[DOI] http://dx.doi.org/10.21487/jrm.2019.03.4.1.151

【연구경향】

Indeterminate Feature of Parameter Estimation in Multilevel Categorical Latent Variable

Jungkyu Park*

ABSTRACT

Structural indeterminacy among the multilevel discrete latent variables in the multilevel latent class model (MLCM) is discussed in this paper. Three scenarios - non-full-rank, independent, and permutation indeterminacy - are presented with theoretical explanations and proofs of each structural indeterminate case. Numerical examples are also included to provide intuitive and conceptual understanding of structural indeterminacy. The awareness of the structural indeterminacy in applying the MLCM to data is highlighted in the discussions. Researchers are giving examples and directions to check for problematic structures to ensure their final model has a theoretically sound latent structure when modeling data with multilevel discrete latent variables.

Keywords: Multilevel Latent Class Model, Model Indeterminacy, Model Identification

I. Introduction

Studies in psychology, education, and the social sciences have utilized latent variables to describe and conceptualize complex phenomena. Researchers

* Assistant Professor, Kyungpook University.

have captured latent variables and elegantly represented their constructs and components. A number of latent variable models have been proposed to establish a link between the observed variables and latent constructs, the best known of which is factor analysis(Jöreskog, 1969; Jöreskog and Sörbon, 1979; Gorsuch, 1983; Bollen, 1989). An important difference among these proposed models is the nature and assumed distribution of the latent components.

Factor analysis models postulate a continuous variable to describe the latent structure of the targeted concepts. However, latent class models(LCM) (Lazarsfeld and Henry, 1968; Goodman, 1974a, 1974b) hypothesized latent variables to be either discrete or categorical. Whether it is the nature of latent variables to be continuous or discrete has been debated, and several articles have focused on methods to differentiate these two opposite distributional assumptions(e.g., Haertel, 1990; Waller and Meehl, 1998; Muthén, 2006; Duncan, Stenbeck, and Brody, 1988; Steinley and McDonald, 2007). Although this differentiation has significance in terms of substantive interpretation, this paper limits the focus to cases in which discrete latent constructs are assumed.

The LCM utilizes a set of latent classes to explain observed relationships among the manifest variables. It is assumed that the relationships among manifest variables are explained by the subject's latent class membership and that different latent classes have specific patterns of probabilities in responding to items. Therefore, the LCM can be conceptualized as having several mutually exclusive and exhaustive internally homogeneous latent classes. Each subject belongs to one and only one latent class. In other words, all subjects in the same latent class have the same probabilities of responding to or endorsing a set of manifest variables. The elegance and simplicity of assumed categorical latent components offer an effective framework for many theories in the social sciences.

Vermunt(2003) proposed multilevel latent class models(MLCM) to take into

account the additional dependency induced by nested data(or multilevel data). He used random-effects approach(Bryk and Raudenbush, 1992; Goldstein, 1995; Snijders and Bosker, 1999; Rabe-Hesketh, Pickles, and Skrondal, 2001) to extend the LCM to adopt a multilevel nested data structure. He distinguished two types of random effects depending on the assumed distribution of the random effects. The first type of random effects follows a parametric continuous distribution. A normal distribution is usually hypothesized in order to facilitate parameter estimation. The second type of random effect is assumed discrete. In one example, multinomial distributed random effects replace normal distributed random effects in explaining the dependency in nested data. Discrete random effects have the advantage of weaker distributional assumptions and have less computational burden(Vermunt and Van Dijk, 2001). In practice, the discrete version of MLCM has been demonstrated to provide a good framework for application in different research fields. For example, Bijmolt, Paas, and Vermunt(2004) applied the framework to marketing studies and Dias, Vermunt, and Ramos(2010) applied it to a study in finance.

Discrete latent components offer an elegant representation for translating the research question into statistical terms. Moreover, hypothesized discrete latent quantities have been used to describe not only the nature of latent components at each level but also the dependency between levels. This paper discusses an important but often overlooked aspect of modeling data with multilevel discrete latent variables: indeterminacy. In LCM, indeterminacy usually relates to a model's identifiability, the issue of label switching, or decisions about the number of classes. However, in the context of MLCM, this paper concerns another aspect of indeterminacy, that is, structural indeterminacy. Specifically, this type of indeterminacy relates to relationships between latent components at different levels. Only a well-defined multilevel latent structure can offer valid and meaningful interpretation of the

relationships among the hypothesized discrete latent constructs. This paper illustrates three scenarios of structural indeterminacy in MLCM. These structural indeterminate scenarios will show that using multilevel discrete latent constructs to interpret relationships observed in data could be compromised.

The purpose of this paper is to draw attention to the potential harm caused by structural indeterminacy. Theoretical explanations and numerical illustrations are given to emphasize the problematic structure of MLCM. The remainder of this paper is structured as follows: First, the model specification and parameter estimation of MLCM are introduced, followed by discussions on selecting the number of latent components. Three scenarios of structural indeterminacy and their theoretical explanations of causing indeterminacy are presented, and numerical examples illustrate each scenario. The paper concludes with discussions and final remarks on the applications of MLCM in practice.

Ⅱ. Formulation and Estimation of Multilevel Latent Class Models

In the MLCM discussed here, the latent variables at different levels are assumed discrete. Therefore, the models are similar to the nonparametric approach of the multilevel latent class model presented in Vermunt's(2003) paper. However, there are several conceptual differences between the two models. Instead of utilizing discrete random effects to account for the dependency, the approach here builds a LCM at each level. Specifically, one set of latent classes is used to explain dependency at a lower level(i.e., individual level), and another LCM is built to explain dependency at a higher

level(i.e., group level). In order to differentiate the components associated with different levels, categories of the discrete latent variable at the lower level are called latent classes, and categories of the discrete latent variable at the higher level are referred to as latent clusters. For formal specification of the MLCM, a classical LCM will be introduced and then extend it to incorporate multilevel latent structures.

Let X be the latent variable consisting of M mutually exclusive and exhaustive latent classes. The latent classes are indexed by m, where $m = 1, 2, \ldots, M$. In addition, Y_{ij} denotes the random variable representing the response of subject i to the jth item($1 \le j \le J$), and y_{ij} represents a realization of the random variable Y_{ij}. The class-specific conditional probability of observing y_{ij} for item j of subject i in class m is $P(Y_{ij} = y_{ij} \mid X = m)$, which is also called the conditional response probability or latent conditional probability. In the LCM, the probability of obtaining the response pattern $\mathbf{y}_i$ for subject i ($P(\mathbf{Y}_i = \mathbf{y}_i)$) is a weighted average of M class-specific conditional response probabilities, $P(\mathbf{Y}_i = \mathbf{y}_i \mid X = m)$:

$$
\begin{aligned}
P(\mathbf{Y}_i = \mathbf{y}_i) &= \sum_{m=1}^{M} P(X = m) P(\mathbf{Y}_i = \mathbf{y}_i \mid X = m) \\
&= \sum_{m=1}^{M} P(X = m) \prod_{j=1}^{J} P(Y_{ij} = y_{ij} \mid X = m),
\end{aligned}
$$

where $P(X = m)$ denotes the probability of a randomly selected subject in the sample belonging to latent class m, which is usually called latent class probability.

Conceptually, latent class probabilities can be conceptualized as the class size of the latent class in the population. A LCM assumes local independence within each class so that the joint probability of response pattern $\mathbf{y}_i$ in class

m is the product of all conditional probabilities of items within this class(line 2 of Equation(1)). A LCM also needs to satisfy the identification constraints. Goodman(1974a, 1974b) showed that latent class models are identifiable by the constraint $\sum_{m=1}^{M} P(X=m)=1$.

In order to identify and differentiate the clustering effects in different levels in MLCM, the nested data structure should be clearly represented in the notation system. Therefore, the observed response vector is now denoted as $\mathbf{Y}_{ig}$ (instead of $\mathbf{Y}_i$) to indicate a subject's higher-level membership g(i.e., subject i is a member of group g).

A discrete random variable H_g with L possible outcomes is hypothesized to represent the L unique higher-level latent clusters. Each outcome of the discrete random variable can be conceptualized as a latent cluster consisting of homogeneous(higher level) groups sharing the same latent class distribution. This discrete random variable requires a parameter $P(H_g=l)$ to describe the distributions(i.e., cluster sizes) of the latent clusters. The differences between clusters are demonstrated in the latent class probabilities of each cluster. In other words, the latent class probabilities $P(X_{ig}=m)$ differ depending on the latent cluster's membership(H_g). They are represented specifically as $P(X_{ig}=m \mid H_g=l)$. The matrix $P(X \mid H)$ of size $M \times L$ is used to represent the estimated M latent class probabilities for the L clusters. The columns of the matrix $P(X \mid H)$ describe the within-cluster distributions of the latent classes. Accordingly, the entries of each column sum to one. Conceptually, each latent cluster can have a unique pattern of conditional response probabilities. Because different conditional response probabilities among clusters can be ambiguous and subject to questionable interpretations, the latent clusters are assumed to have no effect on the conditional response probabilities.

The parameters of MLCM can be estimated using the maximum

likelihood(ML) approach. The likelihood function is based on the probability density for the data of the higher-level unit. The log-likelihood to be maximized equals

$$\log L = \sum_{g=1}^{G} \log P(\mathbf{Y}_g \mid \mathbf{H}_g).$$

The ML estimates of the model parameters are obtained by a modified EM algorithm. The E-step modified by Vermunt(2003) is called the upward-downward procedure, which makes use of the conditional independence assumptions underlying the MLCM. The latent variables are summed out by going from lower- to higher-level units, and the marginal posteriors are subsequently obtained by going from higher- to lower-level units. The details of this algorithm are found in Vermunt(2003), and the outline of this algorithm is described in the Appendix.

Ⅲ. Decisions on the Number of Latent Classes and Clusters

One important step of fitting empirical data to MLCM is to decide on the numbers of latent clusters and classes. This decision is the main research question in the task of model selection. There have been extensive studies on the issue of selecting the number of classes for LCM. For example, researchers have shown that popular chi-squared (χ^2) likelihood ratio statistics are not theoretically correct for model selections in LCM because the ratio between the two latent class models may not follow the χ^2 distribution if the reduced model($M-1$ classes) is obtained from the full model(M classes) by

placing parameters at their boundary values(Everitt and Hand, 1981; Aitkin, Anderson, and Hinde, 1981; Everitt, 1988). Various information-based criteria were suggested as alternative indices for selecting the number of classes for LCM.

The Akaike information criterion(AIC)(Akaike, 1973) and the Bayesian information criterion(BIC)(Schwarz, 1978) are probably the most common methods for model selection. The model with the lowest value of information criteria among several candidate models is considered the best-fit model. Several researches(e.g., Collins, Fidler, Wugalter, and Long, 1993; Lin and Dayton, 1997; Yang, 2006; Nylund, Asparouhov, and Muthén, 2007) have used simulation studies to systematically evaluate and examine the performance of several popular information criteria(e.g., AIC and BIC) and a few modified versions of information criteria(e.g., CAIC and adjusted BIC) on correctly uncovering the true numbers of latent classes in LCM.

Because there are two levels of latent components(i.e., latent classes and latent clusters) in MLCM, deciding the number of latent classes and clusters simultaneously becomes much more challenging. Moreover, the definition of "sample size" is unclear in the context of nested data structure. If the information criteria are used to decide the number of latent classes and clusters, the definition of "sample size" for some criteria that have the sample size as part of the penalty term(e.g., BIC and CAIC) is unclear in the context of nested data structure. For example, the sample size(N) in the formula can be the number of groups, the number of individuals, or the number of total sample size.

The task of simultaneous decisions on the number of both latent classes and clusters has received scant attention in the literature. However, Lukočienė and Vermunt(2010) conducted simulation studies to examine the performance of various model selection methods in deciding the number of higher-level

components. Lukočienė, Varriale, and Vermunt(2010) further extended the investigation on stimulating the performance of various criteria to determine the number of latent classes and clusters. They also proposed a stepwise model-fitting procedure to choose the number classes and clusters more effectively. These two papers shed light on deciding the quantities of latent components at either the lower level or the higher level. The indeterminacy relating to decisions on the number of latent classes and clusters is essential in MLCM. In this paper, however, the focus is on another type of indeterminacy: structural indeterminacy between the lower level and the higher level.

Ⅳ. Structural Indeterminacy among between-level Discrete Latent Components

Structural indeterminacy in MLCM concerns the relationships among discrete latent components between latent clusters and latent classes. Two definitions facilitate discussions of indeterminacy are given first, which are followed by three structural indeterminacy scenarios in MLCM. The theoretical explanations and proofs for each scenario are discussed separately.

Definition 1. An MLCM solution is said to achieve the optimum if the convergence criterion of the MLCM algorithm has been met.

For any given random start, $\{\hat{P}^{(0)}(H_g = l)$, $\hat{P}^{(0)}(X_{ig} = m \mid H_g = l)$, $\hat{f}^{(0)}(y_{ig};\theta)\}$; the solution set $\{\hat{P}^{(t)}(H_g = l)$, $\hat{P}^{(t)}(X_{ig} = m \mid H_g = l)$, $\hat{f}^{(t)}(y_{ig};\theta)\}$ obtained from the MLCM parameter estimation algorithm is said to arrive at the optimum if the algorithm terminates at the iteration t, which is within the maximally

allowable(pre-determined) number of iterations and is smaller than the convergence criterion.

Definition 2. An MLCM problem is identifiable(unambiguous or determined) if and only if its solution is optimal and unique from distinct random starts.

For distinct random starts, $\{\hat{P}^{(0)}(H_g = l)$, $\hat{P}^{(0)}(X_{ig} = m \mid H_g = l)$, $\hat{f}^{(0)}(y_{ig};\theta)\}$; the MLCM is said to be identifiable if the estimation algorithm arrives at the same solution set $\{\hat{P}^{(t)}(H_g = l)$, $\hat{P}^{(t)}(X_{ig} = m \mid H_g = l)$, $\hat{f}^{(t)}(y_{ig};\theta)\}$ within the maximally allowable number of iterations.

1. Non-full-rank Scenario

Let $P(X) \in R_+^{M\times 1}$ and $P(H) \in R_+^{L\times 1}$ denote the latent class probabilities and the latent cluster probabilities, respectively. The dependency of the latent classes and clusters in MLCM, that is, $P(X \mid H) \in R_+^{M\times L}$, are thus expressed as follows:

$$P(X \mid H) \stackrel{\text{def}}{=} \begin{bmatrix} P(X = m_1 \mid H = l_1) & \cdots & P(X = m_1 \mid H = l_L) \\ P(X = m_2 \mid H = l_1) & \cdots & P(X = m_2 \mid H = l_L) \\ \vdots & & \vdots \\ P(X = m_M \mid H = l_1) & \cdots & P(X = m_M \mid H = l_L) \end{bmatrix}.$$

The above matrix relates each latent cluster to its corresponding classes. According to the specifications of MLCM, the sum of the jth column of this matrix is 1 for all $j\,(j = 1,\ldots,L)$.

The number of solution sets of a linear system can be used to explain the indeterminacy of this scenario. The relationship between $P(X)$ and $P(H)$ can be related in terms of a linear system: $P(X \mid H)P(H) = P(X)$. From the theory of linear systems, the following relationships are known: 1) if the linear system

is consistent(i.e., the solution set is non-empty) and if the coefficient matrix $P(X|H)$ is full rank, then $P(X|H)P(H)=P(X)$ has a unique solution; 2) if $P(X|H)P(H)=P(X)$ is consistent but $P(X|H)$ is less than full rank, then $P(X|H)P(H)=P(X)$ has an infinite number of solutions. This implies that if $P(X|H)$ is not full rank, then the solution of the MLCM parameters is not identifiable since the solution set is not unique according to Definition 2. More specifically, when $P(X|H)$ is a non-full-rank matrix, then infinitely many $\hat{P}(H)$ that give $P(X|H)\hat{P}(H)=P(X)$ can be obtained.

2. Independent Scenario

The second case of structural indeterminacy occurs when the relationship between latent classes and clusters is independent. In other words, the compositions and sizes of each class are identical across all latent clusters. This independent scenario violates the fundamental assumption made in multilevel modeling, that is, the lower-level parameters depend on the status of higher-level parameters. Conditional independence cannot be tested using empirical data, but it has conceptual importance in statistical modeling. Without this dependency between levels, applying multilevel models to data may lead to invalid conclusions derived from explanations of dependency observed in the data.

Indeterminacy caused by the independent scenario can be formally explained by simple operations based on probability theory. When the relationships between a latent class m and a latent cluster l are independent of each other, the conditional probability describing the cross-level relationship is

$$P(X \mid H) = \frac{P(X \cap H)P(H)}{P(H)} = \frac{P(X)P(H)}{P(H)} = P(X).$$

Hence, the matrix describing the cross-level relationship $P(X \mid H)$ can be expressed as a matrix form by the concatenate L vector of $P(X)$ horizontally:

$$P(X \mid H) = [\underbrace{P(X) \cdots P(X)}_{\text{L?}}]_{M \times L}.$$

It is obvious that the cross-level relationship summarized as $P(X \mid H)$ in Equation(2) is now rank one. According to the non-full rank scenario discussed earlier, the independent scenario also results in a non-identifiable situation and thus leads to structural indeterminacy. Although this independent scenario can be included as a special case of *non-full-rank* scenario, it is discussed separately because the dependency between levels is the fundamental assumption in multilevel modeling. The independent scenario violates this assumption and thus jeopardizes the usages and applications of MLCM.

3. Permutation Indeterminacy Scenario

Redner and Walker(1984) first introduced the term label switching to describe invariance of likelihood under relabeling the mixture components. This problem is commonly noted in parameter estimation and clustering in the Bayesian approach, and is discussed in LCM and mixture models(e.g., Chung, Loken, and Schafer, 2004; Yang, 2006). Common suggestions to remedy this issue are to impose constraints on the parameter space(e.g., Stephens, 2000; Celeux, Hurn, and Robert, 2000). Within the context of MLCM, the problem of label switching can lead to structural indeterminacy, which is more

severe than classical LCM because it has two levels of latent components.

The simplest examples of this problem are when the posterior probability of observing an individual coming from each of the M classes is equal or when randomly selected groups from each L cluster are equiprobable. In such cases, the labeling of these classes and clusters is arbitrary. Another easy way to comprehend this indeterminacy is the concept of likelihood invariance: permutations of the latent components, that is, rearrangement of the component indices will not change the likelihood value.

The concept of permutation matrix can be utilized to describe this type of structural indeterminacy in MLCM. The function of a permutation matrix is to permute the rows and columns of a matrix, and a permutation matrix is formed by permuting rows or columns of an identity matrix to a desired order. One can find $M \times M$ permutation matrix $\mathbf{R}$; pre-multiplying the matrix $P(X|H)$ with $\mathbf{R}$ reorders the rows of $P(X|H)$. Similarly, anther permutation matrix $\mathbf{C}$ of size $L \times L$ can be defined; post-multiplying $P(X|H)$ by $\mathbf{C}$ reorders the columns of $P(X|H)$.

The operation of pre- and post-multiplication $P(X|H)$ by permutation matrices $\mathbf{R}$ and $\mathbf{C}$ can be repeated many times. For example, if rows are permuted S times with $\mathbf{R}_1, \mathbf{R}_2, \ldots, \mathbf{R}_S$, and $\mathbf{C}_1, \mathbf{C}_2, \ldots, \mathbf{C}_T$ are used to permute columns $P(X|H)$ matrix T times, the configuration of the $P(X|H)$ after the permutation is $\mathbf{R}_S \ldots \mathbf{R}_2 \mathbf{R}_1 P(X|H) \mathbf{C}_1 \mathbf{C}_2 \ldots \mathbf{C}_T$. Some combinations of $\mathbf{R}_s$ and $\mathbf{C}_t$ exist that can permute the matrix back to the original configuration of $P(X|H)$. Specifically, this permutation process is

$$P(X|H) = R_S \ldots R_2 R_1 P(X|H) C_1 C_2 \ldots C_T.$$

Since the combinations of $\mathbf{R}_s$ and $\mathbf{C}_t$ are not unique, the elements of $P(X|H)$

can be permuted in different ways, but they can still permute back to a particular configuration. Note that a constraint is imposed on $P(X|H)$ in MLCM, that is, each column will sum to 1, so the choices of permutation matrix are not as free as in permuting rows and columns of a matrix without constraint. However, this constraint does not confine the combination of $\mathbf{R}_s$ and $\mathbf{C}_t$ as unique for securing structural determinacy in $P(X|H)$.

V. Numerical Examples and Illustrations

In this section, some numerical examples are used to illustrate the three indeterminate scenarios. These examples are designed to serve as intuitive examples to illustrate structural indeterminacy. Therefore, the easiest cases are used as examples. Note that these examples are for demonstration purposes only and many other general cases could be used for each scenario. In addition, because the MLCM is a probabilistic model, the discussions of indeterminacy here are situated at the conceptual level rather than in the context of numerical analysis.

The top panel of Table 1 illustrates examples of the *non-full-rank* scenario. Three numerical examples of different combinations in the number of latent classes(M) and clusters(L) are presented. The first two examples have identical distributions of the latent classes in both latent clusters, and the two-classes and three-clusters case have identical distribution in the first two columns but not in the third one. Simple values are chosen to illustrate this scenario; however, identifying the non-full-rank scenario in empirical data may not be easy or straightforward.

Numerical examples of the *independent scenario* are presented in the middle panel of Table 1. The first two examples are the equal-probable conditions

in which the probability of assigning a randomly sampled subject in each class is the same for both clusters. The next two examples illustrate fixed and identical ordering of class sizes for each latent cluster. Specifically, the likelihood of the assignment of a random sampled subject to a particular class is the same for both clusters.

Table 1. Numerical Examples of the Three Structural Indeterminate Scenarios

Examples of Non-full-Rank Scenario				
M/L	2/2	3/2	2/3	
$P(X \mid H)$	$\begin{bmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{bmatrix}$	$\begin{bmatrix} 1/3 & 1/3 \\ 1/3 & 1/3 \\ 1/3 & 1/3 \end{bmatrix}$	$\begin{bmatrix} 1/2 & 1/2 & 1/3 \\ 1/2 & 1/2 & 2/3 \end{bmatrix}$	
Examples of Independent Scenario				
M/L	2/2	3/2	2/3	3/2
$P(X \mid H)$	$\begin{bmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{bmatrix}$	$\begin{bmatrix} 1/3 & 1/3 \\ 1/3 & 1/3 \\ 1/3 & 1/3 \end{bmatrix}$	$\begin{bmatrix} 1/2 & 1/2 & 1/3 \\ 1/2 & 1/2 & 2/3 \end{bmatrix}$	$\begin{bmatrix} 1/6 & 1/6 \\ 2/6 & 2/6 \\ 3/6 & 3/6 \end{bmatrix}$
Examples of Permutation Indeterminacy				
M/L	2/2	3/2	3/3	
$P(X \mid H)$	$\begin{bmatrix} 1/3 & 2/3 \\ 2/3 & 1/3 \end{bmatrix}$	$\begin{bmatrix} 1/4 & 2/4 \\ 1/4 & 1/4 \\ 2/4 & 1/4 \end{bmatrix}$	$\begin{bmatrix} 1/6 & 2/6 & 3/6 \\ 2/6 & 3/6 & 1/6 \\ 3/6 & 1/6 & 2/6 \end{bmatrix}$	

Examples for permutation indeterminacy are listed in the bottom portion of Table 1. These examples illustrate permutation indeterminacy when the patterns of the latent classes can be rotated into an identical pattern for each latent cluster. In the example of the two-classes and two-clusters model, one-third of individuals are categorized in latent Class 1 and two-thirds are in Class 2 for latent Cluster 1. However, the pattern of the two classes is reversed in Cluster 2: one-third of individuals are in latent Class 2 and two-thirds are in Class 1. Examples of permutation independency for the three-classes with two- and three-clusters case are also shown in Table 1.

Ⅵ. Discussions

The discussion of indeterminacy is not new regarding latent variable models. The factor indeterminacy was an important topic in the earlier development of factor analysis(Schönemann and Wang, 1972; McDonald, 1974, 1977; Steiger, 1979). A review of this topic is found in Steiger(1996). The LCM can be considered a discrete or categorical analogy to factor analysis. However, the discussion of indeterminacy within the traditional factor analysis framework is relatively more straightforward and obvious, but it is not as intuitive and apparent in the probability-based LCM. The earliest discussion on indeterminacy(unidentifiable) in latent class analysis appeared in Goodman's(1974) paper in *Biometrika*. However, since then few studies have focused on this issue when various models were extended from the LCM.

One possible reason for overlooking the issue of structural indeterminacy might be the lack of information about the estimation process. Usually, programs and software used for estimating MLCM do not automatically provide detailed information regarding the estimation process. There are no “warming” messages available in the output or log window when some problematic structures occur. Without information about problematic latent structure, users may easily miss checking for possible structural indeterminacy.

Moreover, different programs may have different routines to handle the problematic operation(e.g., non-full-rank). The solutions are usually based on the default settings of a specific software or program. For example, different kinds of software use different methods to handle the noninvertible matrix in the iteration process. The default settings of software are usually not familiar to users and are explained only in the technical manuals. Therefore, possible indeterminate structures may be masked by the default settings of the software.

When analyzing empirical data, researchers often rely on goodness-of-fit

tests or information criteria to identify the best-fit model among the candidate models. However, these tests and criteria are mainly good for selecting the "numbers" of latent components but not for detecting structural indeterminacy among components. Because there are no indices or tests to detect improper multilevel discrete latent structures, additional evaluations should be performed independently to check for possible structural indeterminacy in addition to making decisions on the number of latent components.

Ⅶ. Conclusion

To the best of my knowledge, since LCMs were extended to accommodate the nested data structure, no paper has discussed this important modeling concept in detail under a multilevel framework. This paper aimed to fill this gap by providing theoretical discussions and interpretations of structural indeterminacy in the context of multilevel discrete latent constructs. Three different scenarios: non-full rank, independent, and permutation indeterminacy were presented to illustrate structural indeterminacy in MLCM. Numerical examples corresponding to the three scenarios were also included to provide intuitive and conceptual understanding of the concept of structural indeterminacy.

In sum, a discrete version of MLCM has many advantages, among which is the potential use in empirical research. One example is its flexibility in translating hypothetical theories into the elements of a statistical model. However, the hypothesized multilevel latent structural model may be an indeterminate structure and thus lead to invalid interpretations of the final model. The awareness of this issue in applying MLCM to data is important in practice. The theoretical explanations of structural indeterminacy are

highlighted in this paper. The numerical examples provided give researchers starting templates and directions to check for problematic structures. I hope that the discussions of structural indeterminacy presented in this paper inspire and encourage researchers to check whether their final model has a theoretically sound structure when modeling data with multilevel discrete latent variables.

References

Aitkin, M., Anderson, D., and Hinde, J. 1981. "Statistical modeling of data on teaching styles with discussion." *Journal of Royal Statistical Society, Series A* 144: 419-461.

Aitkin, M., and Rubin, D. B. 1985. "Estimation and hypothesis testing in finite mixture models." *Journal of Royal Statistical Society, Series B* 47: 67-75.

Akaike, H. 1973. "Maximum likelihood identification of Gaussian autoregressive moving average models." *Biometrika* 60: 255-265.

Bijmolt, T. H., Paas, L. J., and Vermunt, J. K. 2004. "Country and consumer segmentation: Multi-level latent class analysis of financial product ownership." *International Journal of Research in Marketing 21*: 323-340.

Bollen, K. A. 1989. *Structural equations with latent variables*. New York: Wiley.

Bryk, A. S., and Raudenbush, S. W. 1992. *Hierarchical linear models: Applications and data analysis methods*. Newbury Park, CA: Sage.

Celeux, G., Hurn, M., and Robert, C. D. 2000. "Computational and inferential difficulties with mixture posterior distributions." *Journal of the American Statistical Association* 95: 957-979.

Chung, H., Loken, E., and Schafer, J. L. 2004. "Difficulties in drawing inferences with finite-mixture models: A simple example with a simple solution." *The American Statistician* 58: 152-159.

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

Dias, J. G., Vermunt, J. K., and Ramos, S. 2010. "Mixture hidden Markov models in finance research." In Fink, A., Lausen, B., Seidel, W. and Ultsch,

A. eds., *Advances in data analysis, data handling and business intelligence* pp. 451-459. Berlin-Heidelberg: Springer.

Duncan, O. T., Stenbeck, M., and Brody, C. 1988. "Discovering heterogeneity: continuous versus discrete latent variables." *American Journal of Sociology* 93: 1305-1321.

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

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

Goldstein, H. 1995. *Multilevel statistical models*. 2nd edn. New York: Halstead Press.

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

Gorsuch, R. L. 1983. *Factor analysis*. Hillsdale, NJ: Lawrence Erlbaum.

Haertel, E. 1990. "Continuous and discrete latent structure models for item response data." *Psychometrika* 55: 477-494.

Jöreskog, K. G. 1969. "A general approach to confirmatory maximum likelihood factor analysis." *Psychometrika* 34: 183-202.

Jöreskog, K. G., and Sörbom, D. 1979. *Advances in factor analysis and structural equation models*. New York: University Press of America.

Lin, T. H., and Dayton, C. M. 1997. "Model selection information criterion for non- nested latent class models." *Journal of educational and behavioral statistics* 22: 249-264.

Lukočienė, O., Varriale, R., and Vermunt, J. K. 2010. "The simultaneous decision about the number of lower- and higher-level classes in multilevel latent class analysis." *Sociological Methodology* 40: 247-283.

Lukočienė, O., and Vermunt, J. K. 2010. "Determining the number of components in mixture models for hierarchical data." In Fink, A. and Lausen, B. and Seidel, W. and Ultsch Eds., *Advances in data analysis,*

data handling and business intelligence pp. 241-249. Berlin-Heidelberg: Springer.

McDonald, R. P. 1974. "The measurement of factor indeterminacy". *Psychometrika* 39: 203-221.

McDonald, R. P. 1977. "The indeterminacy of components and the definition of common factors." *British Journal of Mathematical and Statistical Psychology* 30: 165-176.

Nylund, K. L., Asparouhov, T., and 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.

Rabe-Hesketh, S., Pickles, A., and Skrondal, A. 2001. "GLLAMM: A general class of multilevel models and a Stata program." *Multilevel Modelling Newsletter* 13: 17-23.

Redner, R. A., andWalker, H. F. 1984. "Mixture densities, maximum likelihood and the EM algorithm." *SIAM Review* 26: 195-239.

Schönemann, P. H., and Wang, M. M. 1972. "Some new results on factor indeterminacy." *Psychometrika* 37: 61-91.

Schwarz, G. 1978. "Estimating the dimensions of a model." *Annals of Statistics* 6: 461-464.

Snijders, T. A. B., and Bosker, R. J. 1999. *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. Thousand Oaks, CA: Sage.

Steiger, J. H. 1979. "Factor indeterminacy in the 1930s and 1970s: Some interesting parallels." *Psychometrika* 44: 157-167.

Steiger, J. H. 1996. "Coming full circle in the history of factor indeterminacy." *Multivariate Behavioral Research* 31: 617-630.

Steinley, D., and McDonald, P. R. 2006. "Examining factor score distributions to determine the number of latent spaces." *Multivariate Behavioral Research* 42: 133-156.

Stephens, M. 2000. “Dealing with label switching in mixture models.” *Journal of the Royal Statistical Society, Series B* 62: 795-809.

Vermunt, J. K. 2003. “Multilevel latent class models.” *Sociological Methodology* 33: 213-239.

Vermunt, J. K., and Van Dijk, L. 2001. “A nonparametric random-coefficients approach: The latent class regression model.” *Multilevel Modelling Newsletter* 13: 6-13.

Waller, N. G., and Meehl, P. E. 1998. *Multivariate taxometric procedures: Distinguishing types from continua*. Newbury Park, CA: Sage.

Yang, C. C. 2006. “Evaluating latent class analysis models in qualitative phenotype identification.” *Computational Statistics and Data Analysis* 50: 1090-1104.

Appendix
MLCM Parameter Estimation Algorithm

Initialization: $t \leftarrow 0$, $\mathbf{log}\boldsymbol{L}(\boldsymbol{\theta})_{\mathbf{max}} \leftarrow 10^7$

Input: $\hat{\boldsymbol{P}}^{(0)}(\boldsymbol{H_g} = \boldsymbol{l})$, $\hat{\boldsymbol{P}}^{(0)}(\boldsymbol{X_{ig}} = \boldsymbol{m} \mid \boldsymbol{H_g} = \boldsymbol{l})$, and $\hat{\boldsymbol{f}}^{(0)}(\boldsymbol{y_{ig}};\boldsymbol{\theta})$

Compute $\log L^{(t)}(\theta)$

While $|\mathbf{log}\boldsymbol{L}^{(t)}(\boldsymbol{\theta}) - \mathbf{log}\boldsymbol{L}(\boldsymbol{\theta})^{\mathbf{max}}| > 10^{-6}$ or $t \le 10^5$ do

$\log L(\theta)^{\max} \leftarrow \log L^{(t)}(\theta)$ and $t \leftarrow t+1$

E-step:

$$h_{igm|l}^{(t)} = \hat{P}^{(t-1)}(X_{ig} = m \mid H_g = l)\hat{f}^{(t-1)}(y_{ig},\theta)$$

$$g_{ig|l}^{(t)} = \sum_{m=1}^{M} h_{igm|l}^{(t)}$$

$$\hat{X}_{igm|l}^{(t)} = h_{igm|l}^{(t)} / g_{ig|l}^{(t)}$$

$$\hat{H}_{gl}^{(t)} = \hat{P}^{(t-1)}(H_g = l)\prod_{i=1}^{n_g} g_{ig|l}^{(t)} / \sum_{l=1}^{L}\hat{P}^{(t-1)}(H_g = l)\prod_{i=1}^{n_g} g_{ig|l}^{(t)}$$

M-step:

$$\hat{P}^{(t)}(H_g = l) = \sum_{g=1}^{G}\hat{H}_{gl}^{(t)} / \sum_{g=1}^{G}\sum_{l=1}^{L}\hat{H}_{gl}^{(t)}$$

$$\hat{P}^{(t)}(X_{ig} = m \mid H_g = l) = \sum_{g=1}^{G}\sum_{i=1}^{n_g}\hat{H}_{gl}^{(t)}\hat{X}_{igm|l}^{(t)} / \sum_{g=1}^{G}\sum_{i=1}^{n_g}\sum_{m=1}^{M}\hat{H}_{gl}^{(t)}\hat{X}_{igm|l}^{(t)}$$

$$\hat{f}^{(t)}(y_{ig};\theta) = \sum_{g=1}^{G}\sum_{i=1}^{n_g}\hat{H}_{gl}^{(t)} y_{ig} / \sum_{g=1}^{G}\sum_{i=1}^{n_g}\sum_{m=1}^{M}\hat{H}_{gl}^{(t)}$$

Compute $\log L^{(t)}(\theta)$

end while

다층 범주형 잠재변수에서 모수추정의 비확정성

박 중 규*

논문요약

본 논문에서는 다층 잠재계층모형에서 모수의 구조적인 비확정성에 대해서 다룬다. 먼저 다층 잠재계층에서 non-full-rank, independent, permutation indeterminacy의 세 가지 경우에 대해 소개하고 각각의 경우에 발생할 수 있는 모수의 비확정성에 대해서 설명한다. 또한 모수의 비확정성에 대한 이해를 높이기 위해 수리적인 예시를 제시한다. 논의에서는다층 잠재계층모형을 이용해서 자료를 분석할 때 구조적 불확정성에 대한 인식이 중요하다는 점을 강조한다. 덧붙여, 연구자가 선택한 최종모형이 적절한 잠재 구조를 가졌는지를 점검하기 위한 지침을 제공한다.

주제어: 다층 잠재계층 모형, 모형 비확정성, 모형 식별

투고일: 2019.02.16. **심사일:** 2019.03.08. **게재확정일:** 2019.03.12.

* 경북대학교 심리학과, 조교수.