# Using Bayesian Priors for More Flexible Latent Class Analysis

Tihomir Asparouhov* Bengt Muthén †

**Abstract**

Latent class analysis is based on the assumption that within each class the observed class indicator variables are independent of each other. We explore a new Bayesian approach that relaxes this assumption to an assumption of approximate independence. Instead of using a correlation matrix with correlations fixed to zero we use a correlation matrix where all correlations are estimated using an informative prior with mean zero but non-zero variance. This more flexible approach easily accommodates LCA model misspecifications and thus avoids spurious class formations that are caused by the conditional independence violations. Simulation studies and real data analysis are conducted using Mplus.

**Key Words:** Bayesian, Latent class analysis, Conditional dependence, Informative priors, Mixtures, Mplus

## 1. Introduction

In this article we describe new modeling possibilities for Latent Class Analysis (LCA) that are now available as a result of methodological advances in Bayesian estimation. The LCA model has traditionally been estimated with the maximum-likelihood estimator via the EM algorithm, see Goodman (1974). Recent advances in Bayesian estimation have made it feasible to estimate the LCA model also within a Bayesian framework, see Elliott et. al. (2005), and Asparouhov and Muthén (2010). In particular the approach of Asparouhov and Muthén (2010) includes algorithms for estimating a correlation matrix within this framework. Using this correlation matrix approach the LCA model can be generalized to a more flexible model where the within class indicators are no longer required to be independent but can be freely correlated through an underlying correlation matrix. Additional modeling possibilities will be discussed here that arise from introducing constraints on this correlation matrix. The most common ways to constraint a correlation matrix is through structural constraints such as those that can be induced through factor analysis within each class or simply constraining some correlation to zero while freely estimating others. However the primary focus of this article will be a different type of constraint, a constraint that is induced through using a specific prior for the correlation matrix. These constraints have been pioneered in Muthén and Asparouhov (2011) within the framework of structural equation models for continuous variables. The constraints are used for models that are traditionally unidentified within the maximum-likelihood estimation framework but within the Bayesian framework those model are identified through very restrictive priors. The idea of using such models and priors is to find the most important model misspecifications among a vast number of such misspecifications. In this article as well we use this restrictive prior approach to identify possible model misspecifications in the LCA framework, however, we only apply this approach to models that are truly identified even within the maximum-likelihood framework.

Traditionally with the Bayesian estimation, the prior distributions of the model parameters are either informative or non-informative. Informative priors are used when there is

*Mplus, 3463 Stoner Ave., Los Angeles, CA, 90066

†Mplus, 3463 Stoner Ave., Los Angeles, CA, 90066

indeed some kind of information already available for the parameters and non-informative priors are used when no such information is available. In this article we use a third type of priors. These priors are informative but are not based on specific information about the parameters but rather reflect the analyst's belief that a certain model should approximate the data well. Specifically in the LCA case, the analyst believes that the LCA model will approximate the data well, i.e., that the class indicators are approximately independent within class. This can be translated as information about the within class correlation matrix. The analyst believes that the correlation matrix should be close to a diagonal matrix. This prior belief can be introduced in the model estimation by specifying a very restrictive prior for correlation matrix that is centered around a diagonal matrix but does allow some wiggle room for the correlations to be different from zero. This way in the Bayesian estimation only those correlations that are truly non-zero, i.e., those for which the data contains evidence that they are non-zero, will escape the restrictive prior, i.e., the data will override the restrictive prior to yield a non-zero correlation. The correlations that are truly zero will not be able to escape the restrictive prior and will be estimated to zero. We call this use of informative prior a model based informative prior and we will discuss below how to specify such priors. Even though we use informative priors, this methodology can be used in practical applications where no prior information is available at all.

In this article we use a non-traditional approach to Bayesian estimation. We consider the Bayesian estimation paired with parameter priors and a method for deriving point estimates and standard errors from the estimated posterior distributions simply as another frequentist estimator. Such an estimator will be evaluated via traditional frequentist means such as bias, mean squared error and coverage of the confidence interval in repeated applications. While traditional Bayesian application based on a specific loss function is of interest as well in this article we only focus on Bayesian estimation as a means to construct new frequentist estimators.

All model estimations presented in this article have been conducted with the Mplus program version 6.11. All of the Mplus inputs and outputs presented here can be found online at statmodel.com.

In Section 2 we describe the conditional dependence LCA models with binary indicator variables and a general tetrachoric correlation matrix. In Section 3 we outline the Bayesian estimation procedure for this model. In Section 4 we describe several conditional dependence models within a Bayesian framework and show the role of the parameter priors in the model definition. In Section 5 we present several simulation studies. In Section 6 we show that if conditional dependence is ignored the LCA analysis can lead to spurious class formations. In Section 7 we present a real data application. We conclude in Section 8.

## 2. LCA with conditional dependence

Let $Y_i$, $i = 1, ..., m$ be a binary observed variable, taking values 0 and 1, and $C$ be a categorical latent class variable taking values 1,...,K. The conditional independence LCA model is given by the following equation

$$P(Y_1, ..., Y_m|C) = P(Y_1|C)...P(Y_m|C) = \prod_{i=1}^{m} p_{ic}^{Y_i}(1 - p_{ic})^{1-Y_i} \quad (1)$$

where $p_{ic}$ are parameters to be estimated as well as the parameter $q_i = P(C = i)$, $i = 1, ..., K$. This model can also be formulated in terms of a multivariate probit model for underlying latent normally distributed variables $Y_i^*$

$$Y_i^*|C \sim N(\mu_{ic}, 1) \quad (2)$$

$$Y_i^* < 0 \Leftrightarrow Y_i = 0 \tag{3}$$

$$P(Y_i = 0|C) = P(Y_i^* < 0|C) = \Phi(\mu_{ic}). \tag{4}$$

As a multivariate model this can be expressed as

$$Y^*|C \sim N(\mu_c, I) \tag{5}$$

where $Y^* = (Y_1^*, ..., Y_m^*)$ and $\mu_c = (\mu_{1c}, ..., \mu_{mc})$. A natural way to introduce conditional dependence models within LCA is to replace equation (5) with

$$Y^*|C \sim N(\mu_c, \Sigma_c) \tag{6}$$

where $\Sigma_c$ is an unrestricted correlation matrix. The independence model of course is a special case of the above model, i.e, it is equivalent to all off-diagonal elements $\Sigma_c$ being equal to 0, i.e., $\Sigma_c = I$. The correlations in $\Sigma_c$ are generally known as the tetrachoric correlations for the observed binary variables, with the exception that the above model is within an LCA framework which means that the tetrachoric correlations are conditional on the class variable and vary across classes. Models where only some tetrachoric correlations in $\Sigma_c$ are estimated, rather than the full correlation matrix, are of interest as well because in practical applications it is very likely that only some pairs of variables show violations of conditional independence. In addition, the above formulation can easily accommodate general structural equation models for $Y_i^*$. In particular it can easily accommodate random effects that can be used to explain conditional dependence among the variables as in Qu et. al. (1996). A model with one random effect for example can be expressed as

$$Y^*|C = \mu_c + \Lambda_c\eta + \varepsilon \tag{7}$$

where $\eta$ is a standard normal latent variable with mean zero and variance 1 and $\varepsilon|C \sim N(0, \Theta_c)$ and $\Theta_c$ is a correlation matrix where some correlations can be free parameters but typically most of the correlations will be fixed to 0, and $\Lambda_c$ are the loading parameters to be estimated. Random effect models can be preferable in practical applications because the random effects explain the conditional dependence within class and may also have substantive interpretations. One such model is the factor mixture model, see for example Muthén (2006; 2008) and Muthén and Asparouhov (2006).

There are other ways to introduce conditional dependence in the LCA model, such as for example the log-linear model or recursive set of logit models, however these models are more difficult to interpret, may require many more parameters and would not naturally accommodate random effects.

Estimating model (6) with the maximum-likelihood estimator is not an easy task because it requires the evaluation of the multivariate probit function which is not feasible for high dimensions. We therefore use the Bayesian methods for this estimation problem. The MCMC estimation for the conditional dependence model (6) is outlined in the following section.

## 3. MCMC Estimation

In this section we describe the MCMC algorithm for estimating the conditional dependence LCA model (6). The model contains the following sets of parameters $\mu_c$, $\Sigma_c$, $q_i$ and the following latent variables $Y_i^*$ and $C$. To use the Bayesian methodology we need to specify priors for the three sets of parameters. For the parameters $\mu_c$ and $q_i$ we simply use the default priors available in Mplus. These priors are not the focus of our investigation. The prior for $\mu_{ic}$ is $N(m_0, \sigma_0)$, where $m_0 = 0$ and $\sigma_0 = 5$. This prior is a weakly informative prior,

see Gelman et al. (2008), and is chosen so that the Bayesian estimation yields estimates close to the maximum-likelihood estimates even when the sample size is not large. Essentially what this prior does however is to eliminate large negative or large positive parameters values which are not needed. In fact even with the maximum-likelihood estimation, these parameters are constrained typically to be within the interval [-15,15]. Beyond this interval the $\mu_{ic}$ parameter are indistinguishable as they all imply conditional probability of 0 or 1.

The prior for the parameters $q_i$ is the Dirichlet distribution $D(a_0, ..., a_0)$, where $a_0 = 10$. In this case again the prior is weakly informative. Intuitively the prior specifies that tiny classes are not of interest, i.e., the prior prevents empty class solutions.

Finally the prior for the correlation matrix $\Sigma_c$ is the marginal correlation distribution of the Inverse Wishart distribution $IW(\Sigma_{0c}, d)$, i.e., to construct this prior the distribution $IW(\Sigma_{0c}, d)$ is used to construct variance covariance matrices which are then reduced to correlation matrices. This prior has no explicit formulation, however, it is a conjugate prior for the PX - algorithm implemented in Mplus for the estimation of the correlation matrix, see Asparouhov and Muthén (2010).

The MCMC algorithm is based on the Gibbs sampler, see Gelman et al. (2004), and it uses the following 5 step generation process

$$[\mu|Y^*, \Sigma, Y, C, q] \sim [\mu|Y^*, \Sigma, C] \tag{8}$$

$$[\Sigma|Y^*, \mu, Y, C, q] \sim [\Sigma|Y^*, \mu, C] \tag{9}$$

$$[C|\mu, \Sigma, Y^*, Y, q] \sim [C|\mu, \Sigma, Y^*, q] \tag{10}$$

$$[q|\mu, \Sigma, Y^*, Y, C] \sim [q|C] \tag{11}$$

$$[Y^*|\mu, \Sigma, Y, C, q] \sim [Y^*|\mu, \Sigma, Y, C] \tag{12}$$

Steps (8) and (9) are performed separately for each class.

The posterior distribution in (8) is given by

$$[\mu_c|Y^*, \Sigma, C] \sim N(Dd, D) \tag{13}$$

where

$$D = \left(n_c\Sigma_c^{-1} + \Sigma_{0c}^{-1}\right)^{-1} \tag{14}$$

$$d = n_c\Sigma_c^{-1}\bar{Y}_c^* + \Sigma_0^{-1}m_0 \tag{15}$$

where $\Sigma_{0c}$ is the prior variance matrix of $\mu_c$, which in our case is a diagonal matrix with 5 for all the diagonal entries, $m_{0c}$ is the prior mean of $\mu_c$, which in our case is a vector of zeros, $n_c$ is the number of observations classified in class $c$ and $\bar{Y}_c^*$ is the sample mean of $Y^*$ in class $c$.

The posterior distribution in (9) is sampled through the PX-algorithm, see van Dyk and Meng (2001) for example. We extend the correlation matrix parameter space to the variance covariance parameter space. Because the observed data is categorical we know that the diagonal entries of $\Sigma_c$ are not really identified, nevertheless these parameters become a part of the Bayesian estimation. The only information available for these parameters is in the prior, i.e., the posterior distribution for these parameters is the same as the prior. We need these parameters to be able to use conjugate priors. If the prior for $\Sigma_c$ is $IW(\Sigma_{0c}, f_c)$ then the posterior for $\Sigma_c$ is

$$IW(E_c + \Sigma_{0c}, n_c + f_c) \tag{16}$$

where $E_c = \sum(Y^* - \mu_c)^T(Y^* - \mu_c)$ and the sum is taken over all observations classified in class $c$.

The posterior distribution in (10) is given by

$$P(C = j|\mu, \Sigma, Y^*, q) = \frac{q_j f(\mu_j, \Sigma_j, Y^*)}{\sum_j q_j f(\mu_j, \Sigma_j, Y^*)} \tag{17}$$

where $f(\mu_j, \Sigma_j, Y^*)$ is the multivariate normal density function

$$(2\pi)^{-m/2}|\Sigma_j|^{-1/2}Exp(-(Y^* - \mu_j)^T\Sigma_j^{-1}(Y^* - \mu_j)/2) \tag{18}$$

The posterior distribution in (11) is the Dirichlet distribution

$$D(a_0 + n_1, ..., a_0 + n_K). \tag{19}$$

The fifth step in the Gibbs sampler given in (12) actually decomposes in additional $m$ steps because we update one $Y_j^*$ conditional on all other $Y^*$ variables. The $m$ Gibbs sampler steps are described as follows

$$[Y_1^*|C, \mu, \Sigma, Y_1, Y_j^*, j \neq 1] \tag{20}$$

$$[Y_2^*|C, \mu, \Sigma, Y_2, Y_j^*, j \neq 2] \tag{21}$$

$$\cdots$$

$$[Y_m^*|C, \mu, \Sigma, Y_m, Y_j^*, j \neq m] \tag{22}$$

Because

$$[Y*|C] \sim N(\mu_c, \Sigma_c) \tag{23}$$

we get that for $i = 1, ..., m$

$$[Y_i^*|C, \mu, \Sigma, Y_j^*, j \neq i] \sim N(\alpha_{ci} + \sum_{j \neq i} \beta_{cij} Y_j^*, \sigma_{ci}). \tag{24}$$

When we condition further on the observed value of $Y_i$ we get that the needed posterior distribution is simply the distribution given in (24) truncated above zero if $Y_i = 1$ or truncated below zero if $Y_i = 0$.

Note here that when $\Sigma = I$, i.e., the estimated model is actually the conditional independence LCA model, a more efficient MCMC algorithm exist, which is also implemented in Mplus 6.11, and that algorithm generates the latent variables $C$ and $Y^*$ together, i.e., steps 3 and 5 are combined in one step

$$[C, Y^*|\mu, Y^*, Y, q] \sim [C|\mu, Y, q][Y^*|\mu, Y, C]. \tag{25}$$

Both distributions on the LHS of (25) are easily derived. The fewer blocks the Gibbs sampler has the more efficient the estimation is, as it avoids high correlations between parameters from different blocks that can produce slow mixing.

Note also that the above algorithm does not easily extend to models for ordered polytomous observed variables. This is because ordered polytomous variable will use 2 or more threshold parameters that cannot be converted to means of $Y^*$ parameters as in the binary case.

In this paper we will not discuss the issue of label switching, which has been a difficult problem to tackle in the past. However, recent methodological advances have largely resolved this problem. There are a number of different solutions available. The two simplest ones are perhaps to fix a certain number of observations to particular classes, or to introduce inequality constraints among the parameters which identify the classes uniquely. In Mplus the second approach is implemented. However, in our simulation studies we used large sample sizes, and at such sample size levels, label switching does not occur or it occurs very rarely. Thus we will not discuss further the issue of label switching.

## 4. Bayes conditional dependence LCA models and the role of the prior in model definition

Here we define several different conditional dependence models and show that by selecting different priors for the correlation matrix we essentially construct different models. The general non-independence model is given by (6) where $\Sigma_c$ is an unrestricted tetrachoric correlation matrix. The prior distribution for $\Sigma_c$ is the marginal correlation distribution of $IW(\Sigma_{0c}, f_c)$. By varying this prior we can construct different models.

**Model 1: Unrestricted LCA.**

For this model we estimate an unrestricted correlation matrix $\Sigma_c$ using as a prior the marginal correlation distribution of $IW(I, m+1)$, where $I$ is the identity matrix and $m$ is the number of class indicators. This prior can be considered uninformative because the marginal distribution of the correlation parameters is the uniform prior on the interval [-1,1]. This prior is the default prior in Mplus.

Note that the unrestricted LCA is an identifiable model even if we use the maximum-likelihood estimator. To see this note that the latent classes are determined primarily by the mean parameters $\mu_c$. If the classes are sufficiently well separated, then estimating the tetrachoric correlations within each class amounts to estimating the tetrachoric correlations in a multiple group analysis. We can also verify that the model is identifiable by computing the total number of parameters in the model and the degrees of freedom. For example, for a 2 class model with $m$ binary variable we have $2^m - 1$ degrees of freedom while the unrestricted LCA model has $m^2+m+1$ parameter. Since $2^m-1 \geq m^2+m+1 \Leftrightarrow m \geq 5$ we conclude that the unrestricted 2-class LCA model can be estimated even with only 5 indicators.

**Model 2: Partial correlation LCA.**

For this model we estimate only some of the tetrachoric correlation parameter in $\Sigma_c$ with prior set to the marginal correlation distribution of $IW(I, m_j+1)$, where $I$ is the identity matrix and $m_j$ is the size of the identity matrix, i.e., the size of the block of free tetrachoric correlation parameters. This prior is again the default prior in Mplus and can also be considered uninformative because the resulting marginal distributions are again the uniform priors on the interval [-1,1]. The estimated partial correlation matrix should be a block diagonal partial correlation matrix, i.e., there are certain restrictions of this correlation matrix that need to be satisfied. Those restrictions are imposed by the MCMC estimation described in the previous section. Essentially the restrictions imply that if we estimate the tetrachoric correlation between $Y_1$ and $Y_2$ and the tetrachoric correlation between $Y_2$ and $Y_3$ we have to also estimate the tetrachoric correlation between $Y_1$ and $Y_3$ to make a full diagonal block in the tetrachoric correlation matrix $\Sigma_c$.

**Model 3: Exploratory LCA.**

For this model we estimate an unrestricted correlation matrix $\Sigma_c$ with prior set to the marginal correlation distribution of $IW(I, f)$, where $I$ is the identity matrix and $f > m+1$. This prior is informative. The value of $f$ changes the model and the level of informativeness of the prior. By increasing the parameter $f$ the prior forces more independence between the indicators variable and decreasing the parameter $f$ results in a model that allows more dependence between the indicators. The marginal distribution for all correlations, see Barnard et. al. (2000), is the symmetric Beta distribution $B((f-m+1)/2, (f-m+1)/2)$ on the interval [-1,1] with mean 0 and variance

$$\frac{1}{f-m+2}. \tag{26}$$

The above variance formula can be used to make a proper choice for the parameter $f$. For

example, if we want the prior for the correlation parameters to have a variance of 0.01 and a standard deviation of 0.1 then $f$ should be $98 + m$.

The Exploratory LCA (ELCA) model formalizes the belief that within class the independence of the indicator variables is only approximate , i.e., the conditional independence is only approximate and that violations of that independence are possible. Nevertheless our prior belief is that conditional independence should hold to a large extent and the parameter $f$ allows us to control to what extent we want to force independence between the indicator variables. In fact if we set $f = \infty$ we specify a model where all tetrachoric correlations are fixed to 0, i.e., the model is the standard conditional independence LCA model. Thus the ELCA model provides a bridge between the unrestricted LCA model ($f = m + 1$) and the conditional independence LCA model ($f = \infty$), i.e., it is the compromise between these two models. One can also interpret the standard conditional independence LCA as an ELCA model with super strong priors for the tetrachoric correlations, i.e., priors with zero variance.

In practical applications different values of $f$ should be explored within a standard sensitivity analysis.

The ELCA model provides a valuable alternative to the unrestricted LCA model, which may be too flexible in some practical situations and thus the ELCA may be much easier to estimate. In principle, as we increase the parameter $f$, the convergence rates, the quality of the mixing of the MCMC chains, and the variability of the estimates for the ELCA model should approach those for the standard conditional independence LCA model.

Also the ELCA model has a distinct advantage over the partial correlation LCA model because it does not require prior knowledge about what tetrachoric correlations should be included in the model. Instead the ELCA model allows the tetrachoric correlations that are not zero to escape the narrow informative priors centered at zero and thus provides a method for automatically detecting the tetrachoric correlations that should be estimated.

The ELCA model can be used as a final model or it can be used as an exploratory model followed up by a partial correlation LCA model. Based on a preliminary ELCA estimation, the largest few tetrachoric correlations (or simply those that are significant) are selected and then those correlations are estimated in the partial correlation LCA model. Such a two stage approach would yield a more parsimonious model that has all the advantages of the ELCA model. Alternatively the ELCA model can be followed up with a random effect model or a combination of a random effect model and a partial correlation model yielding again a more parsimonious model.

In some situations it may be beneficial to use a prior $IW(\Sigma_c, f)$ where $\Sigma_c$ is not the identity matrix, even when there is truly no prior information for the tetrachoric correlations. For example, consider an ELCA estimation with prior $IW(I, m + 99)$ that yields a tetrachoric correlation value of 0.5. Such a value is out of the plausible range for the marginal prior $B(50, 50)$. We can safely conclude in such a situation that the information in the data contradicts the $IW(I, m + 99)$ prior and that a different prior centered around 0.5 for that particular correlation would yield much more appropriate estimation.

## 5. Simulation Studies

In this section we present various simulation studies that illustrate the Bayesian estimation and the various conditional dependence LCA models. In all of the simulations below we use 100 replications, i.e., we generate 100 data sets and conduct the Bayesian estimation for each of these data sets.

**Table 1**: Convergence rates for 2 class unrestricted LCA model with $m = 10$ binary items

| Sample Size | 200 | 500 | 1000 | 2000 | 5000 |
|---|---|---|---|---|---|
| $\mu = 1$ | 31% | 65% | 100% | 100% | 100% |
| $\mu = 1.5$ | 90% | 98% | 97% | 94% | 100% |

### 5.1 Simulation study for the unrestricted LCA model

We generate data from a 2-class LCA model with both classes of equal size, 10 binary indicators, and the parameters $\mu_c$ are all equal within class and opposite in sign across class, i.e., $\mu_{1i} = \mu$ and $\mu_{2i} = -\mu$. The larger the value of $\mu$ the better the separation between the two classes. We generate the data using a tetrachoric correlation matrix in class one that has 3 non-zero values $\sigma_{1,12} = \sigma_{1,39} = \sigma_{1,57} = 0.2$, where $\sigma i, jk$ is the correlation between $Y_j^*$ and $Y_k^*$ in class $i$. The tetrachoric correlation matrix in class two has 1 non-zero value $\sigma_{2,46} = 0.5$. In this simulation study we evaluate the convergence rates of the Bayesian estimator of the unrestricted LCA model. We vary the sample size $n$, using 5 different sample size values: 200, 500, 1000, 2000 and 5000. We also vary the level of class separation using $\mu = 1$ and $\mu = 1.5$ which means that the class septation is 2 or 3 standard deviation units. This is a reasonable level of class separation, see Lubke and Muthén (2007). The results are presented in Table 1. It is clear that the unrestricted LCA model will be difficult to estimate unless the sample size is large or the class separation is good. In Table 2 we also present the results of the simulation study for $n = 1000$ and $\mu = 1$. Only some of the model parameters are included in this table to simplify the presentation. We see that the parameters estimates have no bias and that coverage rates are near the nominal 95%. Thus we conclude that the estimation of the unrestricted LCA model is feasible as long as the sample size is not too small and the classes are not poorly separated. It is worth noting that the positive tetrachoric correlations were not significant in all the replications even at a sample size of $n = 1000$. In fact, the correlations with true value of 0.2 were significant approximately in half of the replications. Therefore even when a correlation is estimated to a positive value there may not be enough power in the data to establish significance for that correlation. Thus a tetrachoric correlation that is not significant should not be automatically discounted as being 0. Instead the size of the correlation should also be taken into account.

Note also that when the values of $\mu$ increase the identifiability of the tetrachoric correlations is reduced. If in one of the classes the value of $\mu_{ci}$ is large by absolute value then the indicator variable $Y_i$ will be close to constant and thus any correlation between that variable and another variable will be difficult to identify.

### 5.2 Simulation study for the ELCA model

In this section we generate the data as in the previous section and estimate the ELCA model where the prior for $\Sigma_c$ is set to be the marginal correlation distribution of $IW(I, 15)$, i.e., all tetrachoric correlations have a marginal beta prior $B(3, 3)$ with mean 0 and variance $1/7$. We estimate the model for various sample sizes and class separations. For the ELCA the convergence rate is 100% in all cases presented in Table 1. Thus we conclude that the ELCA model is easier to estimate than the unrestricted LCA model, i.e., even a slight increase in the second parameter, i.e. the degrees of freedom parameter, of the Inverse Wishart prior can improve the convergence rate. In this example the convergence rate improved by increasing the degrees of freedom parameter from 11 for the unrestricted LCA

**Table 2**: Unrestricted LCA model with 2 classes, $m = 10$ binary items, $n = 1000$ and $\mu = 1$

| Parameter | True Value | Average Estimate | Coverage | Significant |
|---|---|---|---|---|
| $\sigma_{1,12}$ | 0.20 | 0.20 | 93% | 51% |
| $\sigma_{1,13}$ | 0.00 | 0.01 | 89% | 11% |
| $\sigma_{1,38}$ | 0.00 | 0.00 | 93% | 7% |
| $\sigma_{1,39}$ | 0.20 | 0.21 | 96% | 45% |
| $\sigma_{1,56}$ | 0.00 | 0.03 | 91% | 9% |
| $\sigma_{1,57}$ | 0.20 | 0.21 | 93% | 60% |
| $\mu_{1,1}$ | 1.00 | 0.99 | 94% | 100% |
| $\sigma_{2,45}$ | 0.00 | 0.00 | 96% | 4% |
| $\sigma_{2,46}$ | 0.50 | 0.51 | 97% | 100% |
| $\mu_{2,1}$ | -1.00 | -1.00 | 96% | 100% |

to 15 for the ELCA.

In Table 3 we present the results of the simulation study for $n = 1000$ and $\mu = 1$. These results correspond to those presented in Table 2 for the unrestricted LCA model. We see that the parameters estimates have small bias for those tetrachoric correlations that are positive, which also results in reduced coverage as well as reduced power to detect significance. This is of course the result of the informative prior that pushes all tetrachoric correlations towards 0. Nevertheless the positive correlations were singled out, i.e., detected by the ELCA analysis and the zero tetrachoric correlations were estimated correctly to zero.

One way to resolve the small biases of the ELCA for the positive tetrachoric correlations is to adjust the prior so that it is centered around the estimated positive value, i.e., as a bias reduction technique one can use the following approach. First estimate the ELCA model with $IW(I, f)$ then select the largest few correlations and estimate a second ELCA model where the prior is $IW(\Sigma_{0c}, f)$ and $\Sigma_{0c}$ is chosen so that the prior mean values are those positive values found in the first ELCA model. In our example $\Sigma_{0c}$ includes the 3 positive correlations in class one and the 1 positive correlation in class two. In Table 4 we present the results of such a two-stage approach. The convergence rates for this two stage approach are also 100% in all cases and the biases are reduced dramatically and the reduction in coverage and power are eliminated as well.

An alternative follow-up analysis to eliminate the biases in ELCA is to simply estimate the partial correlation LCA model where only the 3 positive correlations in class one and the 1 positive correlation in class two are estimated. All priors for the tetrachoric correlations are uniform on [-1,1], i.e., uninformative priors. The convergence rates for this analysis is again 100%. The results are presented in Table 5. In this follow-up approach the biases are eliminated and the coverage is near the nominal 95% level. Using the partial correlation LCA as a follow up analysis is preferable because it does not require incremental adjustments to the priors. This analysis simply uses uninformative priors for the estimated correlations.

Another possible follow-up analysis to the ELCA model is an ML estimation for an LCA model with random effects. Once the non-zero tetrachoric correlations in the LCA analysis have been identified by the ELCA analysis one can construct an LCA model with one random effect for each positive correlation and estimated that model with the ML estimator. Note however that the ML estimator uses numerical integration and the dimensions

**Table 3**: ELCA model with 2 classes, $m = 10$ binary items, $n = 1000$, $\mu = 1$, and prior $IW(I, 15)$

| Parameter | True Value | Average Estimate | Coverage | Significant |
|---|---|---|---|---|
| $\sigma_{1,12}$ | 0.20 | 0.15 | 95% | 38% |
| $\sigma_{1,13}$ | 0.00 | 0.01 | 94% | 6% |
| $\sigma_{1,38}$ | 0.00 | 0.00 | 99% | 1% |
| $\sigma_{1,39}$ | 0.20 | 0.15 | 96% | 35% |
| $\sigma_{1,56}$ | 0.00 | 0.01 | 97% | 3% |
| $\sigma_{1,57}$ | 0.20 | 0.16 | 98% | 44% |
| $\mu_{1,1}$ | 1.00 | 1.00 | 92% | 100% |
| $\sigma_{2,45}$ | 0.00 | 0.00 | 99% | 1% |
| $\sigma_{2,46}$ | 0.50 | 0.39 | 72% | 100% |
| $\mu_{2,1}$ | -1.00 | -1.00 | 96% | 100% |

**Table 4**: Two-stage ELCA model with 2 classes, $m = 10$ binary items, $n = 1000$, $\mu = 1$ and prior $IW(\Sigma_{0c}, 15)$

| Parameter | True Value | Average Estimate | Coverage | Significant |
|---|---|---|---|---|
| $\sigma_{1,12}$ | 0.20 | 0.18 | 96% | 56% |
| $\sigma_{1,13}$ | 0.00 | 0.01 | 94% | 6% |
| $\sigma_{1,38}$ | 0.00 | 0.00 | 100% | 0% |
| $\sigma_{1,39}$ | 0.20 | 0.18 | 100% | 48% |
| $\sigma_{1,56}$ | 0.00 | 0.01 | 98% | 2% |
| $\sigma_{1,57}$ | 0.20 | 0.19 | 98% | 57% |
| $\mu_{1,1}$ | 1.00 | 1.00 | 93% | 100% |
| $\sigma_{2,45}$ | 0.00 | 0.01 | 99% | 1% |
| $\sigma_{2,46}$ | 0.50 | 0.48 | 100% | 100% |
| $\mu_{2,1}$ | -1.00 | -1.00 | 97% | 100% |

**Table 5**: ELCA follow-up by partial correlation LCA model with 2 classes, $m = 10$ binary items, $n = 1000$, $\mu = 1$ and uninformative priors

| Parameter | True Value | Average Estimate | Coverage | Significant |
|---|---|---|---|---|
| $\sigma_{1,12}$ | 0.20 | 0.19 | 94% | 48% |
| $\sigma_{1,39}$ | 0.20 | 0.20 | 96% | 51% |
| $\sigma_{1,57}$ | 0.20 | 0.19 | 95% | 46% |
| $\mu_{1,1}$ | 1.00 | 1.01 | 92% | 100% |
| $\sigma_{2,46}$ | 0.50 | 0.50 | 97% | 100% |
| $\mu_{2,1}$ | -1.00 | -1.00 | 95% | 100% |

of the numerical integration is the number of random effects. Thus in practical applications not more than 3 or 4 random effects can be included in the model. Note also that the number of random effects is the maximum number of tetrachoric correlations within a class that will be included in the model rather than the total number of tetrachoric correlations. In our example with 3 tetrachoric correlations is class one and 1 in class two one would need 3 random effects rater than 4 as the loadings can be redefined between the classes. Nevertheless, the random effect LCA model estimated with the ML estimator is clearly limited to the number of tetrachoric correlations that can estimate. This limitation of the ML estimator is one of the clear advantages of the Bayesian methodology and the ELCA model which can accommodate any number of tetrachoric correlations.

## 6. Consequences of ignoring the conditional dependence

It is important to understand what the consequences are from ignoring the conditional dependence within class. In this section we demonstrate one such potential problem with a simple simulation study. If the conditional dependence is ignored in an LCA model we could conclude that there are more classes than the true number of classes. Very often in practical applications the analysis seems to find more classes than we can substantively interpret. Thus being able to easily accommodate conditional dependence within class to avoid spurious class formation becomes very important from a practical perspective.

Consider a two class LCA model as in Section 5.2 but only with 6 binary indicator variables and only one non-zero tetrachoric correlation, the correlation between the first two indicators in class one is $\sigma_{1,12} = 0.8$. We generate a data set of size $n = 5000$ using $\mu_{1i} = 1$ and $\mu_{2i} = -1$, for $i = 1, ..., 6$. The two classes are of equal size. One of the most popular methods for selecting the number of classes is to estimate the LCA model with several different classes and select the model with the smallest BIC value, see Nylund et. al. (2007). More precisely we estimate the LCA model with different number of classes and as we increase the number of classes the BIC values typically first decrease and then increase. After the increase of BIC occurs we select the model with the smallest BIC. Table 6 shows the BIC values for the conditional independence LCA model and the conditional dependence LCA model for different number of classes. All these models are estimated with the maximum-likelihood estimation. If the LCA analysis is restricted to conditional independence models only we will conclude that there are 3 classes. If however the conditional dependence is included in the model through a random effect then we conclude the correct number of classes is indeed 2.

Another problem that occurs when the conditional dependence is ignored is that even

**Table 6**: Deciding on the number of classes using conditional independence LCA and conditional dependence LCA

| Number of Classes | Assumption | BIC |
|---|---|---|
| 2 | independence | 32301 |
| 3 | independence | 31893 |
| 4 | independence | 31945 |
| 2 | correlation | **31858** |
| 3 | correlation | 31901 |

when the correct number of classes is analyzed the class formation is incorrect, i.e., the true latent classes are not correctly discovered and many observations are misclassified. To illustrate this problem we generate data just as in the previous example but the two classes are not of equal size. The first class contains 88.1% of the population while the second class contains 11.9% of the population. We then analyze the data using a two class LCA model where the conditional dependence is accounted for and when it is ignored. When the conditional dependence is included in the model the small class is estimated to have 12.6% of the population which is sufficiently close to the true value and we conclude that the classes are correctly identified. When the conditional dependence is ignored the small class is estimated to have 21.5% of the population and we conclude that the true latent classes were misconstrued. The estimates of the mean parameters $\mu$ are also biased when the the conditional dependence is ignored.

There may be other problems that are caused by ignoring the conditional dependence. For example, very often in practical applications too many multiple solutions are found that cannot be easily distinguished and the best solution may even be omitted due to insufficient optimization search of the log-likelihood function. These additional problems are only hypothesized however and additional research should be conducted in this direction.

## 7. The Qu, Tan and Kutner example

In this section we test this new Bayesian methodology with the real data example presented in Qu et. al. (1996). Four diagnostic binary tests are analyzed with a 2-class LCA model. In the Qu et. al. (1996) article the authors determined that the conditional dependence is not satisfied and that a model with a random effect that correlates the second and the third variable in class two provides a good fit for the data. We reach the same conclusion when we analyze the data with a 2-class ELCA model using a prior for the tetrachoric correlation matrix the marginal correlation distribution of $IW(I, 15)$. The results from the 2-class ELCA model estimation are presented in Table 7. If we look at the size of the correlation estimates one parameter stand out $\sigma_{2,23}$. If we look at the significance level, the same parameter stands out. In fact $\sigma_{2,23}$ is the only significant correlation parameter in the analysis. Thus ELCA analysis confirms the findings of Qu, Tan and Kutner (1996). Note that this model is actually not an identifiable model. The ELCA model has 21 parameters but there are only 15 degrees of freedom. Nevertheless the correct tetrachoric correlation was found.

We also conduct a follow up analysis using the partial correlation LCA model where only the $\sigma_{2,23}$ correlation is estimated and the rest of the correlations are fixed to 0. We compare the results of this Bayesian estimation with the ML estimates in Table 8. The results are presented on probability scale. The estimated parameters are $p_{ij} = P(Y_i =$

**Table 7**: ELCA analysis for the 2-class Qu, Tan and Kutner example

| Parameter | Estimate | One-sided P-value | 95% confidence interval |
|---|---|---|---|
| $\sigma_{1,12}$ | 0.14 | 0.24 | [-0.22,0.49] |
| $\sigma_{1,13}$ | 0.00 | 0.50 | [-0.42,0.42] |
| $\sigma_{1,14}$ | -0.02 | 0.47 | [-0.46,0.38] |
| $\sigma_{1,23}$ | -0.01 | 0.49 | [-0.42,0.42] |
| $\sigma_{1,24}$ | -0.11 | 0.30 | [-0.48,0.28] |
| $\sigma_{1,34}$ | 0.00 | 0.49 | [-0.41,0.41] |
| $\sigma_{2,12}$ | 0.01 | 0.49 | [-0.44,0.43] |
| $\sigma_{2,13}$ | 0.00 | 0.50 | [-0.42,0.43] |
| $\sigma_{2,14}$ | -0.02 | 0.47 | [-0.39,0.40] |
| $\sigma_{2,23}$ | **0.35** | **0.00** | **[0.13,0.55]** |
| $\sigma_{2,24}$ | 0.01 | 0.48 | [-0.41,0.42] |
| $\sigma_{2,34}$ | 0.01 | 0.49 | [-0.41,0.42] |

**Table 8**: Comparing ML and Bayes estimates on the Qu, Tan and Kutner example

| Parameter | Bayes Estimate | ML Estimate |
|---|---|---|
| $p_{11}$ | 0.97 | 0.97 |
| $p_{21}$ | 0.97 | 0.96 |
| $p_{31}$ | 1.00 | 1.00 |
| $p_{41}$ | 0.92 | 0.92 |
| $p_{12}$ | 0.00 | 0.00 |
| $p_{22}$ | 0.43 | 0.43 |
| $p_{32}$ | 0.09 | 0.09 |
| $p_{42}$ | 0.00 | 0.00 |
| $\sigma_{2,23}$ | 0.54 | 0.52 |

$0|C = j)$. The ML and the Bayes estimates are nearly identical.

## 8. Conclusions

In this article we demonstrate that with the recent development in Bayesian estimation more flexible LCA models can now be easily estimated to accommodate violations of the conditional independence assumption. In particular the full tetrachoric correlation matrix can be estimated within each class.

We also demonstrate that the ELCA model can easily and automatically discover all conditional independence violations in an LCA model. Alternatively the ELCA model can be used to simply estimate a general tetrachoric correlation matrix that can be used to discover and construct latent factors which explain the correlations. The ELCA model also has the virtue that it yields a more stable estimation than a completely unrestricted LCA model.

The Bayesian estimator, unlike the ML estimator, can accommodate any number of tetrachoric correlations in the model. The ML estimator would generally be limited to 3 or 4 tetrachoric correlations. Allowing the tetrachoric correlations in the model we can

avoid spurious class modeling and obtain LCA models that fit the data better, have fewer classes and are more connected to substantive hypothesis. None of the Bayesian methods and models discussed in this article are computationally intensive. In fact they take less than a minute to estimate in Mplus version 6.11.

The Bayesian methodology allows us to introduce a new concept in statistical modeling. With the maximum-likelihood estimation or any other frequentist estimator a parameter can either be free or fixed. With the Bayesian methodology a parameter can be in between a fixed and a free parameter, i.e., it is a hybrid parameter that is free only if the information in the data requires that parameter to be free. This concept is based on allowing informative priors for the hybrid parameters. A hybrid parameter is a parameter that is free but because it has a strong prior information it can only vary slightly within the wiggle room of the prior.

We also demonstrate in this article a new possibility to build structural models. With frequentist methods if we want to structure for example a variance covariance matrix to obtain a more parsimonious model we can either fix covariance parameters to 0, introduce constraints between the parameters in the matrix, or introduce latent factors that explain the covariances. With the Bayesian methodology we can provide structural restrictions on a model by introducing informative priors, i.e., by using hybrid parameters and providing parameter constraints through informative priors.

Further methodological advances are needed however. Posterior predictive checking is needed to evaluate model fit for the conditional dependence LCA models. Methods for comparing the different conditional dependence LCA models are also needed. The methodology described in this article can be used only with binary variables but not for ordered polytomous variables. A different estimation algorithm is needed to accommodate ordered polytomous variables. Further research is also needed to understand the impact of ignoring the conditional dependence.

## REFERENCES


Asparouhov, T. & Muthén B., (2010), "Bayesian Analysis Using Mplus: Technical Implementation," http://statmodel.com/download/Bayes3.pdf.

Barnard J., McCulloch, R. & Meng, X. L. (2000), "Modeling covariance matrices in terms of standard deviations and correlations, with applications to shrinkage," *Statistica Sinica*, 10, 1281–1311.

Gelman A., Jakulin A., Pittau M. G., Su Y. S. (2008), "A weakly infor- mative default prior distribution for logistic and other regression models," *The Annals of Applied Statistics*, 2, 1360–1383.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B.(2004) *Bayesian Data Analysis*. London, Chapman & Hall.

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

Elliott, M.R., Gallo, J.J., Ten Have, T.R., Bogner H.R., Katz I.R. (2005), "Using a Bayesian latent growth curve model to identify trajectories of positive affect and negative events following myocardial infarction," *Biostatistics*, 6, 119–43.

Lubke, G. & Muthén, B. (2007), "Performance of factor mixture models as a function of model size, covariate effects, and class-specific parameters," *Structural Equation Modeling*, 14, 26–47.

Muthén, B. (2006), "Should substance use disorders be considered as categorical or dimensional?," *Addiction*, 101 (Suppl. 1), 6–16.

Muthén, B. (2008), "Latent variable hybrids: Overview of old and new models," In Hancock, G. R., & Samuelsen, K. M. (Eds.), Advances in latent variable mixture models, 1–24. Charlotte, NC: Information Age Publishing, Inc.

Muthén, B. & Asparouhov, T. (2006), "Item response mixture modeling: Application to tobacco dependence criteria," *Addictive Behaviors*, 31, 1050–1066.

Muthén, B. & Asparouhov, T. (2011), "Bayesian SEM: A more flexible representation of substantive theory," Accepted in *Psychological Methods*

Nylund, K.L., Asparouhov, T., & Muthén, B. (2007), "Deciding on the number of classes in latent class analysis and growth mixture modeling. A Monte Carlo simulation study," *Structural Equation Modeling*, 14, 535–

569.

Qu T., Tan M.,& Kutner M.H. (1996), "Random-effects models in latent class analysis for evaluating accuracy of diagnostic tests," *Biometrics*, 52, 797–810.

van Dyk, D. A., and Meng, X. L. (2001), "The Art of Data Augmentation," *Journal of Computational and Graphical Statistics*, 10, 1–111.