natural extension of the PM is to define additional latent classes and associated situation-response profiles by combining the Q basic profiles in a disjunctive way. This model will be denoted as the disjunctive model (DM). As one may apply the disjunctive combination to each subset of the original Q classes, the DM has $2^Q = T$ classes denoted as $\mathbf{z}_t$ $(t = 1, \ldots, T)$ with each of the elements Z_{tq} being 0 or 1. Note that the DM contains more classes than the PM which means that it may better capture individual differences, while still being a parsimonious model, since the number of basic parameters does not increase. Similar to LC factor models (Magidson and Vermunt, 2001) the DM explains associations between observed variables on the basis of several binary latent variables.

To combine the latent profiles in a disjunctive way, we use the mapping rule

$$D_{ijk} = 1 \Leftrightarrow \exists q : (X_{ki}^{jq} = 1) \wedge (Y_{ji}^{kq} = 1) \wedge (Z_{iq} = 1). \quad (5)$$

Under the assumption that the latent variables $\mathbf{X}_{ki}^{j} = (X_{ki}^{j1}, \ldots, X_{ki}^{jQ})$ and $\mathbf{Y}_{ji}^{k} = (Y_{ji}^{k1}, \ldots, Y_{ji}^{kQ})$ given $\mathbf{z}_i$ are independent Bernoulli variables, one may derive, using a formula similar to (5), that for the DM

$$
\begin{aligned}
&P(D_{ijk} = 1|\mathbf{z}_i, \boldsymbol{\sigma}, \boldsymbol{\rho}) \\
&= \sum_{\mathbf{x}} \sum_{\mathbf{y}} P(D_{ijk} = 1|\mathbf{z}_i, \mathbf{x}_{ki}^j, \mathbf{y}_{ji}^k) p(\mathbf{x}_{ki}^j|\mathbf{z}_i, \boldsymbol{\sigma}) p(\mathbf{y}_{ji}^k|\mathbf{z}_i, \boldsymbol{\rho}) \\
&= 1 - \prod_q (1 - \sigma_{jq}\rho_{kq}z_{iq}). \qquad (6)
\end{aligned}
$$

As may be seen from (6), combining the basic profiles in a disjunctive manner implies a higher response probability than can be expected on the basis of each of the basic profiles. To illustrate, consider a model with two basic person types $z_1 = (1, 0)$ and $z_2 = (0, 1)$ who have probabilities π_{jk1} and π_{jk2} to display behavior k in situation j, respectively. The response probability of a third combined person type $z_3 = (1, 1)$ then equals $\pi_{jk1} + \pi_{jk2} - \pi_{jk1}\pi_{jk2} \geq \max(\pi_{jk1}, \pi_{jk2})$.

Furthermore, we also note that there is a latent class which has all elements z_{iq} equal to zero. This class is usually empty because it contains only persons who have answered a zero for each situation-response combination. Therefore, this latent class may be discarded in order to obtain a more parsimonious model.

Relations to other models. The DM is formally related to three other models that have been proposed for the analysis of binary data. First, if one allows only one latent class which has all elements Z_{iq} $(q = 1, \ldots, Q)$ equal to 1, the disjunctive probability matrix decomposition (PMD) model for the analysis of three-way three-mode binary data is obtained (Maris, De Boeck, & Van Mechelen, 1996; Meulders, De Boeck, Van Mechelen, Gelman, & Maris, 2001; Meulders, De Boeck, & Van Mechelen, 2001). This model does not allow for any individual differences as all persons belong to the same latent class. Second, if the entities of one mode are classified on the basis of the elements of only one other mode then the multiple classification latent class model (MCLCM) is obtained (Maris, 1999). Third, the model presented in this paper may be regarded a probabilistic variant of the INDCLAS model (Leenen, Van Mechelen, De Boeck, & Rosenberg, 1999), which was proposed for the analysis of three-way three-mode data. In particular, this model assumes that the parameters σ and ρ can only be 0 or 1 and that the latent variables X and Y are realized only once for pairs (j, q) and (k, q), respectively, that is, $X_{11}^{jq} = \ldots = X_{KI}^{jq}$ $(j = 1, \ldots, J; q = 1 \ldots, Q)$ and $Y_{11}^{kq} = \ldots = Y_{JI}^{kq}$ $(k = 1, \ldots, K; q = 1, \ldots, Q)$.

Finally, we note that, unlike in the present model, in PMD, MCLCM

and INDCLAS models, the realizations of the latent variables X and Y usually have a specific interpretation. This implies that mapping rules other than the disjunctive one may be meaningful. For example, in the context of situation-response data X_{ki}^{jq} may be used to indicate whether or not person i perceives a certain *latent feature* q in situation j when judging pair (j,k) and Y_{ji}^{kq} may be used to indicate whether or not latent feature q is linked to response k when person i judges pair (j,k). Subsequently, the mapping rule is used to determine whether the person will display the response in the situation given the latent realizations X and Y. A disjunctive rule, for instance, implies that a person will display the response in the situation when there is *at least one* feature perceived in the situation that is also linked to the response. A conjunctive rule, on the other hand, implies that a person will display the response if *all* the features that are perceived in the situation are also linked to the response.

3. Estimation

As the estimation of the ULCM may be considered a standard problem and is supported by widely available software for categorical data analysis (Vermunt, 1997a, 1997b), we will focus on the estimation of the PM and the DM. Furthermore, it suffices to consider estimation of the DM as the only difference with the PM lies in whether or not one accepts overlapping latent classes.

The product representation of the conditional probabilities and the corresponding introduction of latent variables into the model enhances parameter estimation as the joint distribution of observed and latent variables has a simple mathematical structure, that is, $p(\mathbf{d},\mathbf{x},\mathbf{y},\mathbf{z}|\boldsymbol{\sigma},\boldsymbol{\rho},\boldsymbol{\xi}) = p(\mathbf{d}|\mathbf{x},\mathbf{y},\mathbf{z})p(\mathbf{x}|\mathbf{z},\boldsymbol{\sigma})p(\mathbf{y}|\mathbf{z},\boldsymbol{\rho})p(\mathbf{z}|\boldsymbol{\xi}) \propto p(\mathbf{x}|\mathbf{z},\boldsymbol{\sigma})p(\mathbf{y}|\mathbf{z},\boldsymbol{\rho})p(\mathbf{z}|\boldsymbol{\xi})$. As a result, algorithms that especially gain from the simple structure of the complete-data likelihood may be used for parameter estimation. In particular, an EM algorithm (Dempster, Laird and Rubin, 1977) can be used to locate the posterior mode(s) (see Appendix A) and a Gibbs sampling algorithm (Gelfand & Smith, 1990; Geman & Geman, 1984; Tanner & Wong, 1987) can be used to compute a sample of the posterior distribution (see Appendix B). To perform each of these estimation procedures, a Delphi program was written, which may be obtained from the authors upon request. It may be of interest to use both EM and Gibbs sampling for parameter estimation because they have the following strong and weak points:

First, with the Gibbs sampler specifying different types of priors is

well-supported whereas with the EM algorithm a concave prior is needed in order to guarantee the existence of a posterior mode within the interior of the parameter space. In particular, it is convenient to use a conjugate prior:

$$p(\boldsymbol{\sigma}, \boldsymbol{\rho}, \boldsymbol{\xi}) = \prod_j \prod_q \text{Beta}(\sigma_{jq}|\alpha_\sigma, \beta_\sigma) \prod_k \prod_q \text{Beta}(\rho_{kq}|\alpha_\rho, \beta_\rho) \times \text{Dirichlet}(\boldsymbol{\xi}|\gamma_1, \ldots, \gamma_T). \quad (7)$$

A uniform prior is obtained by putting the hyperparameters α, β and γ equal to 1. A concave prior, can be obtained by putting hyperparameters to values larger than 1. Unlike the EM algorithm, the Gibbs sampler is easily adapted to estimate the hyperparameters from the data. The latter may be advantageous as an ill-specified prior may decrease the goodness-of-fit of the model (see Meulders, De Boeck, Van Mechelen, & Gelman, 2000).

Second, monitoring convergence is straightforward when applying the EM algorithm because this algorithm has the strong property to increase the posterior density at each iteration and to converge to a stationnary point. On the other hand, with the Gibbs sampler, monitoring convergence is a difficult topic (for an overview see Cowles & Carlin, 1996). In this paper we will use the method of Gelman and Rubin (1992) to assess the convergence of the Gibbs sampler. According to this method convergence is concluded if, for each parameter of interest, the between-chain-variance plus the within-chain-variance divided by the within-chain-variance is near 1.

Third, when dealing with small samples, the Gibbs sampler provides a more accurate picture of parameter uncertainty than the EM algorithm. In particular, the sample of the posterior allows to compute $(1-\alpha)\%$ posterior intervals that are also valid with small samples whereas the computation of standard errors in the context of the EM algorithm is based on asymptotic theory.

Fourth, dealing with overparameterization or a lack of identifiability is less straightforward when using the Gibbs sampler than when using the EM algorithm. When using EM, one may check whether the Hessian at the obtained solution is negative definite. If this is the case, the model is said to be locally identifiable (Goodman, 1974). When using the Gibbs sampler no such simple check is available. In addition, overparameterization often implies that the Gibbs sampler visits different local maxima within one run. This complicates summarizing the information in the posterior sample and can retard the convergence of the chain to the true posterior.

4. Model Selection and Model Checking

The aim of *model selection* is to select one model out of a set of competing models which best captures the process that underlies the observations. To achieve this goal, model selection criteria seek for an optimal balance between goodness of fit (GOF) and model complexity. Indeed, using only GOF as a criterion for model selection often leads to unnecessarily complex models that overfit the data and for that reason do not generalize to other contexts (Myung, 2000).

Well-known model selection criteria that are often used in latent class analysis are Akaike's information criterion (AIC) (Akaike, 1973, 1974), Bozdogan's consistent AIC (CAIC) (Bozdogan, 1987), and the Bayesian information criterion (BIC) (Schwarz, 1978). These criteria are obtained by taking the sum of a GOF term (i.e., minus twice the log likelihood of the fitted model) and a penalty term, which is a measure of the complexity of the model. For AIC, CAIC and BIC the penalty terms equal $2k$, $k\log(I+1)$ and $k\log(I)$, respectively, with k being the number of free parameters and with I being the "total sample size". In the context of latent class models, I equals the number of persons rather than the number of cells of the contingency table to which the model is applied (Raftery, 1986). As the penalty term of AIC is usually smaller than that of CAIC and BIC, AIC will often select more complex models. On the other hand, since CAIC and BIC only differ by the constant $k\log((I+1)/I))$ they often lead to the same conclusions. In the example presented, we use the BIC as a criterion for model selection.

After having selected the best model out of a set of competing models it is recommended to use *model checking* procedures to investigate whether the selected model actually provides a good overall fit to the data. To check the absolute GOF of latent structure models one may use a Pearson-χ^2 measure which compares observed frequencies with frequencies that are expected under the model in the following way:

$$\chi^2(\mathbf{d},\boldsymbol{\theta}) = \sum_l \frac{(o_l(\mathbf{d}) - e_l(\mathbf{d},\boldsymbol{\theta}))^2}{e_l(\mathbf{d},\boldsymbol{\theta})}, \tag{8}$$

with $\boldsymbol{\theta} = (\boldsymbol{\xi}, \boldsymbol{\sigma}, \boldsymbol{\rho})$ being a vector that comprises the model's parameters and with o_l and e_l being observed and expected frequencies, respectively. In (8) the index l is used to indicate the cells of the contingency table for which the Pearson-χ^2 measure is defined. Two different types of such GOF measures are defined:

First, we define a GOF measure for the 2^{JK} contingency table formed by the full cross-classification of $J \times K$ variables. This measure,

which is labeled χ^2_{Full}, can be used to evaluate the assumption of conditional independence in the joint frequencies of all the variables and as such it is a test for the fit of the whole model.

Second, we define a Pearson-χ^2 measure for the second-order marginal distributions of the set of $J \times K$ dichotomous variables (see also Reiser & Lin, 1999). This measure, which is labeled $\chi^2_{Marginal}$, can be used to investigate the assumption of conditional independence in the marginal frequencies of response patterns on pairs of variables, with each variable being a particular situation-response combination in the present context.

Note that, for the first GOF measure, the index l in (8) is used to indicate each of 2^{JK} possible response patterns $\mathbf{d}_l$, whereas for the second GOF measure l indicates the four response patterns that may be observed for a particular pair of variables, that is, $(0,0)$, $(0,1)$, $(1,0)$, and $(1,1)$. Furthermore, the expected frequency of a response pattern is computed by multiplying the probability of that response pattern by the sample size (I).

It is important to remark that the above described GOF measures differ in several ways. First, the former is a test for the model as a whole whereas the latter focuses on a specific pair of variables. It is important to note that conditional independence in the joint frequencies implies conditional independence in the marginal frequencies (Reiser & Lin, 1999). However, $\chi^2_{Marginal}$ may be a more powerful test than χ^2_{Full} so that, with a non-significant χ^2_{Full}, for some pairs of variables the null hypothesis of conditional independence may still be rejected. An advantage of the specific $\chi^2_{Marginal}$ measures is that they provide more detailed information about the fit of the elements of the classification-basis mode(s). For instance, in the context of three-way three-mode data one may summarize the fit of a particular element(i.e. situation, response) by looking at the fit of the marginal measures in which this element is involved.

Second, unlike the data in a cross-classification table of two variables, the data in the full cross-classification table will generally be extremely sparse. That is, even with fairly large sample sizes (e.g. $I = 1000$) and small values of J and K (e.g. $J = K = 5$) the possible number of response patterns (2^{JK}) in the full cross-classification table will be much larger than I so that a high proportion of cell frequencies will be zero or one. As a result, the χ^2 approximation to the Pearson-χ^2 measure is no longer valid. To solve this problem one could use parametric bootstrap (Efron & Tibshirani, 1993) or posterior predictive checks (PPCs) (Gelman, Meng, & Stern, 1996) to simulate the reference distribution of the GOF measure. In this paper, the latter ap-

proach will be used to evaluate the significance of both $\chi^2_{Marginal}$ and χ^2_{Full}.

Gelman et al. (1996) describe two distinct computational procedures to assess the PPC p-value of measures $T(\mathbf{d}, \boldsymbol{\theta})$ that are a function of both data and parameters. In the first procedure, $\boldsymbol{\theta}$ is replaced by the posterior mode of the model, i.e. $\hat{\boldsymbol{\theta}}$. In this case, the PPC p-value can be computed by generating new data sets $\mathbf{d}^{\text{rep}}$ (using the draws from the posterior) and by computing the proportion of replicated data sets in which $T(\mathbf{d}^{\text{rep}}, \hat{\boldsymbol{\theta}}^{\text{rep}}) \geq T(\mathbf{d}, \hat{\boldsymbol{\theta}})$. Note that $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\theta}}^{\text{rep}}$ are the modes of the observed and the replicated data, respectively. It is worthwhile to note that this procedure only differs from a parametric bootstrap approach in that data are generated using draws from the posterior instead of using the posterior mode. As a result, this procedure may be advantageous in comparison with the parametric bootstrap procedure because the posterior uncertainty in the parameters is taken into account when generating new data sets. The measure $T(\mathbf{d}, \hat{\boldsymbol{\theta}})$ is labeled a statistic because both $\mathbf{d}$ and $\hat{\boldsymbol{\theta}}$ are a function of the data only.

In the second procedure, the PPC p-value is computed as the proportion of replicated data sets in which realized discrepancies $T(\mathbf{d}^{\text{rep}}, \boldsymbol{\theta})$ exceed or equal observed discrepancies $T(\mathbf{d}, \boldsymbol{\theta})$. Note that $\boldsymbol{\theta}$ denotes a draw of the posterior and that it is not replaced by $\hat{\boldsymbol{\theta}}$. An important advantage of this second procedure compared to the one using a statistic (derived from a discrepancy measure) is that it is computationally less demanding because observed and realized discrepancies are computed using a draw from the posterior rather than an estimate of the posterior mode for each replicated data set. A disadvantage, however, is that the procedure may have less power to reject the null hypothesis (see Meulders, De Boeck, & Van Mechelen, 2002).

5. Example

5.1 Data

As an illustration, the presented models are used to analyze the types of reactions that occur when one is angry at a person in a certain situation. The data consist of the answers of 101 first year psychology students who indicated whether or not they would display each of 8 behaviors when being angry at someone in each of 6 situations. The 8 behaviors consist of 4 pairs of reactions that reflect a particular strategy to deal with situations in which one is angry at someone, namely,

Table 1. BIC of unconstrained latent class model (ULCM), product model (PM), and disjunctive model (DM) assuming the number of basic profiles Q equals 1 up to 6

	model		
Q	ULCM	PM	DM
1	5944	6396	6396
2	5937	6314	5943
3	5999	6301	5848
4	6079	6324	5820
5	6203	6356	5846
6	6361	6418	5975

(1) fighting (fly off the handle, quarrel), (2) fleeing (leave, avoid), (3) emotional sharing (pour out one's heart, tell one's story), which is also called "befriending" by Taylor, Klein, Lewis, Gruenewald, Gurung, and Updegraff, (2000) and (4) making up (make up, clear up the matter). Furthermore, the six situations are constructed from two factors with three levels: (1) the extent to which one likes the instigator of anger (like, dislike, unfamiliar), and (2) the status of the instigator of anger (higher, lower, equal). Each situation is presented as one level of a factor, without specifying a level for the other factor.

5.2 Analysis

The EM algorithm is used to locate the posterior mode(s) for ULCM, PM, and DM with values of Q from 1 up to 6. Locating the global maximum is ensured by running the algorithm 20 times using random starting points and by afterwards choosing the best solution. Table 1 shows the values of the BIC criterion for each of the models. The model that balances best between GOF and parsimony (lowest value for BIC) is the DM with four basic profiles ($Q = 4$), defining $2^Q - 1 = 15$ possible latent classes. Further inspection also reveals that high BIC values for the PM are mainly due to low GOF (high value of badness of fit term) and that high BIC values for the ULCM are especially due to high model complexity (high value of penalty term).

For the DM with four basic profiles a sample of the posterior distribution is computed using the Gibbs sampling algorithm. Two chains are simulated using the posterior mode that was located with the EM algorithm as a starting point. After convergence is attained for each parameter, a sample of 5000 draws is constructed by gathering evenly

spaced draws from the second halves of the chains. The obtained sample of the posterior serves a double purpose: (1) via the computation of posterior intervals it may be used to evaluate whether parameters can be reliably estimated (see the section on the interpretation of the selected model), and (2) via the computation of PPC's it may be used to check the fit of the model.

The fit of the model is evaluated with the GOF measures χ^2_{Full} and $\chi^2_{Marginal}$. PPC p-values are computed on the basis of 1000 draws of the posterior and using the computational procedure for discrepancy measures (second procedure). The bootstrap-related procedure in which the GOF measures are considered as a statistic (first procedure) is not applied because it involves the high computational cost of estimating the model for each replicated data set. The results indicate that the model provides an acceptable fit to the data: First, for χ^2_{Full} the estimated PPC p-value equals .72 so that the model cannot be rejected on the basis of this test. Second, to summarize the results of $\chi^2_{Marginal}$ for pairs of variables we compute the number of model checks that are significant at the .01 level and that involve a particular situation or a particular response. For situations 1 to 6, 5%, 4% ,4% ,3% ,6%, and 5% of the model checks are significant, and for responses 1 to 8, 12% 2% 5%, 6%, 2%, 1%, 2%, and 3% are significant. Hence, there are more significant model checks than can be expected under the null hypothesis (i.e. 1%) but, in general, the proportions of significant checks are acceptable.

5.3 Interpretation of Selected Model

Tables 2 and 3 present the posterior mode and the corresponding 95% posterior intervals for mode-specific (situation and response) parameters and marginal latent class probabilities of the DM with four basic profiles (BP). Note that multiplying a situation and a response parameter associated with a basic profile yields the probability that the corresponding basic person type would display that response in that situation. As high situation and high response parameters with a basic profile imply that the corresponding basic person type has a high probability to display a response in a situation, we may also interpret the basic profiles as sensitivities to react with a certain type of behavior in a certain type of situation.

The first basic profile concerns primarily liked persons and persons of lower status. The predominant responses in such situations are making up and clearing up the matter, and to a lesser extent fighting (flying off the handle, quarreling) which gives the impression that making up some-

Table 2. Posterior Mode and 95% Posterior Interval (PI) of the mode-specific parameters for a Disjunctive Model with 4 Basic Profiles (BP).

		BP 1		BP 2		BP 3		BP 4	
set	element	mode	95% PI	mode	95% PI	mode	95% PI	mode	95% PI
situation	like	.95	[.85,1.00]	.52	[.35,.63]	.64	[.27,.98]	.40	[.28,.78]
	dislike	.08	[.03,.20]	.83	[.70,.93]	.60	[.42,.91]	.61	[.47,.95]
	unfamiliar	.04	[.00,.09]	.45	[.21,.52]	.42	[.33,88]	.82	[.59,.99]
	higher status	.04	[.00,.12]	.97	[.89,1.00]	.48	[.37,.87]	.66	[.48,.98]
	lower status	.77	[.68,.88]	.39	[.22,.49]	.49	[.17,.84]	.15	[.07,.37]
	equal status	.43	[.34,.58]	.82	[.73,.93]	.76	[.46,.98]	.36	[.26,.66]
response	fly off the handle	.47	[.37,.60]	.15	[.00,.19]	.77	[.46,.92]	.06	[.01,.23]
	quarrel	.52	[.44,.66]	.09	[.00,.13]	.66	[.35,.82]	.06	[.01,.16]
	leave	.07	[.00,.14]	.26	[.03,.34]	.04	[.00,.30]	.81	[.54,.96]
	avoid	.08	[.00,.17]	.38	[.13,.47]	.09	[.01,.38]	.86	[.57,.98]
	pour out one's heart	.09	[.01,.26]	.96	[.88,1.00]	.53	[.28,.82]	.12	[.02,.29]
	tell one's story	.10	[.00,.27]	.95	[.83,.99]	.70	[.44,.72]	.35	[.19,.61]
	make up	.95	[.83,1.00]	.08	[.00,.14]	.49	[.20,.60]	.03	[.00,.12]
	clear up the matter	.83	[.73,.95]	.09	[.01,.15]	.55	[.20,.66]	.02	[.00,.10]

Table 3. Posterior Mode and 95% Posterior Interval (PI) of the Marginal Latent Class Probabilities for a Disjunctive Model with 4 Basic Profiles.

class	Z_1	Z_2	Z_3	Z_4	mode	95%PI
1	0	0	0	1	.01	[.00,.04]
2	0	0	1	0	.03	[.00,.06]
3	0	0	1	1	.03	[.00,.08]
4	0	1	0	0	.01	[.00,.03]
5	0	1	0	1	.02	[.00,.05]
6	0	1	1	0	.05	[.00,.10]
7	0	1	1	1	.01	[.00,.06]
8	1	0	0	0	.01	[.00,.03]
9	1	0	0	1	.08	[.04,.17]
10	1	0	1	0	.05	[.01,.21]
11	1	0	1	1	.10	[.01,.19]
12	1	1	0	0	.22	[.02,.28]
13	1	1	0	1	.28	[.19,.49]
14	1	1	1	0	.09	[.04,.22]
15	1	1	1	1	.02	[.00,.08]

times follows a fight. Hence, this first sensitivity is related to situations of closeness and low resistance. The predominant response is one of reconciliation and making up.

The second basic profile includes primarily situations with persons of an equal or higher status or persons one dislikes. The predominant behavior seems to be sharing one's feelings with someone else than the instigator of anger (pour out one's hart, tell one's story) . This second sensitivity is somewhat contrasted to the first in that it relates to distance (high status, dislike) and higher resistance while the predominant behavior is again social but directed to a third party one is not angry with. In other words, one turns away from the person one is angry with and shares feelings with someone else.

The distinction between the first and the second sensitivity might be whether one experiences inhibition, and this may determine the kind of affiliative behavior. In case of no inhibition, one affiliates with the person one is angry with, but in case of inhibition, one affiliates with someone else.

The third basic profile is especially characterized by a low probability of flight. All other types of reactions have substantial but moderate probabilities to occur and this is relatively independent of the type of situation. In fact, the situation parameters for this basic profile have wide 95% posterior intervals, which means that situation parameters are not

reliably estimated. This third sensitivity is a all-but-flight (or approach) sensitivity.

Finally, the fourth basic profile is the complement of the third, with two high probabilities for the flight reactions (leave, avoid). In particular, one has a rather high probability to flight when the instigator of anger is unfamiliar, of higher status, or a person one dislikes. This fourth sensitivity is a flight-sensitivity.

Like the first and second sensitivities are complementary, also the third and fourth are. In sum, it seems that there are two dimensions of sensitivities. The first dimension contrasts two kinds of situations (inhibition or not) and two kinds of affiliative behaviors. Although contrasted, people can show both types of sensitivities as they relate to different situations. The second dimension is an approach versus avoidance dimension. In principle, one may combine both dimensions but this seems less likely from a psychological point of view.

As shown in Table 3, the single basic profiles (classes 1, 2, 4, and 8) have only small probabilities to occur in the population (i.e .01, .03, .01, and .01, respectively). This may partly explain the low fit of the PM, which only includes basic person types. Most persons combine two or more sensitivities. Especially combinations of the first two basic profiles are popular (61% in total): 22%, 28%, 9%, and 2% respectively combine $\{1,2\}$, $\{1,2,4\}$, $\{1,2,3\}$, $\{1,2,3,4\}$. As expected, combinations of 3 and 4 are much less popular (16% in total): 3%, 1%, 10%, and 2% respectively combine $\{3,4\}$, $\{2,3,4\}$, $\{1,3,4\}$, and $\{1,2,3,4\}$. Hence, the second "dimension" represents a much stronger contrast than the first one.

To illustrate the situation-response profiles of the relevant classes of the DM we graphically display, for each type of reaction, the profiles of the 5 person types that are most likely to be observed in the population (i.e. classes 9, 11, 12, 13, and 14) (see Figure 1). In Figure 1, the profiles for fight, flight, share and make up are obtained by averaging over the two reactions that were used to measure these response styles. This is justified as the pairs of reactions that measure the same response style have similar parameter estimates for the four basic profiles (see Table 2).

Two aspects are useful to describe the profiles of the different person types in Figure 1. First, person types may differ in the way they *differentiate* between situations. For instance, person types $\{9,11\}$ have about the same probability of sharing in different situations, whereas person types $\{12,13,14\}$ have a higher probability of sharing when being angry at a disliked person or at a person of higher or equal status, than when being angry at a liked person, an unfamiliar person, or a person of

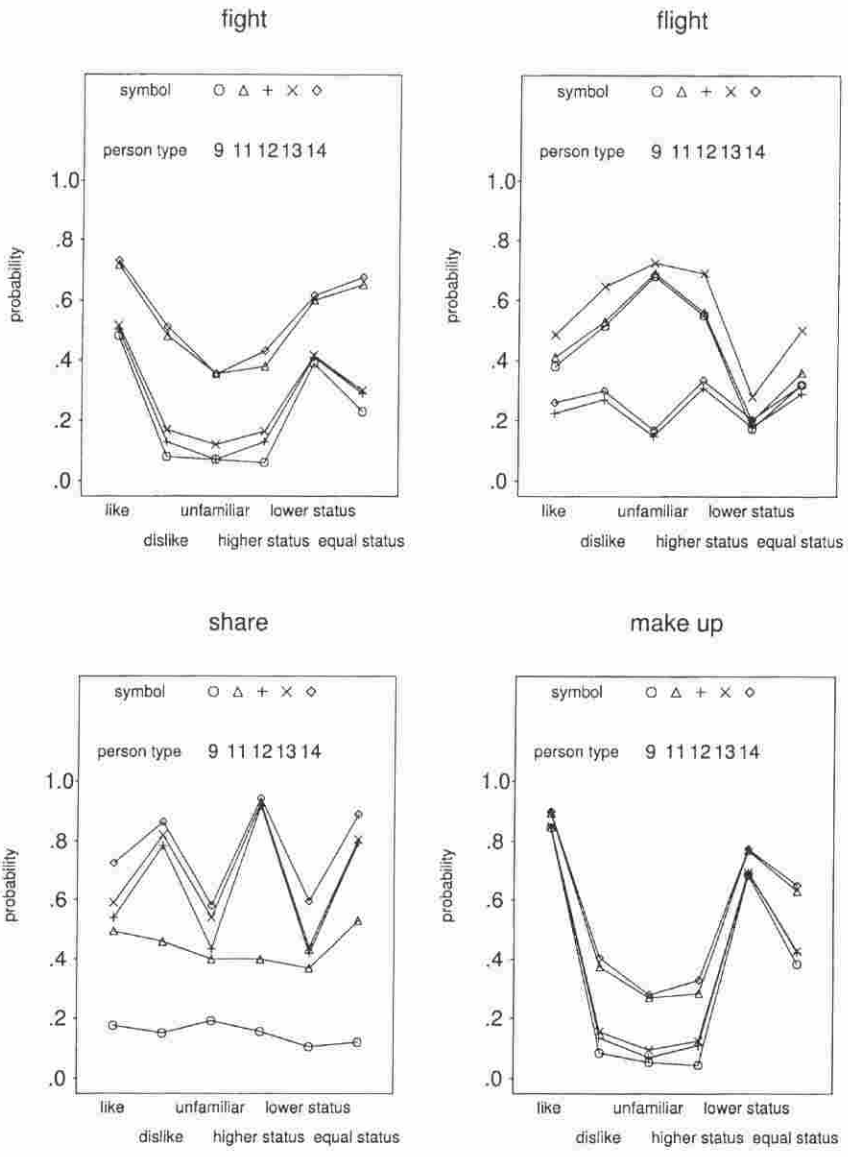

Figure 1. Situation-Response profiles of five person types that are most likely to be observed in the population.

lower status. Note also that person types $\{9, 11\}$ generally have a low probability to share. Hence it seems that differentiation is related to level, that is, sharing one's feelings with others depends on the situation only when one is very much inclined to share.

Moreover, the similarity of the profiles (i.e. equivalent pattern) associated with different person types can be explained by the fact that profiles of different person types are derived from the same basic profiles. In particular, two person types have a similar profile for a certain response if both person types are based on the same combination of the basic profiles that are associated with the response. For example, the two basic profiles that are associated with fighting (BP1 and BP3) are combined in the same way for person types $\{11, 14\}$ $(Z_1 = 1, Z_3 = 1)$ and for person types $\{9, 12, 13\}$ $(Z_1 = 1, Z_3 = 0)$, and these sets of person types have very similar profiles for fighting. Furthermore, it also follows that person types with a similar profile for one response type have different profiles for another response type if other basic profiles are associated with this response. In this way the different associations of situations and responses to basic profiles can be used to describe the different behaviors different person types display depending on the situation, that is, to capture the three-way interactions between persons, situations, and responses.

Second, apart from the way person types differentiate between situations, different types may be characterized by different levels of the *tendency* to display a particular behavior in a certain situation. For example, person types $\{11, 14\}$ and $\{9, 11, 13\}$ have about the same situation-response profiles for making up, but the former person types generally have higher probabilities to show this reaction.

Finally, we note that the profiles of the different person types for "fight" and "make up" are rather similar in that the same person types can be clustered on the basis of differentiation between situations and general tendency to display the response.

6. Discussion

In this paper, a constrained latent class model for three-way three-mode data is presented. The model can be used to classify entities of one mode on the basis of the elements of the other two modes simultaneously. The ULCM can often not serve this purpose because, when increasing the number of latent classes, the number of conditional probabilities rapidly increases so that parameters are no longer reliably estimated. Two mod-

els are proposed to solve this problem. First, in the PM, the conditional probabilities of the ULCM are constrained to be a product of basic parameters that are associated with the elements of the classification-basis modes. Second, in the DM, new latent classes are added to those of the PM by considering disjunctive combinations of situation-response profiles of the PM. In this way, individual differences can be better captured without increasing the number of basic parameters in the model. It is shown with an empirical application that the DM may yield a better balance between fit and model complexity (as measured by the BIC) compared to the ULCM or the PM.

For parameter estimation, a combined approach is developed which includes both a mode-finding algorithm (EM) to locate the posterior mode(s) and a sampling-based approach (Gibbs) to compute a (local) sample of the posterior distribution. The implementation of both types of algorithms is enhanced because the proposed models can be derived from a latent data process in which the observed data are considered to be a specific mapping of latent data. The advantages of computing a sample of the posterior and hence going beyond the posterior mode(s) are twofold: First, the posterior sample provides an accurate picture of the posterior uncertainty of the parameters that does not rely on an asymptotic normal approximation to the posterior. Second, the sample of the posterior can serve as a basis for posterior predictive assessment with PPCs. The latter approach is used for checking the GOF with two types of χ^2 measures: one for the contingency table that is formed by the full cross-classification of $J \times K$ variables and one for the marginal distributions of pairs of variables.

Finally, we mention some issues for future research. A first issue concerns the fixed relation between the number of basic profiles (Q) and the number of latent classes (2^Q). As this assumption may be suboptimal in a particular application, it may be interesting to relax this assumption by allowing more or less latent classes for models with a certain number of basic profiles. More specifically, when assuming a relatively large number of basic profiles (4 or more) it may be interesting to omit latent classes that are almost empty, and with a small number of basic profiles (less than 4) it may be interesting to extend the number of latent classes. A restriction can be obtained by discarding the vector $\mathbf{z}_t$ associated with latent class t from the design matrix $\mathbf{Z}$. The extension can for instance be obtained by assuming that person types weigh the basic profiles according to a prespecified fixed number of weights in the $[0, 1]$ interval. If only binary weights are used, the original DM in (6) is obtained with weights

z_{iq}. If, for instance, 0,1 and 0.5 weights are used, the model is extended to include 3^Q person types. Both types of modifications (restricting and extending the number of classes) can be easily implemented with the Gibbs sampling algorithm in Appendix A; however, for the model with non-binary weights the EM algorithm is complex as in this case the M-step has no longer a closed form solution.

A second issue for future research is to compare the presented models (PM and DM) with models that use another type of link function to model the (transformed) conditional probabilities as a linear function of mode-specific basic parameters. In particular, a useful alternative for the log link of the PM (recall that according to the PM $\log(\pi_{jkq}) = \sigma_{jq} + \rho_{kq}$) is to consider a logit link (i.e. $\text{logit}(\pi_{jkq}) = \sigma_{jq} + \rho_{kq}$) as is used in linear logistic latent class analysis. Analogous to extending the PM to the DM one may extend the logistic model with basic profiles by defining additional latent classes and associated situation-response profiles that combine basic profiles in a certain way. In particular, one could define the following model for the conditional probabilities:

$$\text{logit}P(D_{ijk} = 1|\mathbf{z}_i, \boldsymbol{\sigma}, \boldsymbol{\rho}) = \sum_q (\sigma_{jq} + \rho_{kq}) z_{iq} \tag{9}$$

Note that (9) can be considered a compensatory model as the parameters of different basic profiles are simply added. The DM, on the other hand, can be considered a non-compensatory or configural model because it combines basic profiles in a disjunctive way. Furthermore, model (9) can be considered a restricted variant of the LC factor model developed by Magidson and Vermunt (2001). It can be estimated with standard software for the analysis of categorical data such as LEM (Vermunt, 1997a, 1997b) or Latent Gold (Vermunt & Magidson, 2000).

A third issue concerns extensions of the presented models for polytomous data. For the logistic model (9) this kind of extension is straightforward as the LC factor model can handle polytomous data (see Magidson & Vermunt, 2001). For the DM, one possible strategy that could be followed to build a model for polytomous data is to use multinomially distributed polytomous latent variables X and Y and to define a suitable mapping $C(\cdot)$ of X, Y and Z into polytomous observed variables D.

Appendix A: Computation of the Posterior Mode

To describe the EM algorithm it is convenient to consider a model for frequencies of response patterns that are observed with all the entities

in the sample. In case of $J \times K$ binary observed variables there are 2^{JK} possible patterns which are indicated as $\mathbf{d}_l$ $(l = 1, \ldots, L)$. Furthermore, the vector of random variables $\mathbf{N} = (N_1, \ldots, N_L)$ is used to denote the frequencies of those patterns and $\mathbf{n} = (n_1, \ldots, n_L)$ is used to denote a specific realization. Assuming a multinomial model for $\mathbf{N}$, the observed posterior distribution may be expressed as

$$p(\boldsymbol{\sigma}, \boldsymbol{\rho}, \boldsymbol{\xi}|\mathbf{n}) \propto p(\boldsymbol{\sigma}, \boldsymbol{\rho}, \boldsymbol{\xi}) \prod_l [p(\mathbf{d}_l|\boldsymbol{\sigma}, \boldsymbol{\rho}, \boldsymbol{\xi})]^{n_l}.$$

The EM algorithm can be used to maximize the observed posterior distribution because the observed data can be considered a mapping of latent data. More specifically, the data that are observed for a specific entity can be considered a mapping of a $J \times K \times Q$ array of latent observations X_k^{jq} denoted as $\mathbf{X}$, a $J \times K \times Q$ array of latent observations Y_j^{kq} denoted as $\mathbf{Y}$, and a Q-element vector of latent classifications Z_q denoted as $\mathbf{Z}$. With dichotomous latent variables there are R $(= 2^{JKQ})$ different arrays $\mathbf{X}$, denoted as $\mathbf{x}_r$ $(r = 1, \ldots, R)$; S $(= 2^{JKQ})$ different arrays $\mathbf{Y}$ denoted as $\mathbf{y}_s$ $(s = 1, \ldots, S)$; and T $(= 2^Q)$ different vectors $\mathbf{Z}$ denoted as $\mathbf{z}_t$ $(t = 1, \ldots, T)$. Furthermore, the random variable V_{rst} is used to denote the number of entities that have latent data patterns $(\mathbf{x}_r, \mathbf{y}_s, \mathbf{z}_t)$.

Assuming a multinomial model for $\mathbf{V}$, the augmented posterior can be expressed as

$$p(\boldsymbol{\sigma}, \boldsymbol{\rho}, \boldsymbol{\xi}|\mathbf{n}, \mathbf{v}) \propto p(\boldsymbol{\sigma}, \boldsymbol{\rho}, \boldsymbol{\xi}) \prod_s \prod_r \prod_t [p(\mathbf{x}_s, \mathbf{y}_r, \mathbf{z}_t|\boldsymbol{\sigma}, \boldsymbol{\rho}, \boldsymbol{\xi})]^{v_{srt}}.$$

In iteration (m+1), the EM algorithm consists of two steps:

1. *Expectation Step:* Compute the expected value of the logarithm of the augmented posterior with respect to the distribution of the latent data, $\mathbf{v}$, conditional on the observed data, $\mathbf{n}$, and the current guess of the posterior mode, $(\boldsymbol{\sigma}^{(m)}, \boldsymbol{\rho}^{(m)}, \boldsymbol{\xi}^{(m)})$, that is, compute

$$Q(\boldsymbol{\sigma}, \boldsymbol{\rho}, \boldsymbol{\xi}; \boldsymbol{\sigma}^{(m)}, \boldsymbol{\rho}^{(m)}, \boldsymbol{\xi}^{(m)}) =$$
$$\int_{\mathbf{V}} \ln[p(\boldsymbol{\sigma}, \boldsymbol{\rho}, \boldsymbol{\xi}|\mathbf{n}, \mathbf{v})]\, p(\mathbf{v}|\boldsymbol{\sigma}^{(m)}, \boldsymbol{\rho}^{(m)}, \boldsymbol{\xi}^{(m)}, \mathbf{n}) d\mathbf{v}.$$

2. *Maximization Step:* Maximize $Q(\boldsymbol{\sigma}, \boldsymbol{\rho}, \boldsymbol{\xi}; \boldsymbol{\sigma}^{(m)}, \boldsymbol{\rho}^{(m)}, \boldsymbol{\xi}^{(m)})$ with respect to $(\boldsymbol{\sigma}, \boldsymbol{\rho}, \boldsymbol{\xi})$

The algorithm has converged to a stationary point (local maximum or saddle point) if

$$\left.\frac{\partial p(\boldsymbol{\sigma}, \boldsymbol{\rho}, \boldsymbol{\xi}|\mathbf{n})}{\partial(\boldsymbol{\sigma}, \boldsymbol{\rho}, \boldsymbol{\xi})}\right|_{\hat{\boldsymbol{\sigma}}, \hat{\boldsymbol{\rho}}, \hat{\boldsymbol{\xi}}} \approx 0.$$

As the log augmented posterior is a linear function of the V_{srt}'s both the E-step and the M-step of the EM algorithm are easily implemented (see also Maris, 1999). More specifically, in the E-step one only has to substitute the V_{srt}'s by their conditional expected values and in the M-step one has to maximize the log augmented posterior.

Assuming a conjugate prior (see 7) it is easy to derive that the maximization of the log augmented posterior in the M-step yields the following closed form solution:

$$\hat{\sigma}_{jq} = \frac{(\alpha_\sigma - 1) + \sum_k \sum_s \sum_r \sum_t v_{srt} z_{tq} x_{ks}^{jq}}{(\alpha_\sigma - 1) + (\beta_\sigma - 1) + \sum_k \sum_s \sum_r \sum_t v_{srt} z_{tq}} \tag{10}$$

$$\hat{\rho}_{kq} = \frac{(\alpha_\rho - 1) + \sum_j \sum_s \sum_r \sum_t v_{srt} z_{tq} y_{jr}^{kq}}{(\alpha_\rho - 1) + (\beta_\rho - 1) + \sum_j \sum_s \sum_r \sum_t v_{srt} z_{tq}} \tag{11}$$

$$\hat{\xi}_t = \frac{(\gamma_t - 1) + \sum_s \sum_r v_{srt}}{\sum_c (\gamma_t - 1) + \sum_s \sum_r \sum_t v_{srt}} \tag{12}$$

In the E-step we have to compute the conditional expected values of the statistics in (10), (11), and (12). We first consider the numerator of (12). To compute $\mathrm{E}[\sum_s \sum_r V_{srt} | \mathbf{n}, \boldsymbol{\sigma}, \boldsymbol{\rho}, \boldsymbol{\xi}]$ we note that

$$\mathrm{E}[V_{srt} | \mathbf{n}, \boldsymbol{\sigma}, \boldsymbol{\rho}, \boldsymbol{\xi}] = \sum_l n_l p(\mathbf{x}_s, \mathbf{y}_r, \mathbf{z}_t | \mathbf{d}_l, \boldsymbol{\sigma}, \boldsymbol{\rho}, \boldsymbol{\xi}) \tag{13}$$

so that

$$\begin{aligned} \mathrm{E}[\textstyle\sum_s \sum_r V_{srt} | \mathbf{n}, \boldsymbol{\sigma}, \boldsymbol{\rho}, \boldsymbol{\xi}] &= \textstyle\sum_l n_l \sum_s \sum_r p(\mathbf{x}_s, \mathbf{y}_r, \mathbf{z}_t | \mathbf{d}_l, \boldsymbol{\sigma}, \boldsymbol{\rho}, \boldsymbol{\xi}) \\ &= \textstyle\sum_l n_l p(\mathbf{z}_t | \mathbf{d}_l, \boldsymbol{\sigma}, \boldsymbol{\rho}, \boldsymbol{\xi}) \end{aligned}$$

with

$$p(\mathbf{z}_t | \mathbf{d}_l, \boldsymbol{\sigma}, \boldsymbol{\rho}, \boldsymbol{\xi}) = \frac{p(\mathbf{d}_l | \mathbf{z}_t, \boldsymbol{\sigma}, \boldsymbol{\rho}) p(\mathbf{z}_t | \boldsymbol{\xi})}{\sum_c p(\mathbf{d}_l | \mathbf{z}_c, \boldsymbol{\sigma}, \boldsymbol{\rho}) p(\mathbf{z}_c | \boldsymbol{\xi})}$$

Second, we consider the denominator of (10). Computation of the conditional expected value of the denominator in (11) is analogous. Using (13) it is easy to derive that

$$\mathrm{E}(\sum_k \sum_s \sum_r \sum_t V_{srt} z_{tq} | \mathbf{n}, \boldsymbol{\sigma}, \boldsymbol{\rho}, \boldsymbol{\xi}) = K \sum_t z_{tq} \sum_l n_l p(\mathbf{z}_t | \mathbf{d}_l, \boldsymbol{\sigma}, \boldsymbol{\rho}, \boldsymbol{\xi})$$

Third, we consider the numerator of (10). Computation of the conditional expected value of the numerator in (11) is analogous. Using (13) it is easy to derive that $\mathrm{E}[\sum_k \sum_s \sum_r \sum_t V_{srt} z_{tq} x_{ks}^{jq} | \mathbf{n}, \boldsymbol{\sigma}, \boldsymbol{\rho}, \boldsymbol{\xi}]$ equals

$$\sum_t z_{tq} \sum_l n_l p(\mathbf{z}_t | \mathbf{d}_l, \boldsymbol{\sigma}, \boldsymbol{\rho}, \boldsymbol{\xi}) \sum_k [\sum_s x_{ks}^{jq} p(\mathbf{x}_s | \mathbf{z}_t, \mathbf{d}_l, \boldsymbol{\sigma}, \boldsymbol{\rho})],$$

with

$$\sum_s x_{ks}^{jq} p(\mathbf{x}_s|\mathbf{z}_t, \mathbf{d}_l, \boldsymbol{\sigma}, \boldsymbol{\rho}) = P(X_k^{jq} = 1|d_{ijk}, \mathbf{z}_t, \boldsymbol{\sigma}, \boldsymbol{\rho}).$$

Appendix B: Computation of a Sample of the Posterior Distribution

In this section we describe the Gibbs sampler to compute a sample of the observed posterior distribution $p(\boldsymbol{\sigma}, \boldsymbol{\rho}, \boldsymbol{\xi}|\mathbf{d})$. A conjugate prior (7) is used. In each iteration, the algorithm involves the following steps:

1. For each entity i draw the vector $\mathbf{z}_i$ from

$$p(\mathbf{z}_i|\mathbf{d}_i, \boldsymbol{\sigma}, \boldsymbol{\rho}, \boldsymbol{\xi}) \propto [\prod_j \prod_k p(d_{ijk}|\mathbf{z}_i, \boldsymbol{\sigma}_j, \boldsymbol{\rho}_k)] p(\mathbf{z}_i|\boldsymbol{\xi})$$

2. For each triple (i, j, k) draw the vector $\mathbf{x}_{ki}^j$ from

$$p(\mathbf{x}_{ki}^j|d_{ijk}, \mathbf{z}_i, \boldsymbol{\sigma}_j, \boldsymbol{\rho}_k) \propto p(d_{ijk}|\mathbf{x}_{ki}^j, \mathbf{z}_i, \boldsymbol{\rho}_k) p(\mathbf{x}_{ki}^j|\mathbf{z}_i, \boldsymbol{\sigma}_j)$$

3. For each triple (i, j, k) draw the vector $\mathbf{y}_{ji}^k$ from

$$p(\mathbf{y}_{ji}^k|d_{ijk}, \mathbf{z}_i, \mathbf{x}_{ki}^j, \boldsymbol{\rho}_k) \propto p(d_{ijk}|\mathbf{z}_i, \mathbf{x}_{ki}^j, \mathbf{y}_{ji}^k) p(\mathbf{y}_{ji}^k|\mathbf{z}_i, \boldsymbol{\rho}_k)$$

4. For each pair (j, q) draw σ_{jq} from

$$\text{Beta}(\alpha_\sigma + \sum_k \sum_i x_{ki}^{jq} z_{iq}, \beta_\sigma + \sum_k \sum_i (1 - x_{ki}^{jq}) z_{iq})$$

5. For each pair (k, q) draw ρ_{kq} from

$$\text{Beta}(\alpha_\rho + \sum_j \sum_i y_{ji}^{kq} z_{iq}, \beta_\rho + \sum_j \sum_i (1 - y_{ji}^{kq}) z_{iq})$$

6. Draw $\boldsymbol{\xi}$ from Dirichlet$(\gamma_1 + g_1, \ldots, \gamma_T + g_T)$ with g_t $(t = 1, \ldots, T)$ the number of entities with pattern $\mathbf{z}_t$.

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrov and F. Csaki (Eds.), *Second international symposium on information theory*, pp. 271–281. Budapest: Academiai Kiado.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19, 716–723.

Bozdogan, H. (1987). Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions. *Psychometrika, 52*, 345–370.

Cowles, K., & Carlin, B. P. (1996). Markov Chain Monte Carlo convergence diagnostics: A comparative review. *Journal of the American Statistical Association, 91*, 883–904.

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

Efron, B, & Tibshirani, R. J. (1993). *An introduction to the bootstrap*. New York: Chapmann & Hall.

Hunt, L. A., & Basford, K. E. (1999). Fitting a mixture model to three-mode three-way data with categorical and continuous variables. *Journal of Classification, 16*, 283–296.

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

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

Formann, A. K. (1989). Constrained latent class models: Some further applications. *British Journal of Mathematical and Statistical Psychology, 42*, 37–54.

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

Gelfand, A. E., & Smith, A. F. M. (1990). Sampling based approaches to calculating marginal densities. *Journal of the American Statistical Association, 85*, 398–409.

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science, 7*, 457–472.

Gelman, A., Meng, X. M., & Stern, H. (1996). Posterior predictive assessment of model fitness via realized discrepancies. *Statistica Sinica, 4*, 733–807.

Geman, S., & Geman, D. (1984). Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 6*, 721–741.

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

Leenen, I., Van Mechelen, I., De Boeck, P., & Rosenberg, S. (1999). INDCLAS: A three-way hierarchical classes model. *Psychometrika, 64*, 9–24.

Magidson, J., & Vermunt, J. K. (2001). Latent class factor and cluster models, biplots, and related graphical displays. *Sociological methodology, 31*, 223-264.

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

Maris, E., De Boeck, P., & Van Mechelen, I. (1996). Probability matrix decomposition models. *Psychometrika, 61*, 7–29.

Meulders, M., De Boeck, P., Van Mechelen, I., & Gelman, A. (2000) *Hierarchical extensions of probability matrix decomposition models.* Manuscript submitted for publication.

Meulders, M, De Boeck, P., Van Mechelen, I. (2001). Probability matrix decomposition models and main-effects generalized linear models for the analysis of replicated binary associations. *Computational Statistics and Data Analysis, 38*, 217–233.

Meulders, M., De Boeck, P., Van Mechelen, I., Gelman, A., & Maris, E. (2001). Bayesian inference with probability matrix decomposition models. *Journal of Educational and Behavioral Statistics, 26*, 153–179.

Meulders, M., De Boeck, P., Van Mechelen, I. (2002). A taxonomy of latent structure assumptions for probability matrix decomposition models. Accepted by *Psychometrika.*

Myung, J. (2000). The importance of complexity in model selection. *Journal of Mathematical Psychology, 44*, 190-204.

Raftery, A. E. (1986). A note on Bayes factors for log-linear contingency table models with vague prior information. *Journal of the Royal Statistical Society, Series B, 48*, 249–250.

Reiser, M., & Lin, Y. (1999). A goodness-of-fit test for the latent class model when expected frequencies are small. *Sociological Methodology, 29*, 81–111.

Schwarz, G. (1978). Estimating the dimensions of a model. *Annals of Statistics, 6*, 461–464.

Tanner, M. A., & Wong, W. H. (1987). The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association, 82*, 528–540.

Taylor, S. E., Klein, L. C., Lewis, B. P., Gruenewald, T. L., Gurung, R. A. R., & Updegraff, J. A. (2000). Biobehavioral responses to stress in females: tend-and-befriend, not fight-or-flight. *Psychological Review, 107(3)*, 411-429.

Vermunt, J. K. (1997a). *LEM: A general program for the analysis of categorical data. User's manual.* Tilburg University, The Netherlands.

Vermunt J. K. (1997b). *Log-linear models for event histories.* Thousand Oaks, CA: Sage.

Vermunt, J. K., & Magidson, J. (2000). *Latent GOLD 2.0 user's guide.* Belmont, MA: Statistical Innovations.

Copyright © 2003 EBSCO Publishing