



UW Biostatistics Working Paper Series

4-11-2014

Nonparametric Identifiability of Finite Mixture Models with Covariates for Estimating Error Rate without a Gold Standard

Zheyu Wang

Johns Hopkins University, wangzy@jhu.edu

Xiao-Hua Zhou

University of Washington, azhou@u.washington.edu

Suggested Citation

Wang, Zheyu and Zhou, Xiao-Hua, "Nonparametric Identifiability of Finite Mixture Models with Covariates for Estimating Error Rate without a Gold Standard" (April 2014). *UW Biostatistics Working Paper Series*. Working Paper 403.
<http://biostats.bepress.com/uwbiostat/paper403>

This working paper is hosted by The Berkeley Electronic Press (bepress) and may not be commercially reproduced without the permission of the copyright holder.

Copyright © 2011 by the authors

1 Introduction

Finite mixture models include a set of models that describe a heterogeneous population as a mixture of some homogenous but potentially unknown components. These models provide a flexible approach to modeling unobserved constructs and to explaining population heterogeneity, therefore are gaining popularity in many disciplines. For example, in diagnostic medicine, the evaluation of a diagnostic test can be difficult when the true disease status, also called the gold standard, is unknown due to time or cost constraints, ethics issues or the lack of necessary biotechnology. In such situation, latent class models (finite mixture models with discrete manifest variables), have been adopted to study the relation between the test results and the unobserved disease status [1, 2, 3]. Recent years, the advance in biomarker discovery promoted another family of finite mixture models, latent profile models (finite mixture models with continuous manifest variables), in biomarker evaluation without a gold standard. Numerous applications can also be found in psychometrics, social sciences, economics and other fields whose respective research problems involve entities that are hard to measure directly.

The popularity of finite mixture models is partially due to their flexible structure, which can be easily modified to meet the specific need of a research question. This flexibility makes finite mixture models a powerful tool, but it also renders them vulnerable to problems associated with identifiability. However, this fundamental problem of model identifiability is often ignored in practice while researchers tailoring complicated model structures. Loosely speaking, a non-identifiable model can have multiple parameter values that all correspond to the same likelihood, which means that one cannot identify the true parameter values among these choices based only on the observed data. Although identifiability is not the same as consistency, estimates from a model lacking identifiability will not be consistent. Thus, establishing identifiability is a crucial issue for establishing the validity of the study when using a finite mixture approach, and for the interpretation of its results.

1.1 “Label Switching” and Local Maxima

Theoretically, finite mixture models are not strictly identifiable, as they suffer from a “label switching” problem—that is, the distribution remains identical if the labels of the mixture components are switched. In a sense, finite mixture models only help researchers to “group” together subjects that are similar; they cannot further “label” the groups. For this reason, applications in diagnostic testing often assume that higher test results correspond to more severe conditions [4]. Practically speaking, this “label switching” problem of the finite mixture models is not an issue of great importance, because it is usually easy to correctly associate each mixture component with its label. For example, when a finite mixture approach is used in diagnostic medicine to model the diagnostic test results from subjects with or without a certain medical condition, it is usually apparent in the results which group is the diseased group and which is the healthy one.

In addition to the “label switching” problem, another difficulty in proving the identifiability of the finite mixture model is that the likelihood functions of such models usually have multiple local maxima. In other words, one cannot show that the likelihood is a bijection—a common approach to establish model identification.

1.2 Previous Results

Due to the difficulties caused by “label switching” and multiple local maxima, research on the identifiability issue of finite mixture models has mainly focused on local identifiability [5, 6, 7]. It is important to recognize that the commonly used criterion that the degrees of freedom in the data must meet or exceed the number of parameters in the model is not enough to guarantee model identifiability [6, 8]. McHugh [5] proposed sufficient conditions for local identifiability of models with dichotomous observed variables. Goodman [6] extended these conditions to polytomous variables, and other work has focused on finding sufficient conditions to guarantee that the Jacobian matrix of the mixture distribution has full column rank [9, 10]. However, conditions for ensuring identifiability for more general models, such as models that allow for covariate effects or continuous test results have not been established.

Moreover, although local identifiability is necessary for the implementation of the model and for establishing the validity of the asymptotic approximation, it only addresses the problem at an infinitely small neighborhood in the entire parameter space. No conclusion can be made about the same model with a different set of parameters. The identifiability of the model for one researcher cannot be carried on to another researcher who wants to repeat the experiment. Local identifiability looks at model identification at a given point of parameters. In a sense, it is more specific to a given problem in which the true parameters are fixed (even though they are unknown) than to an unrestricted model itself. Additionally, since the true parameters are unknown, current methods evaluate local identifiability using the estimated parameter values. This means that identifiability cannot be assessed until after data collection has occurred. However, in many cases the ability to establish the identifiability before collecting the data is necessary or desirable, as the study may only be meaningful when the model is identifiable.

In the special application of finite mixture models to diagnostic tests, more work on the identifiability issue has been accomplished. For the L -component mixture of independent binomial distributions, $B(K, p_l)$, it has been shown that $K \geq 2L - 1$ is both necessary and sufficient for model identification [11, 12]. Based on this result, with an additional assumption of identically distributed ordinal or continuous tests, several researchers [13, 14] dichotomized the test results and concluded that $K \geq 2L - 1$ is sufficient for the identification of such models. Hall and his colleagues [15, 16] studied the nonparametric identifiability of models of K tests with conditional independence assumptions within each of the M subpopulations, and provided sufficient conditions for model identification of $K \geq (1 + o(1))6L \log L$. Jones et al. [17] discussed identifiability issue for multiple binary but possibly conditional dependent tests. However, conditions for more general models have not yet been established.

1.3 Outlines

In this paper, we provide conditions for both local identifiability and global identifiability of a finite mixture model. We consider general structures of finite mixture models, which allow for 1) ordinal or nominal latent groups to accommodate disease severity and subtypes; 2) continuous, discrete or mix-typed manifest variables which can combine information from continuous biomarkers and discrete test readings; 3) inclusion of covariates in both the structure part (model about the latent

groups) and the measurement part (model about the manifest variables within a latent group) of the model; 4) general link functions when including covariate effects. The paper is presented as follows: Section 2 provides a general description of a finite mixture model that we considered here. Section 3 and 4 establish conditions for local and global identifiability of the model, respectively. For a clear presentation, the results are first provided for models without covariates and then for models with covariates. Section 5 concludes the paper with a discussion.

2 Description of a Finite Mixture model

A finite mixture model is characterized by its component distributions, its number of components and its mixing proportion. Consider a model with L components with label $d = 0, \dots, L-1$ (such as disease severity), and K manifest variables $\vec{T} = (T_1, \dots, T_K)$ (such as diagnostic tests). Assuming conditional independence among manifest variables T_k within a given component, we can write the model as:

$$P(\vec{T}) = \sum_{d=0}^{L-1} P(\vec{T}, D) = \sum_{d=0}^{L-1} [P(D = d) \prod_{k=1}^K P(T_k | D = d)],$$

where $P(\cdot)$ is the marginal distribution function of manifest variable \vec{T} , $P(D = d)$ is the mixture proportion of the d th subpopulation, and $P(T_k | D = d)$ is the conditional probability of T_k in the d th subpopulation, either as a conditional density function $f_{T_k|D}(\cdot)$ for continuous T_k , or as a probability mass function for categorical T_k . For example, when all of the manifest variables are categorical and take on the values $\{1, \dots, J_k\}$, $k = 1, \dots, K$, the model is a latent class model as follows,

$$P(T_1 = t_1, \dots, T_K = t_K) = \sum_{d=0}^{L-1} \{P(D = d) \prod_{k=1}^K \prod_{j=1}^{J_k} [P(T_k = j | D = d)]^{I[t_k=j]}\}, \quad (1)$$

where $I[t_k = j]$ is an indicator, which equals 1 if $t_k = j$ and 0 otherwise.

Model (1) is a finite mixture model without covariates. When covariates are involved, the extended model can be expressed as

$$P(\vec{T} | \vec{X}, \vec{Z}) = \sum_{d=0}^{L-1} P(\vec{T}, D | \vec{X}, \vec{Z}) = \sum_{d=0}^{L-1} [P(D = d | \vec{Z}) \prod_{k=1}^K P(T_k | D = d, \vec{X})],$$

where \vec{Z} is the set of covariates associated with mixture membership and \vec{X} is the set of covariates associated with manifest variables within each group. The two sets of covariates may be mutually exclusive or overlapping, and they can include continuous and categorical variables.

Similarly as before, $P(T_k | D = d, \vec{X})$ can be the conditional density function or probability mass function. In the latter case, the model is,

$$P(T_1 = t_1, \dots, T_K = t_K | \vec{X}, \vec{Z}) = \sum_{d=0}^{L-1} \{P(D = d | \vec{Z}) \prod_{k=1}^K \prod_{j=1}^{J_k} [P(T_k = j | D = d, \vec{X})]^{I[t_k=j]}\}. \quad (2)$$

3 Local Identifiability

Local identifiability is important for the implementation of the model and the validity of its asymptotic approximation. Although, as mentioned earlier, it only address the identifiability of the model at a given point and thus may not be adequate in some situations; we nevertheless discuss local identifiability here.

We argue here that for local identifiability, it is sufficient to only consider models with discrete manifest variables, as is the case in model (1) and model (2). This is because local identifiability essentially considers the mapping between fixed parameter values in the parameter space and their induced data sets in the data space. When some or all of the manifest variables are continuous, their distributions can be modeled empirically on the observed data points, which are discrete. In other words, the same techniques used to establish the local identifiability of a latent class model can also be used to establish the nonparametric local identifiability of a finite mixture model with some or all of its manifest variables being continuous. The parametric assumptions about the conditional distribution of a manifest variable $P(T_k|D = d)$ or $P(T_k = j|D = d, \vec{X})$, if any, can be viewed as additional constraints in the estimation procedure.

3.1 Definition

By definition, a function F is locally identifiable at parameter $\theta_0 \in \Theta$ if there exists some neighborhood U_{θ_0} of θ_0 , such that

$$F(\theta) \neq F(\theta_0) \quad \forall \theta \in U_{\theta_0} \setminus \{\theta_0\}.$$

This suggests that F is a one-to-one map, or locally invertible in U_{θ_0} .

As argued before, without loss of generality, we consider a situation in which all of the manifest variables are categorical. Let $\vec{t}_h = (t_{h1}, \dots, t_{hK})$ be the h th possible in lexicographic order among $(\prod_{k=1}^K J_k) - 1$ distinct response patterns of the manifest variables, excluding a reference pattern. We stacked the probability $P(\vec{T} = \vec{t}_h)$ into a vector p of length $(\prod_{k=1}^K J_k) - 1$. A given model specifies a function F , which determines how p is calculated from parameters θ ,

$$p = F(\theta).$$

The model is locally identifiable at θ_0 if F is invertible in a neighborhood of θ_0 . When the number of parameters is less than $(\prod_{k=1}^K J_k) - 1$, F is potentially invertible, and local invertibility at θ_0 can be evaluated by examining the Jacobian matrix of F at θ_0 , $J(\theta_0) = \frac{\partial F}{\partial \theta} \Big|_{\theta=\theta_0}$. By the weak inversion theorem, if $J(\theta_0)$ has full column rank, F is locally invertible at θ_0 , and thus the model was locally identifiable at θ_0 .

3.2 Models without Covariates

In this section we consider local identifiability for models without covariates, using model (1). The basic idea is to find sufficient conditions that guarantee the local invertibility of the Jacobian matrix

of the model.

Conditions for Local Identifiability

Let $\pi_d = P(D = d)$, $g_{kjd} = P(T_k = j | D = d)$, Ψ_d be a vector of length $(\prod_{k=1}^K J_k) - 1$, $d = 0, \dots, L - 1$, with the h th element

$$\psi_{dh} = P(\vec{T} = \vec{t}_h | D = d) = \prod_{k=1}^K g_{kt_{hk}d}.$$

Further, let $\eta_d = \Psi_d - \Psi_0$, $d = 1, \dots, L - 1$, and Γ_{kjd} be a vector of length $(\prod_{k=1}^K J_k) - 1$, $d = 0, \dots, L - 1$, $k = 1, \dots, K$ and $j = 1, \dots, J_k - 1$, with the h th element

$$\gamma_{kjdh} = \pi_d \psi_{dh} \left[\frac{I(t_{hk} = j)}{g_{kjd}} - \frac{I(t_{hk} = J_k)(g_{kjd} - \sum_{j=1}^{J_k-1} g_{kjd})}{g_{kJ_kd}} \right],$$

where $I(t_{hk} = j)$ is an indicator function that equals 1 if $t_{hk} = j$ and 0 if otherwise. Then, we have the following theorem:

Theorem 1:

The finite mixture model (1) is locally identifiable at parameter $\theta = \{\pi_d, g_{kjd} \mid d = 0, \dots, L - 1; k = 1, \dots, K; j = 1, \dots, J_k\}$ if the following conditions hold.

- (i) $(\prod_{k=1}^K J_k) - 1 \geq L \times \sum_{k=1}^K (J_k - 1) + L - 1$;
- (ii) $P(\vec{T} = \vec{t}_h) = \sum_{d=0}^{L-1} \pi_d \psi_{dh} > 0 \quad \forall h$, and $\pi_d > 0 \quad \forall d$;
- (iii) vectors $\{\eta_d \mid d = 1, \dots, L - 1\}$, $\{\Gamma_{kjd} \mid d = 0, \dots, L - 1; k = 1, \dots, K; j = 1, \dots, J_k\}$ are linearly independent.

Proof:

Condition (i) requires that the degrees of freedom in the data are greater than the number of parameters. Moreover, $(\prod_{k=1}^K J_k) - 1$ is the number of rows of the Jacobian matrix of model (1), and $L \times \sum_{k=1}^K (J_k - 1) + L - 1$ is the number of columns of the Jacobian matrix. When condition (i) is satisfied, the Jacobian matrix can potentially have full column rank. Condition (ii) is included to ensure that the probability of observing every response pattern is positive. It is the third condition in Theorem 1 of McHugh (1956) [5] that pertains to the local identifiability of latent class models with binary manifest variables. Here, we only need to prove that Condition (iii) is equivalent to requiring that the Jacobian matrix of model (1) has full column rank.

Based on model (1), the function between parameter θ and all possible response patterns of the manifest variables \vec{t}_h (excluding a reference pattern) can be expressed as

$$F(\vec{t}_h; \theta) = P_\theta(\vec{T} = \vec{t}_h) = \sum_{d=0}^{L-1} \pi_d \prod_{k=1}^K g_{kt_{hk}d}, \quad h = 1, \dots, H,$$

where $H = (\prod_{k=1}^K J_k) - 1$. Since the component probabilities sum up to 1, $\pi_0 = 1 - \sum_{d=1}^{L-1} \pi_d$. Therefore

$$F(\vec{t}_h; \theta) = \sum_{d=1}^{L-1} \pi_d \prod_{k=1}^K g_{kt_{hk}d} + (1 - \sum_{d=1}^{L-1} \pi_d) \prod_{k=1}^K g_{kt_{hk}0}, \quad h = 1, \dots, H.$$

Taking the derivative of F with respect to free component probability parameters π_d , $d = 1, \dots, L-1$, we have that

$$\frac{\partial F(\vec{t}_h; \theta)}{\partial \pi_d} = \prod_{k=1}^K g_{kt_{hk}d} - \prod_{k=1}^K g_{kt_{hk}0} = \psi_{dh} - \psi_{0h}.$$

Therefore, the first $L-1$ columns of the Jacobian matrix of model (1) are,

$$\frac{\partial F}{\partial \pi_d} = \Psi_d - \Psi_0 = \eta_d, \quad d = 1, \dots, L-1.$$

Meanwhile, since $\sum_{j=1}^{J_k} g_{kjd} = 1$, the function F can be rewritten as follows,

$$\begin{aligned} F(\vec{t}_h; \theta) &= \sum_{d=0}^{L-1} \pi_d \prod_{k=1}^K g_{kt_{hk}d} \\ &= \sum_{d=0}^{L-1} \pi_d \prod_{k=1}^K \left[I(t_{hk} \neq J_k) g_{kt_{hk}d} + I(t_{hk} = J_k) \left(1 - \sum_{j=1}^{J_k-1} g_{kjd} \right) \right], \quad h = 1, \dots, H. \end{aligned}$$

Taking the derivative of F with respect to the free parameter g_{kjd} , $d = 0, \dots, L-1$, $k = 1, \dots, K$ and $j = 1, \dots, J_k-1$, we have

$$\begin{aligned} \frac{\partial F(\vec{t}_h; \theta)}{\partial g_{kjd}} &= \pi_d \psi_{hd} \left[\frac{I(t_{hk} = j)}{g_{kjd}} - \frac{I(t_{hk} = J_k) (1 - \sum_{j=1}^{J_k-1} g_{kjd} + g_{kjd} - 1)}{g_{kJ_kd}} \right] \\ &= \pi_d \psi_{hd} \left[\frac{I(t_{hk} = j)}{g_{kjd}} - \frac{I(t_{hk} = J_k) (g_{kjd} - \sum_{j=1}^{J_k-1} g_{kjd})}{g_{kJ_kd}} \right]. \end{aligned}$$

Thus, the last $L \times \sum_{k=1}^K (J_k - 1)$ columns of the Jacobian matrix of model (1) are,

$$\frac{\partial F}{\partial g_{kjd}} = \Gamma_{kjd}, \quad d = 0, \dots, L-1, \quad k = 1, \dots, K, \quad j = 1, \dots, J_k-1.$$

Therefore, the Jacobian matrix of model (1) is a $(\prod_{k=1}^K J_k) - 1$ by $L \times \sum_{k=1}^K (J_k - 1) + L - 1$ matrix $J(\theta)$ with columns $(\eta_1, \dots, \eta_{L-1}, \Gamma_{111}, \dots, \Gamma_{KJ_{K-1}L-1})$. As a result, condition (iii) is equivalent to requiring that the Jacobian matrix of model (1) has full column rank, which in turn guarantees that the finite mixture model (1) is locally identifiable at parameter $\theta = \{\pi_d, g_{kjd} \mid d = 0, \dots, L-1; k = 1, \dots, K; j = 1, \dots, J_k\}$. \square

More on Condition (iii)

Conditions (i) and (ii) in Theorem 1 are relatively straight forward to examine, condition (iii) requires more attention. Some efforts have been made to provide equivalent but simplified conditions to condition (iii) with additional model specifications, such as constraining all of the manifest variables to be binary.

It is worth noting that having Ψ_0, \dots, Ψ_D being linearly independent is not sufficient for the Jacobian matrix of model (1) to have full rank, in contrary to some previous believes. For example, consider the situation where all but one of the manifest variables are totally non-informative about the latent subgroup membership.

Without loss of generality, we assumed that the last manifest variable is the only informative one. Let \mathbf{g}_{Kd} be the vector of probability mass of this last manifest variable in subgroup d ,

$$\mathbf{g}_{Kd} = (P(T_K = 1|D = d), \dots, P(T_K = J_K|D = d))^T = (g_{K1d}, \dots, g_{KJ_Kd})^T, \quad d = 0, \dots, L-1.$$

Suppose that $J_K \geq L$ and vectors $\mathbf{g}_{K0}, \dots, \mathbf{g}_{K(L-1)}$ are linearly independent. Additionally, suppose that all other manifest variables are uniformly distributed among the subgroups, and are thus non-informative about subgroup membership:

$$P(T_k = j|d = 0) = \dots = P(T_k = j|d = D), \quad \forall k = 1, \dots, K-1; j = 1, \dots, J_k,$$

or equivalently, $g_{kj0} = \dots = g_{kj(L-1)} \equiv \bar{g}_{kj} \neq 0, \quad \forall k = 1, \dots, K-1; j = 1, \dots, J_k.$

Then the first J_K elements of Ψ_d is

$$\begin{pmatrix} P(\vec{T} = (1, 1, \dots, 1) | D = d) \\ P(\vec{T} = (1, 1, \dots, 2) | D = d) \\ \vdots \\ P(\vec{T} = (1, 1, \dots, J_K) | D = d) \end{pmatrix} = \begin{pmatrix} (\prod_{k=1}^{K-1} \bar{g}_{k1}) g_{K1d} \\ (\prod_{k=1}^{K-1} \bar{g}_{k1}) g_{K2d} \\ \vdots \\ (\prod_{k=1}^{K-1} \bar{g}_{k1}) g_{KJ_Kd} \end{pmatrix} = \left(\prod_{k=1}^{K-1} \bar{g}_{k1} \right) \mathbf{g}_{Kd}.$$

Because $\prod_{k=1}^{K-1} \bar{g}_{k1} \neq 0$ and $\mathbf{g}_{K0}, \dots, \mathbf{g}_{K(L-1)}$ are linearly independent, we have that

$$\text{vectors } \left(\prod_{k=1}^{K-1} \bar{g}_{k1} \right) \mathbf{g}_{K0}, \dots, \left(\prod_{k=1}^{K-1} \bar{g}_{k1} \right) \mathbf{g}_{K(L-1)} \text{ are linearly independent.} \quad (3)$$

With more elements appended below each vector in 3, the extension groups $\Psi_0, \dots, \Psi_{(L-1)}$ are linearly independent.

Consequently, if having $\Psi_0, \dots, \Psi_{(L-1)}$ being linearly independent is sufficient for the Jacobian matrix of model (1) to have full rank, the above example suggests that having only one “good” manifest variable with several non-informative manifest variables is sufficient to achieve local identifiability. This is certainly not the case. Otherwise, in diagnostic testing settings with binary disease groups, for example, this would mean that one informative binary test with two random guesses is sufficient for model identifiability. However, the only informative test needs to provide estimates for disease prevalence as well as for its own sensitivity and specificity. There are three parameters but only two degrees of freedom for the results of the first binary test. Therefore, the model is not identifiable.

In fact, let π be the disease prevalence and Se_k and Sp_k be the sensitivity and specificity of the k th binary test, respectively. The Jacobian for model (1) with a binary disease group and 3 binary tests is a 7 by 7 matrix with determinant:

$$|J| = \pi^3(\pi - 1)^3(Se_1 + Sp_1 - 1)^2(Se_2 + Sp_2 - 1)^2(Se_3 + Sp_3 - 1)^2. \quad (4)$$

When a test is non-informative, we have that

$$Se = P(T + |D+) = P(T + |D-) = 1 - Sp.$$

Consequently, it is easy to see that when one or more of the tests are non-informative, Jacobian 4 is singular. However, by the same construction described above, we still have that Ψ_0 and Ψ_1 are

linearly independent as long as one test is informative.

An Algebraic Geometry Point of View

It is interesting to revisit this problem from an algebraic geometry point of view. Model (1) can be expressed as follows,

$$p = F(\theta) = \sum_{d=0}^{L-1} \pi_d \Psi_d. \quad (5)$$

Therefore, p is a linear combination of vectors $\Psi_0, \dots, \Psi_{L-1}$. It may be natural to consider that if $\Psi_0, \dots, \Psi_{L-1}$ are linearly independent, it becomes a basis of its span over the real field. Thus, the decomposition is unique and the model is identifiable. In fact, this is the essential idea of Yakowitz and Spragins [18], when they studied the identifiability of finite mixture models. However, they required that the component distributions belonged to a pre-specified family \mathcal{F} , and that all elements in \mathcal{F} were linearly independent over the field of real numbers. For example, as they showed, the exponential family and the Gaussian family are such families.

However, their results do not apply when considering the identifiability of model (1), where all manifest variables are categorical and are not constrained as reasoned below. In this case, component distributions belong to a multinomial family. For any m variate multinomial family, the probability mass function can be expressed as a factor with $m - 1$ elements. Therefore, at most m such vectors will be linearly dependent. Meanwhile, there is an infinite number of such m element probability vectors in the m variate multinomial family, so the multinomial family does not satisfy the conditions set forth in [18].

Another way to understand this problem is by directly examining equation 5. In this model, $\Psi_0, \dots, \Psi_{L-1}$ may only expand a subspace in \mathcal{F} , and therefore is not its basis, and even if it were, this is not the only basis of the vector space of $F(\theta)$. In fact, any simultaneous rotation of $\Psi_0, \dots, \Psi_{L-1}$ along vector p can lead to a different decomposition of p while maintaining the necessary property for probability mass that the sum of all of the elements is equal to 1.

For example, one can consider the problem in a 3-dimensional space for simplicity. Define the length of a vector $\vec{a} = (a_1, a_2, a_3)$ as $|\vec{a}| = a_1 + a_2 + a_3$. Then, as vectors of probability mass, Ψ_1, Ψ_2, Ψ_3 are all unit vectors with nonnegative elements, suppose that they are linearly independent and that,

$$p = F(\theta) = \pi_1 \Psi_1 + \pi_2 \Psi_2 + \pi_3 \Psi_3.$$

A rotation of angle ϕ along vector $p = (p_1, p_2, p_3)$ has a rotation matrix as follows:

$$R(\phi) = \begin{pmatrix} [\cos \phi + p_1(1 - \cos \phi)]^2 & [\sqrt{p_1 p_2}(1 - \cos \phi) - \sqrt{p_3} \sin \phi]^2 & [\sqrt{p_1 p_3}(1 - \cos \phi) + \sqrt{p_2} \sin \phi]^2 \\ [\sqrt{p_1 p_2}(1 - \cos \phi) + \sqrt{p_3} \sin \phi]^2 & [\cos \phi + p_2(1 - \cos \phi)]^2 & [\sqrt{p_2 p_3}(1 - \cos \phi) - \sqrt{p_1} \sin \phi]^2 \\ [\sqrt{p_1 p_3}(1 - \cos \phi) - \sqrt{p_2} \sin \phi]^2 & [\sqrt{p_2 p_3}(1 - \cos \phi) + \sqrt{p_1} \sin \phi]^2 & [\cos \phi + p_3(1 - \cos \phi)]^2 \end{pmatrix}.$$

Note that R is different from the rotation matrix of angle ϕ along vector $a = (a_1, a_2, a_3)$ in Euclidean space

$$\begin{pmatrix} \cos \phi + a_1^2(1 - \cos \phi) & a_1 a_2(1 - \cos \phi) - a_3 \sin \phi & a_1 a_3(1 - \cos \phi) + a_2 \sin \phi \\ a_1 a_2(1 - \cos \phi) + a_3 \sin \phi & \cos \phi + a_2^2(1 - \cos \phi) & a_2 a_3(1 - \cos \phi) - a_1 \sin \phi \\ a_1 a_3(1 - \cos \phi) - a_2 \sin \phi & a_2 a_3(1 - \cos \phi) + a_1 \sin \phi & \cos \phi + a_3^2(1 - \cos \phi) \end{pmatrix}$$

because of the different definition of length .

Then, for any angle $\phi \in [0, 2\pi)$, we have that,

$$p = F(\theta) = \pi_1 \Psi_1^*(\phi) + \pi_2 \Psi_2^*(\phi) + \pi_3 \Psi_3^*(\phi),$$

where $\Psi_k^*(\phi) = R(\phi)\Psi_k$, $k = 1, 2, 3$, remain unit vectors with nonnegative elements, and thus are vectors of probability mass.

This example further illustrates that having $\Psi_0, \dots, \Psi_{L-1}$ being linearly independent is not in itself a sufficient condition to guarantee the identifiability of finite mixture model (1).

3.3 Models with Covariates

Now we will consider local identifiability for latent class models with covariates, model (2), as follows:

$$P(T_1 = t_1, \dots, T_K = t_K | \vec{X}, \vec{Z}) = \sum_{d=0}^{L-1} \{P(D = d | \vec{Z}) \prod_{k=1}^K \prod_{j=1}^{J_k} [P(T_k = j | D = d, \vec{X})]^{I[t_k=j]}\}.$$

we further assume that the covariate effects are linear under certain pre-specified transformations, such as under a logit transformation. For example, one may use polytomous regression models for both the latent group membership and the manifest variables, as both are categorical,

$$\begin{aligned} \pi_d(\vec{z}^T \alpha_d) &= P(D_i = d | \vec{Z}_i = \vec{z}), \quad \log \frac{\pi_d(\vec{z}_i^T \alpha_d)}{\pi_0(\vec{z}_i^T \alpha_0)} = \vec{z}_i^T \alpha_d \quad d = 1, \dots, L-1 \\ g_{kjd}(\vec{x}^T \beta_{kjd}) &= P(T_{ik} = j | D_i = d, \vec{X}_i = \vec{x}), \quad \log \frac{g_{kjd}(\vec{x}_i^T \beta_{kjd})}{g_{kJkd}(\vec{x}_i^T \beta_{kJkd})} = \vec{x}_i^T \beta_{kd} \\ &d = 0, \dots, L-1; \quad k = 1, \dots, K; \quad j = 1, \dots, J_k - 1. \end{aligned}$$

Under the linear covariate effect assumption, model (2) can be rewritten in the following form:

$$P(\vec{T} = \vec{t} | \vec{X}, \vec{Z}) = \sum_{d=0}^{L-1} \pi_d(\vec{Z}^T \alpha_d) \prod_{k=1}^K \prod_{j=1}^{J_k} g_{kjd}(\vec{X}^T \beta_{kjd})^{I[t_k=j]}, \quad (6)$$

$$\text{where } \pi_d(\vec{Z}^T \alpha_d) = P(D = d | \vec{Z}), \quad g_{kjd}(\vec{X}^T \beta_{kjd}) = P(T_{ik} = j | D_i = d, \vec{X}).$$

We can see that the function above has a similar form to the model without covariates, except that π_d and g_{kjd} are predefined functions while the parameters to be estimated are α_d and β_{kdj} , $d = 0, \dots, L-1$; $k = 1, \dots, K$; $j = 1, \dots, J_k$. In fact, a similar procedure can be used when examining the local invertibility of this function.

Let N by q matrix \mathbf{X} and N by p matrix \mathbf{Z} be the design matrix in model (6), with the i th row \mathbf{X}_i and \mathbf{Z}_i , respectively. Let $\eta_d(\mathbf{X}_i) = \Psi_d(\mathbf{X}_i) - \Psi_0(\mathbf{X}_i)$, $d = 1, \dots, L-1$, where $\Psi_d(\mathbf{X}_i)$ is a vector

of length $(\prod_{k=1}^K J_k) - 1$, with the h th element

$$\psi_{dh}(\mathbf{X}_i) = P(\vec{T} = \vec{t}_h | D = d, \mathbf{X}_i) = \prod_{k=1}^K g_{kt_{hk}d}(\mathbf{X}_i \beta_{kt_{hk}d}).$$

Additionally, let $\Gamma_{kjd}(\mathbf{X}_i, \mathbf{Z}_i)$ be a vector of length $(\prod_{k=1}^K J_k) - 1$, $d = 0, \dots, L - 1$, $k = 1, \dots, K$ and $j = 1, \dots, J_k - 1$, with the h th element

$$\gamma_{kjd}(\mathbf{X}_i, \mathbf{Z}_i) = \pi_d(\mathbf{Z}_i \alpha_d) \psi_{dh}(\mathbf{X}_i) \left[\frac{I(t_{hk} = j)}{g_{kjd}(\mathbf{X}_i \beta_{kjd})} - \frac{I(t_{hk} = J_k)(g_{kjd}(\mathbf{X}_i \beta_{kjd}) - \sum_{j=1}^{J_k-1} g_{kjd}(\mathbf{X}_i \beta_{kjd}))}{g_{kJkd}(\mathbf{X}_i \beta_{kJkd})} \right]$$

where $I(t_{hk} = j)$ is an indicator function, which equals 1 if $t_{hk} = j$ and 0 otherwise.

Further, define $N[(\prod_{k=1}^K J_k) - 1]$ by p matrix A_d , $d = 1, \dots, L - 1$, and $N \times ((\prod_{k=1}^K J_k) - 1)$ by q matrix B_{kjd} , $d = 0, \dots, L - 1$, $k = 1, \dots, K$, $j = 1, \dots, J_k$, as follows,

$$A_d = \begin{pmatrix} \eta_d(\mathbf{X}_1) \pi'_d(\mathbf{Z}_1 \alpha_d) \mathbf{Z}_1 \\ \vdots \\ \eta_d(\mathbf{X}_N) \pi'_d(\mathbf{Z}_N \alpha_d) \mathbf{Z}_N \end{pmatrix}, \quad B_{kjd} = \begin{pmatrix} \Gamma_{kjd}(\mathbf{X}_1, \mathbf{Z}_1) g'_{kjd}(\mathbf{X}_1 \beta_{kjd}) \mathbf{X}_1 \\ \vdots \\ \Gamma_{kjd}(\mathbf{X}_N, \mathbf{Z}_N) g'_{kjd}(\mathbf{X}_N \beta_{kjd}) \mathbf{X}_N \end{pmatrix},$$

where $\pi'_d(\cdot)$ and $g'_{kjd}(\cdot)$ are the derivative of $\pi_d(\cdot)$ and $g_{kjd}(\cdot)$.

Theorem 2:

Finite mixture model (6) is locally identifiable at parameter $\theta = \{\alpha_d, \beta_{kjd} \mid d = 0, \dots, L - 1; k = 1, \dots, K; j = 1, \dots, J_k\}$ if the following conditions hold.

- (i) $N[(\prod_{k=1}^K J_k) - 1] \geq q \times L[\sum_{k=1}^K (J_k - 1)] + p(L - 1)$;
- (ii) $P(\vec{T} = \vec{t}_h | \mathbf{X}_i, \mathbf{Z}_i) = \sum_{d=0}^{L-1} \pi_d \psi_{dh}(\mathbf{X}_i) > 0 \quad \forall h$, and $\pi_d > 0 \quad \forall d$;
- (iii) column vectors in matrices $\{A_d \mid d = 1, \dots, L - 1\}$, $\{B_{kjd} \mid d = 0, \dots, L - 1; k = 1, \dots, K; j = 1, \dots, J_k\}$ all together are linearly independent.

To prove Theorem 2, we will first prove the following lemma.

Lemma 1:

Suppose that A is a matrix, and that A^* is a matrix obtained by stacking some of the rows in A vertically l times,

$$A^* = \left(\begin{matrix} A_1 \\ A_2^* \end{matrix} \right), \quad A_2^* = \left(\begin{matrix} A_2 \\ \vdots \\ A_2 \end{matrix} \right) \Bigg\} l \text{ times}$$

where A_1, A_2 are sub-matrices of A , such that $A = (A_1^T, A_2^T)^T$, l is a given positive integer. Then the column vectors of A being linearly independent is equivalent to the column vectors of A^* , linearly independent.

Proof of Lemma 1:

Denote the set of column vectors of A by A_{col} , and the set of column vectors of A^* by A_{col}^* . Because each of the vectors in A_{col}^* append more elements to each of the vectors in A_{col} , A_{col}^* is an extension group of A_{col} . Therefore, if the vectors in A_{col} are linearly independent, we have that the vectors in A_{col}^* are linearly independent.

Now we prove that the reverse is also true by contradiction. Suppose that the vectors in A_{col}^* are linearly independent, but that the vectors in A_{col} are linearly dependent. Then there exists a vector α , such that $A\alpha = 0$. Thus, $A_1\alpha = 0$ and $A_2\alpha = 0$. It follows that,

$$A^*\alpha = \begin{pmatrix} A_1 \\ A_2 \\ \vdots \\ A_2 \end{pmatrix} \alpha = \begin{pmatrix} A_1\alpha \\ A_2\alpha \\ \vdots \\ A_2\alpha \end{pmatrix} = 0.$$

However, this contradicts our supposition that the vectors in A_{col} are linearly independent. \square

Proof of Theorem 2:

Conditions (i) and (ii) are similar as before. We only need to prove condition (iii) here.

Based on model (6), the function between parameter θ and all possible response patterns of the manifest variables \vec{t}_h (excluding a reference pattern) is,

$$\begin{aligned} p &= F(\vec{t}_h; \theta | \vec{X}, \vec{Z}) = P_\theta(\vec{T} = \vec{t} | \vec{X}, \vec{Z}) \\ &= \sum_{d=0}^{L-1} \pi_d(\vec{Z}^T \alpha_d) \prod_{k=1}^K g_{kjd}(\vec{X}^T \beta_{kt_{hk}d}), \end{aligned}$$

where $h = 1, \dots, H$ with $H = (\prod_{k=1}^K J_k) - 1$.

First, we consider the situation where all of the covariate vectors $(\vec{X}_i^T, \vec{Z}_i^T)^T$ are different for $i = 1, \dots, N$. In this case, F defines a mapping between all H possible response patterns of \vec{t} for each of the covariate vectors $(\vec{X}_i^T, \vec{Z}_i^T)^T$. Therefore, the Jacobian matrix has $N \times H$ rows. It can be divided into N blocks, each containing H contiguous rows. Then the i th block is the derivative of F with respect to each of the parameters when the covariate vectors are $(\vec{X}_i^T, \vec{Z}_i^T)^T$. In other words, it is the Jacobian matrix of F when there is only a single observation with covariates $(\vec{X}_i^T, \vec{Z}_i^T)^T$. Since each block can compute each block separately, we only compute the i th block with covariates $(\vec{X}_i^T, \vec{Z}_i^T)^T$.

Take the derivative of F with respect to free parameters α_d , $d = 1, \dots, L - 1$, and by the chain rule, we have

$$\begin{aligned} \frac{\partial F(\vec{t}_h; \theta | \mathbf{X}_i, \mathbf{Z}_i)}{\partial \alpha_d} &= \frac{\partial F(\vec{t}_h; \theta | \mathbf{X}_i, \mathbf{Z}_i)}{\partial \pi_d(\mathbf{Z}_i \alpha_d)} \cdot \frac{\partial \pi_d(\mathbf{Z}_i \alpha_d)}{\partial (\mathbf{Z}_i \alpha_d)} \cdot \frac{\partial (\mathbf{Z}_i \alpha_d)}{\partial \alpha_d} \\ &= \left[\prod_{k=1}^K g_{kt_{hk}d}(\mathbf{X}_i \beta_{kt_{hk}d}) - \prod_{k=1}^K g_{kt_{hk}0}(\mathbf{X}_i \beta_{kt_{hk}0}) \right] \pi'_d(\mathbf{Z}_i \alpha_d) \mathbf{Z}_i \\ &= [\psi_{dh}(\mathbf{X}_i) - \psi_{0h}(\mathbf{X}_i)] \pi'_d(\mathbf{Z}_i \alpha_d) \mathbf{Z}_i. \end{aligned}$$

Therefore, the first $p \times (L - 1)$ columns of the i th block of the Jacobian matrix of model (6) is,

$$\frac{\partial F}{\partial \alpha_d} = [\Psi_d(\mathbf{X}_i) - \Psi_0(\mathbf{X}_i)] \mathbf{Z}_i = \eta_d(\mathbf{X}_i) \pi'_d(\mathbf{Z}_i \alpha_d) \mathbf{Z}_i, \quad d = 1, \dots, L - 1.$$

Then, taking the derivative of F with respect to free parameter β_{kjd} , $d = 0, \dots, L-1$, $k = 1, \dots, K$ and $j = 1, \dots, J_k - 1$, we have

$$\begin{aligned} \frac{\partial F(\vec{t}_h; \theta | \mathbf{X}_i, \mathbf{Z}_i)}{\partial \beta_{kjd}} &= \frac{\partial F(\vec{t}_h; \theta | \mathbf{X}_i, \mathbf{Z}_i)}{\partial g_{kjd}(\mathbf{X}_i^T \beta_{kjd})} \cdot \frac{\partial g_{kjd}(\mathbf{X}_i^T \beta_{kjd})}{\partial (\mathbf{X}_i^T \beta_{kjd})} \cdot \frac{\partial (\mathbf{X}_i^T \beta_{kjd})}{\partial \beta_{kjd}} \\ &= \left\{ \pi_d(\mathbf{Z}_i \alpha_d) \psi_{dh}(\mathbf{X}_i) \left[\frac{I(t_{hk} = j)}{g_{kjd}(\mathbf{X}_i^T \beta_{kjd})} - \frac{I(t_{hk} = J_k)(g_{kjd}(\mathbf{X}_i^T \beta_{kjd}) - \sum_{j=1}^{J_k-1} g_{kjd}(\mathbf{X}_i^T \beta_{kjd}))}{g_{kJkd}(\mathbf{X}_i^T \beta_{kJkd})} \right] \right\} g'_{kjd}(\mathbf{X}_i^T \beta_{kjd}) \mathbf{X}_i \\ &= \gamma_{kjd}(\mathbf{X}_i, \mathbf{Z}_i) g'_{kjd}(\mathbf{X}_i^T \beta_{kjd}) \mathbf{X}_i \end{aligned}$$

Thus, the last $q \times L[\sum_{k=1}^K (J_k - 1)]$ columns of the i th block of the Jacobian matrix of model (6) are,

$$\frac{\partial F}{\partial \beta_{kjd}} = \Gamma_{kjd}(\mathbf{X}_i, \mathbf{Z}_i) g'_{kjd}(\mathbf{X}_i^T \beta_{kjd}) \mathbf{X}_i, \quad d = 0, \dots, L-1, \quad k = 1, \dots, K, \quad j = 1, \dots, J_k - 1.$$

Therefore, the Jacobian matrix of model (6) is a $N \times [(\prod_{k=1}^K J_k) - 1]$ by $q \times L[\sum_{k=1}^K (J_k - 1)] + p(L-1)$ matrix

$$J^*(\theta) = (A_1, \dots, A_{L-1}, B_{111}, \dots, B_{KJ_{K-1}L-1}).$$

As a result, condition (iii) is equivalent to requiring that the Jacobian matrix of model (6) has full column rank, which in turn guarantees that the finite mixture model (6) is locally identifiable at parameter $\theta = \{\alpha_d, \beta_{kjd} \mid d = 0, \dots, L-1; k = 1, \dots, K; j = 1, \dots, J_k\}$.

Now suppose that some of the covariate vectors $(\vec{X}_i^T, \vec{Z}_i^T)^T$ are the same. Then the Jacobian matrix of model (6) is a sub-matrix of $J(\theta)$, obtained by excluding the repeated blocks. By Lemma 1, column vectors linearly independent are equivalent for these two matrices. \square

Note that, in the above proof, the number of rows in the Jacobian matrix $J^*(\theta)$ for a finite mixture model with covariates is N -fold of that in the Jacobian matrix $J(\theta)$ for a finite mixture model without covariates, where N is the total sample size. If $J^*(\theta)$ is divided vertically into N blocks with equal sizes, each block is essentially $J(\theta)$ for subjects with the same covariate $\mathbf{X}_i, \mathbf{Z}_i$ and then multiplied by the corresponding design matrix. As a result, if one of these N blocks has full column rank, the longer matrix $J^*(\theta)$ will have full column rank. Moreover, even if $J(\theta)$ does not have full column rank for any of the covariate patterns, when the design matrices \mathbf{X}_i and \mathbf{Z}_i have full column rank, they may help restore full column rank of $J^*(\theta)$. This result is very interesting since it suggests that finite mixture models with covariates may be easier to identify than models without covariates. The difficulty with including covariates in finite mixture models is that it increases the number of parameters to be estimated, especially when the covariates are categorical, and especially when it is of interest to consider their interactions with each manifest variable and each latent group. Because of this, researchers adopted various constraints to ensure model identification, such as assuming that either the latent structure model or the measurement model is indexed with covariates, or assuming that covariate effects are constant across latent groups to remove some interaction terms, etc. In contrast to previous beliefs, these results suggest that for models considered here, a model with covariates is more likely to be identifiable than a model without covariates. This result is not surprising when considering that the additional degrees of freedom in the data introduced by covariates are much higher than the increase in the number of parameters – although having sufficient degrees of freedom does not guarantee model identification, it is a necessary condition. As a simple example, consider that a model with two independent binary tests for binary disease

status is not identifiable, but Hui and Walter [19] showed that with two populations, this model can be identifiable. The population here can be viewed as a covariate that they included in their latent structure model.

Specially, we have the following corollary.

Corollary 1:

Finite mixture model (6) is locally identifiable at parameter $\theta = \{\alpha_d, \beta_{kjd} \mid d = 0, \dots, L-1; k = 1, \dots, K; j = 1, \dots, J_k\}$ if the following conditions hold.

- (i) $N[(\prod_{k=1}^K J_k) - 1] \geq q \times L[\sum_{k=1}^K (J_k - 1)] + p(L-1)$;
- (ii) $P(\vec{T} = \vec{t}_h | \mathbf{X}_i, \mathbf{Z}_i) = \sum_{d=0}^{L-1} \pi_d \psi_{dh}(\mathbf{X}_i) > 0, \quad \forall h$;
- (iii) column vectors in matrices $\{\eta_d(\mathbf{X}_0) | d = 1, \dots, L-1\}$, $\{\Gamma_{kjd}(\mathbf{X}_0, \mathbf{Z}_0) | d = 0, \dots, L-1; k = 1, \dots, K; j = 1, \dots, J_k\}$ all together are linearly independent for some $\mathbf{X}_0 \in \{\mathbf{X}_1, \dots, \mathbf{X}_N\}$ and $\mathbf{Z}_0 \in \{\mathbf{Z}_1, \dots, \mathbf{Z}_N\}$; and
- (iv) design matrix \mathbf{X} and \mathbf{Z} both have full rank.

Proof:

We only need to show that conditions (iii) and (iv) in Corollary 1 \Rightarrow condition (iii) in Theorem 2. Let

$$\Theta = \begin{pmatrix} \eta_1(\mathbf{X}_1) & \dots & \eta_{L-1}(\mathbf{X}_1) & \Gamma_{111}(\mathbf{X}_1, \mathbf{Z}_1) & \dots & \Gamma_{KJ_K-1L-1}(\mathbf{X}_1, \mathbf{Z}_1) \\ \vdots & \dots & \vdots & \vdots & \dots & \vdots \\ \eta_1(\mathbf{X}_N) & \dots & \eta_{L-1}(\mathbf{X}_N) & \Gamma_{111}(\mathbf{X}_N, \mathbf{Z}_N) & \dots & \Gamma_{KJ_K-1L-1}(\mathbf{X}_N, \mathbf{Z}_N) \end{pmatrix},$$

and \mathbb{X} be a block diagonal matrix,

$$\mathbb{X} = \text{diag}\left\{ \underbrace{\mathbf{Z}, \dots, \mathbf{Z}}_{L-1 \text{ times}}, \underbrace{\mathbf{X}, \dots, \mathbf{Z}}_{L[\sum_{k=1}^K (J_k-1)] \text{ times}} \right\}.$$

Then the Jacobian matrix of model (6) can be rewritten as follows,

$$J(\theta) = (A_1, \dots, A_{L-1}, B_{111}, \dots, B_{KJ_K-1L-1}) = \Theta \mathbb{X}.$$

Condition (iii) guarantees that Θ has full column rank, since the column vectors in Θ are an extension group of the vectors in condition (iii). Meanwhile, condition (iv) guarantees that \mathbb{X} has full rank. Therefore, $J(\theta) = \Theta \mathbb{X}$ has full column rank. \square

4 Global Identifiability

Global identifiability considers the identifiability of a model in its entire parameter space. It does not just focus on a single parameter value that may result from a particular sample, but instead considers all possible parameter choices for a given model structure, and is thus more fundamental than local identifiability. However, proving the global identifiability of a function is generally very hard; with the additional difficulty introduced by “label switching” in a finite mixture model, this problem has not been well addressed in the literature.

4.1 Definition

The global identifiability of a model is defined equivalently as the global invertibility of the model induced function $p = F(\theta)$. In the classic definition, a function F is globally invertible for all parameter $\theta \in \Theta$, if

$$F(\theta) = F(\theta^*) \Rightarrow \theta = \theta^*$$

or equivalently, $F(\theta) \neq F(\theta^*) \quad \forall \theta \neq \theta^*, \theta \in \Theta, \theta^* \in \Theta.$

This suggests that F is a one-to-one map on its domain Θ .

Due to the “label switching” problem mentioned in the introduction, finite mixture models are not strictly identifiable. However, since group labels are usually not difficult to determine in practice, it is still worthwhile to consider whether the model induced function $p = F(\theta)$ can uniquely determine a parameter value, up to permutations of group labels.

To express this idea clearly, we first define an equivalent class and an equivalent relationship. Let $\theta = \{\theta_1, \dots, \theta_{L-1}\}$, where θ_d , $d = 1, \dots, L-1$, contain the parameters related to the d th group. We define an equivalent class of θ , denoted by $[\theta]$, as

$$[\theta] = \{\theta_{\sigma(1)}, \dots, \theta_{\sigma(L-1)} \mid \sigma(1), \dots, \sigma(L-1) \text{ is a permutation of } 1, \dots, L-1\}.$$

Moreover, define an equivalent relationship on Θ , denoted by \sim , as

$$\theta \sim \theta^* \text{ if } \theta \in [\theta^*], \text{ or equivalently } [\theta] = [\theta^*].$$

The global identifiability of a finite mixture model can then be defined as follows. A finite mixture model is globally identifiable if its induced function $p = F(\theta)$ satisfies that,

$$F(\theta) = F(\theta^*) \Rightarrow \theta \sim \theta^*$$

or equivalently, $F(\theta) \neq F(\theta^*) \quad \forall \theta \notin [\theta^*], \theta \in \Theta, \theta^* \in \Theta.$

In other words, if a finite mixture model is globally identifiable, its parameter θ can be uniquely determined by the model induced function $p = F(\theta)$, regardless the value of θ , which is in contrast to local identifiability. Clearly, if a model is globally identifiable on Θ , then it is locally identifiable at each $\theta \in \Theta$. Consequently, the Jacobian matrix $F(\theta)$ has full column rank for all $\theta \in \Theta$. However, the reverse is not true. This is because having a Jacobian of full column rank in a region does not guarantee F is invertible in that region. As an example, we can consider the following function on the unit circle $D = \{(x, y) \mid 0 < x^2 + y^2 < 1\}$,

$$F(x, y) = \begin{cases} x^2 - y^2 = u \\ 2xy = v \end{cases}$$

The Jacobian is

$$\frac{\partial(u, v)}{\partial(x, y)} = \begin{vmatrix} 2x & -2y \\ 2y & 2x \end{vmatrix} = 4(x^2 + y^2) > 0.$$

However, $F(0, \frac{1}{2}) = F(0, -\frac{1}{2})$, so F is not globally invertible. Therefore, examining the Jacobian in the entire parameter space does not lead to global identifiability of the model. Consequently, methods in the previous section cannot be extended to show global identifiability of a finite mixture model.

Because of the difficulty in establishing global identifiability, an alternative concept that some researchers have considered, is generic identifiability, which requires that the set of parameter values on which the model is not identifiable has measure zero. More formally, a model is called generically identifiable if its induced function $p = F(\theta)$ satisfies that,

$$\theta \in \Theta \setminus U, \theta^* \in \Theta \setminus U, F(\theta) = F(\theta^*) \Rightarrow \theta = \theta^*$$

or equivalently, $F(\theta) \neq F(\theta^*) \quad \forall \theta \neq \theta^*, \theta \in \Theta \setminus U, \theta^* \in \Theta \setminus U,$

where U is a set of measure zero.

Due to the “label switching” problem, the definition for generic identifiability of a finite mixture model is in fact as follows: a finite mixture model is called generically identifiable if its induced function $p = F(\theta)$ satisfies that,

$$\theta \in \Theta \setminus U, \theta^* \in \Theta \setminus U, F(\theta) = F(\theta^*) \Rightarrow \theta \sim \theta^*$$

or equivalently, $F(\theta) \neq F(\theta^*) \quad \forall \theta \notin [\theta^*], \theta \in \Theta \setminus U, \theta^* \in \Theta \setminus U,$

where U is a set of measure zero.

This concept is useful because, with generic identifiability, one has probability one to reach an identifiable model. Thus, the parameter values can be uniquely determined up to label switching. However, this conclusion is weaker than the global identifiability considered in our work.

4.2 Models without Covariates

In this section we consider global identifiability for a finite mixture model without covariates. An important result that our proof based on is the uniqueness of trilinear decomposition given by Kruskal [20]. We summarize Kruskal’s work on this subject below. We then proceed with our proof, which follows the idea of Allman, et al. [21], except that we will focus on global identifiability instead of generic identifiability of mixture models.

Kruskal’s Result

Kruskal’s result originated from the decomposition of a three-way contingency table. To summarize the result, we first introduce some notation and definitions.

Let M_k be a L by J_k matrix, $k = 1, 2, 3$, with the d th row $\mathbf{m}_d^k = (m_{d1}^k, \dots, m_{dJ_k}^k)$. Let $[M_1, M_2, M_3]$ denote a three-dimensional array with the (u, v, w) element

$$[M_1, M_2, M_3](u, v, w) = \sum_{d=0}^{L-1} m_{du}^1 m_{dv}^2 m_{dw}^3.$$

Additionally, define the Kruskal rank of a matrix M , denoted by $\text{rank}_K M$, as the largest integer I such that every set of I rows of M are linearly independent. Consequently, $\text{rank}_K M$ is less than or equal to the row rank of M , with equality if and only if M has full row rank. Kruskal showed the following result.

Lemma 2

If $\text{rank}_K M_1 + \text{rank}_K M_2 + \text{rank}_K M_3 \geq 2L + 2$, then array $[M_1, M_2, M_3]$ uniquely determines M_k , up to simultaneous permutation and re-scaling of the rows.

we first consider the identifiability of model (1), where all the manifest variables are discrete. If $K = 3$, the following theorem holds.

Theorem 3

Let M_k be a L by J_k matrix, with the $(d + 1, j)$ element

$$M_k(d + 1, j) = P(T_k = j | D = d), \quad k = 1, 2, 3, \quad d = 0, \dots, L - 1, \quad j = 1, \dots, J_k.$$

Then finite mixture model (1) with $K = 3$ is globally identifiable if

$$\text{rank}_K M_1 + \text{rank}_K M_2 + \text{rank}_K M_3 \geq 2L + 2.$$

Proof:

Let L be a diagonal matrix, $L = \text{diag}\{P(D = 0), \dots, P(D = L - 1)\}$. Let $\tilde{M}_1 = LM_1$, then the (u, v, w) element of array $[M_1, M_2, M_3]$ is

$$\begin{aligned} [M_1, M_2, M_3](u, v, w) &= \sum_{d=0}^{L-1} P(D = d) M_1(d, u) M_2(d, v) M_3(d, w) \\ &= P(T_1 = u, T_2 = v, T_3 = w). \end{aligned}$$

According to Lemma 2, the finite mixture model (1) uniquely determines \tilde{M}_1 , M_2 and M_3 up to a simultaneous permutation and re-scaling of the rows. Meanwhile, since $\sum_{d=0}^{L-1} P(D = d) = 1$, $\sum_{j=1}^{J_k} P(T_k = j | D = d) = 1$ for all $d = 0, \dots, L - 1$, it follows that: $\sum_{d=0}^{L-1} \sum_{j=1}^{J_1} \tilde{M}_1 = 1$, $\sum_{d=0}^{L-1} \sum_{j=1}^{J_2} M_2 = L$ and $\sum_{d=0}^{L-1} \sum_{j=1}^{J_3} M_3 = L$. Thus, the scaling of the rows is uniquely determined. Moreover, suppose that \tilde{M}_1 is properly scaled, then $P(D = d) = \sum_{j=1}^{J_1} \tilde{M}_1(d, j)$, $d = 0, \dots, L - 1$, and $M_1 = L^{-1} \tilde{M}_1$. Therefore, M_1 , M_2 and M_3 are uniquely determined up to simultaneous permutation of the rows, and thus finite mixture model (1) is globally identifiable. \square

When $K \geq 3$ in model (1), several univariate tests can be viewed as a single multivariate test and Theorem 3 can be used to examine the global identifiability of the model. Specifically, we have the following theorem,

Theorem 4

Suppose $K \geq 3$. Let A, B, C be a partition of $\{1, \dots, K\}$, where $A = \{a(1), \dots, a(p)\}$, $B = \{b(1), \dots, b(q)\}$ and $C = \{c(1), \dots, c(r)\}$, with $p \geq 1$, $q \geq 1$, $r \geq 1$, and $p + q + r = K$. Let M_A be a L by $\prod_{i=1}^p J_{a(i)}$ matrix with the $(d + 1, j)$ element

$$M_A(d + 1, j) = P((T_{a(1)} \dots, T_{a(p)}) = \vec{t}_{Aj} | D = d),$$

where \vec{t}_{Aj} is the j th possible pattern in lexicographic order among $\prod_{i=1}^p J_{a(i)}$ distinct response patterns of $(T_{a(1)}, \dots, T_{a(p)})$. Similarly, define M_B as a L by $\prod_{i=1}^q J_{b(i)}$ matrix with the $(d+1, j)$ element $M_B(d+1, j) = P((T_{b(1)}, \dots, T_{b(q)}) = \vec{t}_{Bj} | D = d)$, and M_C as a L by $\prod_{i=1}^r J_{c(i)}$ matrix with the $(d+1, j)$ element $M_C(d+1, j) = P((T_{c(1)}, \dots, T_{c(r)}) = \vec{t}_{Cj} | D = d)$. Then the finite mixture model (1) is globally identifiable if $\text{rank}_K M_A + \text{rank}_K M_B + \text{rank}_K M_C \geq 2L + 2$.

Proof:

Applying Theorem 3, parameters $P(D = d)$, $P((T_{a(1)}, \dots, T_{a(p)}) | D = d)$, $P((T_{b(1)}, \dots, T_{b(q)}) | D = d)$ and $P((T_{c(1)}, \dots, T_{c(r)}) | D = d)$, $D = 0, \dots, L-1$, can be uniquely identified up to label switching. Moreover, the marginal distributions $P(T_k = j | D = d)$, $k = 1, \dots, K$, $d = 0, \dots, L-1$, $j = 1, \dots, J_k$ can be obtained from these joint distributions, thus the theorem holds. \square

Theorem 4 also illustrates why information from additional tests is helpful for model identification, since the matrix for multiple manifest variables has a larger Kruskal rank than the one for a single manifest variable. To see this more clearly, let M_1 be a L by J_1 matrix, with the $(d+1, j)$ element $M_1(d+1, j) = P(T_1 = j | D = d)$, and M_2 be a L by J_2 matrix, with the $(d+1, j)$ element $M_2(d+1, j) = P(T_2 = j | D = d)$. Additionally, let L_j be a diagonal matrix $L_j = \text{diag}\{M_2(0, j), \dots, M_2(L-1, j)\}$, $j = 1, \dots, J_2$. Then the matrix according to the joint distribution of T_1 and T_2 is a L by $J_1 J_2$ matrix as follows,

$$M = (L_1 M_1, \dots, L_{J_2} M_1).$$

Because L_j is a diagonal matrix with all diagonal elements positive, if any of the row vectors in M_1 are linearly independent, the same rows of the vectors in $L_1 M_1$ are also linearly independent, and vice versa. Additionally, the row vectors in M are an extension group of the row vectors in $L_1 M_1$, thus $\text{rank}_K M \geq \text{rank}_K M_1$. This argument can be generalized to several manifest variables. Consequently, we have the following corollary.

Corollary 2

Let M_k be a L by J_k matrix, with the $(d+1, j)$ element

$$M_k(d+1, j) = P(T_k = j | D = d), \quad k = 1, \dots, K, \quad d = 0, \dots, L-1, \quad j = 1, \dots, J_k.$$

If there exist $k_1, k_2, k_3 \in \{1, \dots, K\}$ such that

$$\text{rank}_K M_{k_1} + \text{rank}_K M_{k_2} + \text{rank}_K M_{k_3} \geq 2L + 2,$$

finite mixture model (1) is globally identifiable.

Moreover, if a L by J_k matrix M has full row rank, $\text{rank}_K M = L$, this gives the following corollary.

Corollary 3

Let M_k be a L by J_k matrix, with the $(d+1, j)$ element

$$M_k(d+1, j) = P(T_k = j | D = d), \quad k = 1, \dots, K, \quad d = 0, \dots, L-1, \quad j = 1, \dots, J_k.$$

If there exist $k_1, k_2, k_3 \in \{1, \dots, K\}$ such that M_{k_1} , M_{k_2} , M_{k_3} have full row rank, then the finite mixture model (1) is globally identifiable.

Now we consider the identifiability of a finite mixture model with continuous manifest variables. Let $f_{kd}(\cdot) = P(T_k|D = d)$ be the conditional density function of the k th manifest variable in group d , and $F_{kd}(\cdot)$ be the corresponding cumulative distribution function (CDF), $k = 1, \dots, K$, $D = 0, \dots, L - 1$. The model can be expressed as follows,

$$P(T_1 = t_1, \dots, T_K = t_K) = \sum_{d=0}^{L-1} P(\vec{T}, D) = \sum_{d=0}^{L-1} [P(D = d) \prod_{k=1}^K f_{kd}(t_k)]. \quad (7)$$

When $K = 3$, we show that the following theorem holds.

Theorem 5

If there exists integer $J_1, J_2, J_3 \geq L$ and points $t_{11}, \dots, t_{1(J_1-1)}$, $t_{21}, \dots, t_{2(J_2-1)}$ and $t_{31}, \dots, t_{3(J_3-1)}$, such that matrix M_1 , M_2 and M_3 satisfy that

$$\text{rank}_K M_1 + \text{rank}_K M_2 + \text{rank}_K M_3 \geq 2L + 2,$$

where the $(d + 1, j)$ element in M_k is $F_{kd}(t_{kj})$, $j = 1, \dots, J_k - 1$, and $F_{kd}(t_{J_k}) = 1$, $k = 1, 2, 3$, $d = 0, \dots, L - 1$, then the finite mixture model (7) is globally identifiable.

To prove Theorem 5, we first show that the following lemma holds.

Lemma 3

Let A and A^* be two P by Q matrices, defined below,

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1Q} \\ \vdots & \vdots & & \vdots \\ a_{P1} & a_{P2} & \dots & a_{PQ} \end{pmatrix} \quad A^* = \begin{pmatrix} a_{11} & a_{12} - a_{11} & \dots & a_{1Q} - a_{1Q-1} \\ \vdots & \vdots & & \vdots \\ a_{P1} & a_{P2} - a_{P1} & \dots & a_{PQ} - a_{PQ-1} \end{pmatrix},$$

then $\text{rank}_K A = \text{rank}_K A^*$.

Proof of Lemma 3

We only need to show that if any rows in A are linearly independent, the corresponding rows in A^* are also linearly independent, and vice versa. Equivalently, we can show that if any rows in A are linearly dependent, the corresponding rows in A^* are also linearly dependent, and vice versa. Without loss of generality, we assume the first $p \leq P$ rows, denoted by A_p , are linearly dependent. Then there exist k_1, \dots, k_p , such that $(k_1, \dots, k_p)A_p = 0$, in other words, the following equations hold

$$\begin{cases} k_1 a_{11} + \dots + k_p a_{p1} = 0 \\ k_1 a_{12} + \dots + k_p a_{p2} = 0 \\ \dots \\ k_1 a_{1Q} + \dots + k_p a_{pQ} = 0 \end{cases}.$$

These are equivalent to the following equations

$$\begin{cases} k_1 a_{11} + \dots + k_p a_{p1} = 0 \\ k_1 (a_{12} - a_{11}) + \dots + k_p (a_{p2} - a_{p1}) = 0 \\ \dots \\ k_1 (a_{1Q} - a_{1Q-1}) + \dots + k_p (a_{pQ} - a_{pQ-1}) = 0 \end{cases}.$$

Thus $(k_1, \dots, k_p)A_p^* = 0$, where A_p^* are the first p rows of A^* . As a result, if any rows in A are linearly independent, the corresponding rows in A^* are also linearly independent, and vice versa.

Therefore, $\text{rank}_K A = \text{rank}_K A^*$. \square

Proof of Theorem 5

Matrix M_k , $k = 1, 2, 3$, in theorem 5 is

$$M_k = \begin{pmatrix} F_{k0}(t_{k1}) & F_{k0}(t_{k2}) & \dots & F_{k0}(t_{kJ_k-1}) & 1 \\ \vdots & \vdots & & \vdots & \vdots \\ F_{kL-1}(t_{k1}) & F_{kL-1}(t_{k2}) & \dots & F_{kL-1}(t_{kJ_k-1}) & 1 \end{pmatrix}.$$

Create new matrices M_k^* , $k = 1, 2, 3$, as follows:

$$M_k^* = \begin{pmatrix} F_{k0}(t_{k1}) & F_{k0}(t_{k2}) - F_{k0}(t_{k1}) & \dots & 1 - F_{k0}(t_{kJ_k-1}) \\ \vdots & \vdots & & \vdots \\ F_{kL-1}(t_{k1}) & F_{kL-1}(t_{k2}) - F_{kL-1}(t_{k1}) & \dots & 1 - F_{kL-1}(t_{kJ_k-1}) \end{pmatrix}.$$

Then according to Lemma 3,

$$\text{rank}_K M_1^* + \text{rank}_K M_2^* + \text{rank}_K M_3^* = \text{rank}_K M_1 + \text{rank}_K M_2 + \text{rank}_K M_3 \geq 2L + 2.$$

Meanwhile, create categorical variables T_k^* , $k = 1, 2, 3$, such that $T_k^* = j$ if $t_{kj-1} \leq T_k \leq t_{kj}$, where $j = 1, \dots, J_k$ and $a_0 = -\infty$. Then according to Theorem 3, the finite mixture model (1) with T_1^* , T_2^* and T_3^* as manifest variables is globally identifiable. Specially, parameters $P(D = d)$, $F_{kd}(t_{k1})$, $F_{kd}(t_{k2}) - F_{kd}(t_{k1})$, \dots , $1 - F_{kd}(t_{kJ_k-1})$ are globally identifiable, which in turn leads to parameters $P(D = d)$, $F_{kd}(t_{k1})$, $F_{kd}(t_{k2})$, \dots , $F_{kd}(t_{kJ_k-1})$ being globally identifiable, $d = 0 \dots, D$ and $k = 1, 2, 3$.

Let t_k^* be any number in the domain of $F_{kd}(\cdot)$. Without loss of generality, we assume $t_k^* < t_{k1}$, and create matrices \tilde{M}_k , $k = 1, 2, 3$ as follows:

$$\tilde{M}_k = \begin{pmatrix} F_{k0}(t_k^*) & F_{k0}(t_{k1}) & \dots & F_{k0}(t_{kJ_k-1}) & 1 \\ \vdots & \vdots & & \vdots & \vdots \\ F_{kL-1}(t_k^*) & F_{kL-1}(t_{k1}) & \dots & F_{kL-1}(t_{kJ_k-1}) & 1 \end{pmatrix}.$$

The row vectors in \tilde{M}_k are extension groups of the row vectors in M_k , $k = 1, 2, 3$. Therefore,

$$\text{rank}_K \tilde{M}_1 + \text{rank}_K \tilde{M}_2 + \text{rank}_K \tilde{M}_3 \geq \text{rank}_K M_1 + \text{rank}_K M_2 + \text{rank}_K M_3 \geq 2L + 2.$$

Constructing the corresponding finite mixture model and following the same argument above, we can show that parameters $P(D = d)$, $F_{kd}(t_k^*)$, $F_{kd}(t_{k1})$, \dots , $F_{kd}(t_{kJ_k-1})$ are globally identifiable, $d = 0 \dots, D$ and $k = 1, 2, 3$. Since t_k^* is any number in the domain of $F_{kd}(\cdot)$, the function $F_{kd}(\cdot)$ is globally identifiable. As a result, the finite mixture model (7) is globally identifiable. \square

Similarly as in Theorem 4, several univariate tests as a multivariate test to obtain global identifiability of model (7) when $K \geq 3$. Specifically, we have the following theorem.

Theorem 6

Suppose $K \geq 3$. Let A , B , C be a partition of $\{1, \dots, K\}$, where $A = \{a(1), \dots, a(p)\}$, $B = \{b(1), \dots, b(q)\}$ and $C = \{c(1), \dots, c(r)\}$, with $p \geq 1$, $q \geq 1$, $r \geq 1$, and $p + q + r = K$. Let G_{1d} be the joint CDF for manifest variables $(T_{a(1)}, \dots, T_{a(p)})$ in group d , G_{2d} be the joint CDF for

manifest variables $(T_{b(1)}, \dots, T_{b(q)})$ in group d and G_{3d} be the joint CDF for manifest variables $(T_{c(1)}, \dots, T_{c(r)})$ in group d , $d = 0, \dots, L - 1$.

If there exists integers $J_1, J_2, J_3 \geq L$ and points $t_{11}, \dots, t_{1(J_1-1)}, t_{21}, \dots, t_{2(J_2-1)}$ and $t_{31}, \dots, t_{3(J_3-1)}$, such that matrices M_1 , M_2 and M_3 satisfy

$$\text{rank}_K M_1 + \text{rank}_K M_2 + \text{rank}_K M_3 \geq 2L + 2,$$

where the $(d + 1, j)$ element in M_k is $G_{kd}(t_{kj})$, $j = 1, \dots, J_k - 1$, and $G_{kd}(t_{J_k}) = 1$, $k = 1, 2, 3$, $d = 0, \dots, L - 1$, then the finite mixture model (7) is globally identifiable. \square

Moreover, with the same argument for proving Corollary 2 and 3, we know that if at least three of the manifest variables satisfy the condition in Theorem 5, then the finite mixture model (7) is globally identifiable. Therefore, we have the following corollaries.

Corollary 4

If there exist $s_1, s_2, s_3 \in \{1, \dots, K\}$, integers $J_1, J_2, J_3 \geq L$ and points $t_{11}, \dots, t_{1(J_1-1)}, t_{21}, \dots, t_{2(J_2-1)}$ and $t_{31}, \dots, t_{3(J_3-1)}$, such that matrices M_1 , M_2 and M_3 satisfy that,

$$\text{rank}_K M_1 + \text{rank}_K M_2 + \text{rank}_K M_3 \geq 2L + 2,$$

where the $(d + 1, j)$ element in M_k is $F_{s_k d}(t_{kj})$, $j = 1, \dots, J_k - 1$, and $F_{s_k d}(t_{J_k}) = 1$, $k = 1, 2, 3$, $d = 0, \dots, L - 1$, then the finite mixture model (7) is globally identifiable. \square

Corollary 5

If there exists $s_1, s_2, s_3 \in \{1, \dots, K\}$, integers $J_1, J_2, J_3 \geq L$ and points $t_{11}, \dots, t_{1(J_1-1)}, t_{21}, \dots, t_{2(J_2-1)}$ and $t_{31}, \dots, t_{3(J_3-1)}$, such that matrices M_1 , M_2 and M_3 have full row rank, where the $(d + 1, j)$ element in M_k is $F_{s_k d}(t_{kj})$, $j = 1, \dots, J_k - 1$, and $F_{s_k d}(t_{J_k}) = 1$, $k = 1, 2, 3$, $d = 0, \dots, L - 1$, then the finite mixture model (7) is globally identifiable. \square

From the theorem above, we can see that one important condition to guarantee global identifiability of a finite mixture model is about the row ranks of a matrix whose $(d + 1, j)$ element is $P(T = j|D = d)$. We give an intuitive explanation here before moving to be next section. Each row of this matrix is about the conditional distribution of manifest variable T in a certain group. If the rows are linearly independent, it means the manifest variable T can reveal some nontrivial differences among the latent groups that can then be used to distinguish between them. Identifying latent groups is more fundamental, as once the group labels are determined, other parameters can be obtained by, in some sense, regular regression. Another explanation can be obtained by observing that the marginal distribution of T is a linear combination of its conditional distributions in each of the latent groups. The factorization is unique only when the conditional distributions are linearly independent. However, since we have multiple manifest variables, we may not need to require that every one of them be informative for all latent groups – for example, if some manifest variables are informative for all groups except groups 1 and 2, the model may still be identifiable if there is a manifest variable that can distinguish between these two groups. Clearly, the task is harder if the number of latent groups is bigger. The theorems here show a balance between the information needed from the manifest variables and the number of latent groups needed for the model to be identifiable.

4.3 Models with Covariates

In this section we examine the global identifiability of a finite mixture model with covariates. Again, we assume that covariate effects are linear on a certain transformed scale. First we consider model (6) – a model in which all manifest variables are categorical. The model is given as follows,

$$P(\vec{T} = \vec{t} | \vec{X}, \vec{Z}) = \sum_{d=0}^{L-1} \pi_d(\vec{Z}^T \alpha_d) \prod_{k=1}^K \prod_{j=1}^{J_k} g_{kjd}(\vec{X}^T \beta_{kjd})^{I[t_k=j]},$$

$$\text{where } \pi_d(\vec{Z}^T \alpha_d) = P(D = d | \vec{Z}), \quad g_{kjd}(\vec{X}^T \beta_{kjd}) = P(T_{ik} = j | D_i = d, \vec{X}).$$

When $K = 3$ we obtain the following theorem.

Theorem 7

In model (6), suppose α_d has $u \leq N$ elements and β_{kd} has $v \leq N$ elements. Let $\max\{u, v\} \leq W \leq N$ and $\{i(1), \dots, i(W)\}$ be a subset of $\{1, \dots, N\}$. Let $M_{k[w]}$ be a L by J_k matrix with the $(d+1, j)$ element $g_{kjd}(\mathbf{X}_{i(w)} \beta_{kjd})$, where $\mathbf{X}_{i(w)}$ is the $i(w)$ th row of the design matrix \mathbf{X} , $d = 0, \dots, L-1$, $k = 1, 2, 3$, $j = 1, \dots, J_k$, $w = 1, \dots, W$. Let M_k be a block diagonal matrix defined as follows,

$$M_k = \begin{pmatrix} M_{k[1]} & 0 & \dots & 0 \\ 0 & M_{k[2]} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & M_{k[W]} \end{pmatrix}, \quad k = 1, 2, 3.$$

Let $\mathbf{X}_w = (\mathbf{X}_{i(1)}, \dots, \mathbf{X}_{i(W)})^T$ and $\mathbf{Z}_w = (\mathbf{Z}_{i(1)}, \dots, \mathbf{Z}_{i(W)})^T$. Then, the finite mixture model (6), with $K = 3$, is globally identifiable if the following conditions hold.

- (i) $\text{rank}_K M_1 + \text{rank}_K M_2 + \text{rank}_K M_3 \geq 2WL + 2$;
- (ii) $\text{rank } \mathbf{X}_w \geq u$ and $\text{rank } \mathbf{Z}_w \geq v$.

Proof:

The idea is to “absorb” the observed covariate pattern into manifest variables, and then use the results about models without covariates to complete the proof.

Create a new categorical variable D^* based on pairs (D, \mathbf{x}_k) such that $D^* = L(w-1) + d$ if $D = d$ and $\mathbf{x}_k = \mathbf{X}_{i(w)}$, $d = 0, \dots, L-1$, $w = 1, \dots, W$. Additionally, create new categorical variables T_k^* based on pairs (T_k, \mathbf{x}_k) such that, $T_k^* = J_k(w-1) + j$ if $T_k = j$ and $\mathbf{x}_k = \mathbf{X}_{i(w)}$, $k = 1, 2, 3$, $j = 1, \dots, J_k$, $w = 1, \dots, W$. Then the marginal probability of triplet (T_1^*, T_2^*, T_3^*) is

$$P(T_1^*, T_2^*, T_3^*) = \begin{cases} P(T_1, T_2, T_3 | \mathbf{X}_{i(w)}) & \text{if } \mathbf{x}_1 = \mathbf{x}_2 = \mathbf{x}_3 = \mathbf{X}_{i(w)} \\ 0 & \text{Otherwise} \end{cases}.$$

Construct a finite mixture model of form (2) with manifest variables T_1^*, T_2^*, T_3^* and latent variable D^* . Then, based on Theorem 3, condition (i) guarantees that this new model is globally identifiable. It follows that quantities $\pi_d(\mathbf{Z}_{i(w)}^T \alpha_d)$, $g_{kjd}(\mathbf{X}_{i(w)}^T \beta_{kjd})$ in the original model are globally identifiable, where $d = 0, \dots, L-1$, $k = 1, 2, 3$, $j = 1, \dots, J_k$, $w = 1, \dots, W$. Moreover, since $\pi_d(\cdot)$ and $g_{kjd}(\cdot)$

are pre-specified monotone link functions, the following equations hold:

$$\left\{ \begin{array}{c} \mathbf{Z}_{i(1)}^T \alpha_d = \pi_d^{-1}(a_{d1}) \\ \vdots \\ \mathbf{Z}_{i(W)}^T \alpha_d = \pi_d^{-1}(a_{dW}) \end{array} \right. \quad \text{and} \quad \left\{ \begin{array}{c} \mathbf{X}_{i(1)}^T \beta_{kd} = g_{kj d}^{-1}(b_{kj d1}) \\ \vdots \\ \mathbf{X}_{i(W)}^T \beta_{kd} = g_{kj d}^{-1}(b_{kj dW}) \end{array} \right.,$$

where a_{dw} is the value of $\pi_d(\mathbf{Z}_{i(w)}^T \alpha_d)$ and $b_{kj dw}$ is the value of $g_{kj d}(\mathbf{X}_{i(w)}^T \beta_{kd})$, $d = 0, \dots, L-1$, $k = 1, 2, 3$, $j = 1, \dots, J_k$, $w = 1, \dots, W$. Then, condition (ii) guarantees that at least u equations on the left hand side are linearly independent, and at least v equations on the right hand side are linearly independent. As a result, α_d , $\beta_{kj d}$ have unique solutions. Therefore, the finite mixture model (6) with $K = 3$ is globally identifiable. \square

Since multiple univariate manifest variables can be combined into a single multivariate manifest variable, we have the following theorem.

Theorem 8

Suppose $K \geq 3$, α_d has $u \leq N$ elements and β_{kd} has $v \leq N$ elements. Let $\max\{u, v\} \leq W \leq N$, and $\{i(1), \dots, i(W)\}$ be a subset of $\{1, \dots, N\}$. Let A, B, C be a partition of $\{1, \dots, K\}$, where $A = \{a(1), \dots, a(p)\}$, $B = \{b(1), \dots, b(q)\}$ and $C = \{c(1), \dots, c(r)\}$, with $p \geq 1$, $q \geq 1$, $r \geq 1$, and $p + q + r = K$. Let $M_{A[w]}$ be a L by $\prod_{i=1}^p J_{a(i)}$ matrix with the $(d+1, j)$ element

$$M_{A[w]}(d+1, j) = P((T_{a(1)}, \dots, T_{a(p)}) = \vec{t}_{Aj} | D = d, \mathbf{x} = \mathbf{X}_{i(w)}),$$

where \vec{t}_{Aj} is the j th possible in lexicographic order among $\prod_{i=1}^p J_{a(i)}$ distinct response patterns of $(T_{a(1)}, \dots, T_{a(p)})$. Let M_k be a block diagonal matrix defined as follows,

$$M_A = \begin{pmatrix} M_{A[1]} & 0 & \dots & 0 \\ 0 & M_{A[2]} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & M_{A[W]} \end{pmatrix}.$$

Define M_B and M_C similarly; then, the finite mixture model (6) is globally identifiable if the following conditions hold.

- (i) $\text{rank}_K M_A + \text{rank}_K M_B + \text{rank}_K M_C \geq 2WL + 2$.
- (ii) $\text{rank } \mathbf{X}_w \geq u$ and $\text{rank } \mathbf{Z}_w \geq v$. \square

Observing that the row vectors in the “M” matrix generated by multiple manifest variables is an extension group of the row vectors in the “M” matrix generated by any one of these manifest variables, we have the following corollary.

Corollary 6

When $K \geq 3$, if there are at least three manifest variables that satisfy the conditions in Theorem 7, then the finite mixture model (6) is globally identifiable. \square

Moreover, for a block diagonal matrix, the row vectors that intersect with different blocks are clearly linearly independent. Therefore, the Kruskal rank of a block diagonal matrix equals the summation of the Kruskal rank of each of the block. The following corollary results.

Corollary 7

Suppose $K \geq 3$. In the same set up as in Theorem 7, if there exist $k_1, k_2, k_3 \in \{1, \dots, K\}$ such that
 (i) $\text{rank}_K M_{k_1[w]} + \text{rank}_K M_{k_2[w]} + \text{rank}_K M_{k_3[w]} \geq 2L + 2$, for all $w = 1, \dots, W$;
 (ii) $\text{rank } \mathbf{X}_w \geq u$ and $\text{rank } \mathbf{Z}_w \geq v$,
 then the finite mixture model (6) is globally identifiable. \square

Now we consider the identifiability of a finite mixture model with continuous manifest variables. Again, we assumed that the covariate effects are linear on a certain transformed scale. Let $f_{kd}(t|\mathbf{X}_i\beta_{kd}) = P(T_k = t|D = d, \mathbf{X}_i)$ be the conditional density function of the k th manifest variable in group d and $F_{kd}(t|\mathbf{X}_i\beta_{kd})$ be the corresponding CDF, $k = 1, \dots, K$, $D = 0, \dots, L - 1$. we further assumed that after an unknown transformation H_{kd} , manifest variable T_k satisfied

$$(H_{kd}(T_k) | \mathbf{X}_i) = \mathbf{X}_i\beta_{kd} + \epsilon_{ikd}, \quad \epsilon_{ikd} \sim G_{kd}(\cdot),$$

where $G_{kd}(\cdot)$ is a pre-specified distribution function with corresponding density function $g_{kd}(\cdot)$, $k = 1, \dots, K$, $D = 0, \dots, L - 1$. The model can be expressed as follows,

$$P(T_1 = t_1, \dots, T_K = t_k | \mathbf{X}_i, \mathbf{Z}_i) = \sum_{d=0}^{L-1} \pi_d(\mathbf{Z}_i^T \alpha_d) \prod_{k=1}^K f_{kd}(t_k | \mathbf{X}_i\beta_{kd}). \quad (8)$$

When $K = 3$ we claim the following theorem holds.

Theorem 9

Suppose α_d has $u \leq N$ elements and β_{kd} has $v \leq N$ elements. Let $\max\{u, v + 1\} \leq W \leq N$ and $\{i(1), \dots, i(W)\}$ be a subset of $\{1, \dots, N\}$. Let J_1, J_2, J_3 be some positive integers. Let $M_{k[w]}$ be a L by J_k matrix with the $(d + 1, j)$ element $F_{kd}(t_{kj} | \mathbf{X}_i\beta_{kd})$, where t_{k1}, \dots, t_{kJ_k} are a set of points in the domain of $F_{kd}(t | \mathbf{X}_i\beta_{kd})$, and $\mathbf{X}_{i(w)}$ is the $i(w)$ th row of the design matrix \mathbf{X} , $d = 0, \dots, L - 1$, $k = 1, 2, 3$, $j = 1, \dots, J_k$, $w = 1, \dots, W$. Let M_k be a block diagonal matrix defined as follows,

$$M_k = \begin{pmatrix} M_{k[1]} & 0 & \dots & 0 \\ 0 & M_{k[2]} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & M_{k[W]} \end{pmatrix}, \quad k = 1, 2, 3.$$

Let $\mathbf{X}_w = (\mathbf{X}_{i(1)}, \dots, \mathbf{X}_{i(W)})^T$ and $\mathbf{Z}_w = (\mathbf{Z}_{i(1)}, \dots, \mathbf{Z}_{i(W)})^T$. Then, finite mixture model (8) with $K = 3$ is globally identifiable if the following conditions hold.

- (i) $\text{rank}_K M_1 + \text{rank}_K M_2 + \text{rank}_K M_3 \geq 2WL + 2$;
- (ii) $\text{rank } \mathbf{X}_w \geq u$ and $\text{rank } \mathbf{Z}_w \geq v + 1$.

Proof:

Using Lemma 3 and following the same logic as in the proof of Theorem 5, we can show that for any t_k^* in the domain of $F_{kd}(t | \mathbf{X}_i\beta_{kd})$, quantities $\pi_d(\mathbf{Z}_{i(w)}^T \alpha_d)$ and $F_{kd}(t_k^* | \mathbf{X}_i\beta_{kd})$, $F_{kd}(t_{kj} | \mathbf{X}_i\beta_{kd})$ are identifiable, $d = 0, \dots, L - 1$, $k = 1, 2, 3$, $j = 1, \dots, J_k$, $w = 1, \dots, W$. Moreover, since

$$\begin{aligned} F_{kd}(t_k^* | \mathbf{X}_i\beta_{kd}) &= P(T_k \leq t_k^* | \mathbf{X}_i\beta_{kd}) \\ &= P(H_{kd}(T_k) \leq H_{kd}(t_k^*) | \mathbf{X}_i\beta_{kd}) = G_{kd}(H_{kd}(t_k^*) - \mathbf{X}_i\beta_{kd}), \end{aligned}$$

we have

$$\left\{ \begin{array}{c} \mathbf{Z}_{i(1)}^T \alpha_d = \pi_d^{-1}(a_{d1}) \\ \vdots \\ \mathbf{Z}_{i(W)}^T \alpha_d = \pi_d^{-1}(a_{dW}) \end{array} \right. \quad \text{and} \quad \left\{ \begin{array}{c} H_{kd}(t_k^*) - \mathbf{X}_{i(1)}^T \beta_{kd} = G_{kd}^{-1}(b_{kjd1}) \\ \vdots \\ H_{kd}(t_k^*) - \mathbf{X}_{i(W)}^T \beta_{kd} = G_{kd}^{-1}(b_{kjdW}) \end{array} \right.,$$

where a_{dw} is the value of $\pi_d(\mathbf{Z}_{i(w)}^T \alpha_d)$ and b_{kjdw} is the value of $F_{kd}(t_k^* | \mathbf{X}_i \beta_{kd})$, $d = 0, \dots, L-1$, $k = 1, 2, 3$, $j = 1, \dots, J_k$, $w = 1, \dots, W$. Note that the equations on the right hand side are still linear equations of \mathbf{X}_i , so condition (ii) guarantees that α_d , $\beta_{kj d}$ and $H_{kd}(t_k^*)$ have unique solutions. Since t_k^* is arbitrary, the finite mixture model (8) with $K = 3$ is globally identifiable. \square

Comparing Theorem 9 to Theorem 7, the additional rank of \mathbf{Z}_w was used for identification of $H_{kd}(\cdot)$. Applying the same techniques as before, it was straightforward to obtain the following theorem and corollaries.

Theorem 10

Suppose that $K \geq 3$, α_d has $u \leq N$ elements and β_{kd} has $v \leq N$ elements. Let $\max\{u, v+1\} \leq W \leq N$, and $\{i(1), \dots, i(W)\}$ be a subset of $\{1, \dots, N\}$. Let J_1, J_2, J_3 be some positive integers and t_{k1}, \dots, t_{kJ_k} be a set of points in the domain of $F_{kd}(t | \mathbf{X}_i \beta_{kd})$, $k = 1, \dots, K$. Let A, B, C be a partition of $\{1, \dots, K\}$, where $A = \{a(1), \dots, a(p)\}$, $B = \{b(1), \dots, b(q)\}$ and $C = \{c(1), \dots, c(r)\}$, with $p \geq 1$, $q \geq 1$, $r \geq 1$, and $p + q + r = K$. Let $M_{A[w]}$ be a L by $\prod_{i=1}^p J_{a(i)}$ matrix with the $(d+1, j)$ element $\mathbb{F}_{kd}(\vec{t}_{A_j} | \mathbf{X}_i \beta_{kd})$, where $\mathbb{F}_{kd}(\cdot | \mathbf{X}_i \beta_{kd})$ is the CDF of $(T_{a(1)}, \dots, T_{a(p)})$ conditional on \mathbf{X}_i , and \vec{t}_{A_j} is the j th possible in lexicographic order among $\prod_{i=1}^p J_{a(i)}$ distinct response patterns of $(T_{a(1)}, \dots, T_{a(p)})$, generated by point $t_{a(s)1}, \dots, t_{a(s)J_{a(s)}}$, $s = 1, \dots, p$. Let M_k be a block diagonal matrix defined as follows,

$$M_A = \begin{pmatrix} M_{A[1]} & 0 & \dots & 0 \\ 0 & M_{A[2]} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & M_{A[W]} \end{pmatrix}.$$

Define M_B and M_C similarly; then the finite mixture model (8) is globally identifiable if the following conditions hold.

- (i) $\text{rank}_K M_A + \text{rank}_K M_B + \text{rank}_K M_C \geq 2WL + 2$.
- (ii) $\text{rank } \mathbf{X}_w \geq u$ and $\text{rank } \mathbf{Z}_w \geq v + 1$. \square

Corollary 8

When $K \geq 3$, if there are at least three manifest variables that satisfy the conditions in Theorem 9, then the finite mixture model (8) is globally identifiable. \square

Corollary 9

Suppose $K \geq 3$. In the same set-up as used in Theorem 9, if there exist $k_1, k_2, k_3 \in \{1, \dots, K\}$, such that

- (i) $\text{rank}_K M_{k_1[w]} + \text{rank}_K M_{k_2[w]} + \text{rank}_K M_{k_3[w]} \geq 2L + 2$, for all $w = 1, \dots, W$;
- (ii) $\text{rank } \mathbf{X}_w \geq u$ and $\text{rank } \mathbf{Z}_w \geq v + 1$,

then the finite mixture model (8) is globally identifiable. \square

5 Summary

In this paper we provided conditions under which a finite mixture model is locally identifiable or globally identifiable. The derivation was considered under general structures of finite mixture models, including models with categorical manifest variables, models with continuous manifest variables, and both types of models when covariates are included. One of the assumptions that we made here is that the covariate effects are linear on some transformed scales. This assumption is common in the literature on finite mixture models. Moreover, these transformations do not necessarily need to be pre-specified, because we established conditions for nonparametric identifiability of these models. Therefore, the results can be applied to a wide range of models. Additionally, we did not impose any constraints about the covariate effects among different groups or on different manifest variables. Therefore, the models considered here represent many general situations.

For local identifiability, a key idea in the proofs was to show that the Jacobian matrix of the model induced function had full column rank. The results suggest that, contrary to common belief, including covariates in the model may in fact help model identification. Additionally, by considering a continuous distribution in its empirical form for a given data set, we unified the proofs for models with categorical manifest variables or with continuous manifest variables, and showed nonparametric identifiability of the models. Consequently, the results can also be applied to cases when manifest variables have mixed types.

For global identifiability, the proof is established on a previous result about the uniqueness of trilinear decomposition [20]. We only discussed the situation where the number of manifest variables K is greater than or equal to 3. When $K = 2$, Hall and Zhou [15] showed that the model is not nonparametrically identifiable. The proof for models with continuous manifest variables used the results about models with categorical variables, and was accomplished by showing that the CDF was globally identifiable at every point in its domain. The results easily extend to models with mixed types of manifest variables.

In addition to allowing for flexible covariate structure, another useful direction in which finite mixture models has been developing is to include random effects to describe correlations introduced by unobserved covariates other than the latent components. We did not discuss the identifiability issue for this kind of models. Although the local identifiability of these models can be examined the same way using Jacobian matrices, it is usually critical to utilize background information and evaluate random effects in a case by case base to ensure model validity and interpretation. Aside from the potential computational burden, it is relatively easy and perhaps tempting to include more complicated and perhaps hierarchical random effects further account for possible residual dependence among the manifest variables within a latent group. These models should be adopted with caution, because when dealing with an unobserved latent structure, many modeling assumptions can hardly be verified. The possible danger of model misspecification or over-fitting results from building a complicated model may compromise its potential gains relative to using a simpler model. If a more complicated model were to be used, careful examination of the identifiability issue would be necessary.

References

- [1] Paul S Albert, Lisa M McShane, and Joanna H Shih. Latent class modeling approaches for assessing diagnostic error without a gold standard: with applications to p53 immunohistochemical assays in bladder tumors. *Biometrics*, 57(2):610–619, 2001.
- [2] Xiao-Hua Zhou, Pete Castelluccio, and Chuan Zhou. Nonparametric estimation of roc curves in the absence of a gold standard. *Biometrics*, 61(2):600–609, 2005.
- [3] Zheyu Wang, Xiao-Hua Zhou, and Miqu Wang. Evaluation of diagnostic accuracy in detecting ordered symptom statuses without a gold standard. *Biostatistics*, 12(3):567–581, 2011.
- [4] Margaret Sullivan Pepe. *The statistical evaluation of medical tests for classification and prediction*. Oxford University Press, 2003.
- [5] Richard B McHugh. Efficient estimation and local identification in latent class analysis. *Psychometrika*, 21(4):331–347, 1956.
- [6] Leo A Goodman. Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika*, 61(2):215–231, 1974.
- [7] Anton K Formann. Linear logistic latent class analysis for polytomous data. *Journal of the American Statistical Association*, 87(418):476–486, 1992.
- [8] Ryan Elmore, Peter Hall, and Amnon Neeman. An application of classical invariant theory to identifiability in nonparametric mixtures. In *Annales de l’institut Fourier*, volume 55, pages 1–28. Chartres: L’Institut, 1950-, 2005.
- [9] Guan-Hua Huang and Karen Bandeen-Roche. Building an identifiable latent class model with covariate effects on underlying and measured variables. *Psychometrika*, 69(1):5–32, 2004.
- [10] Antonio Forcina. Identifiability of extended latent class models with individual covariates. *Computational Statistics & Data Analysis*, 52(12):5263–5268, 2008.
- [11] Henry Teicher. Identifiability of finite mixtures. *The Annals of Mathematical Statistics*, pages 1265–1269, 1963.
- [12] WR Blischke. Estimating the parameters of mixtures of binomial distributions. *Journal of the American Statistical Association*, 59(306):510–528, 1964.
- [13] TP Hettmansperger and Hoben Thomas. Almost nonparametric inference for repeated measures in mixture models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 62(4):811–825, 2000.
- [14] IR Cruz-Medina, TP Hettmansperger, and H Thomas. Semiparametric mixture models and repeated measures: the multinomial cut point model. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 53(3):463–474, 2004.
- [15] Peter Hall and Xiao-Hua Zhou. Nonparametric estimation of component distributions in a multivariate mixture. *Annals of Statistics*, pages 201–224, 2003.
- [16] Peter Hall, Amnon Neeman, Reza Pakyari, and Ryan Elmore. Nonparametric inference in multivariate mixtures. *Biometrika*, 92(3):667–678, 2005.

- [17] Geoffrey Jones, Wesley O Johnson, Timothy E Hanson, and Ronald Christensen. Identifiability of models for multiple diagnostic testing in the absence of a gold standard. *Biometrics*, 66(3):855–863, 2010.
- [18] Sidney J Yakowitz, John D Spragins, et al. On the identifiability of finite mixtures. *The Annals of Mathematical Statistics*, 39(1):209–214, 1968.
- [19] Sui L Hui and Steven D Walter. Estimating the error rates of diagnostic tests. *Biometrics*, pages 167–171, 1980.
- [20] Joseph B Kruskal. Three-way arrays: rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics. *Linear algebra and its applications*, 18(2):95–138, 1977.
- [21] Elizabeth S Allman, Catherine Matias, and John A Rhodes. Identifiability of parameters in latent structure models with many observed variables. *The Annals of Statistics*, pages 3099–3132, 2009.