Identifiability of Finite Mixture Models

Jang SCHILTZ (University of Luxembourg)

joint work with

Cédric NOEL(University of Lorraine & University of Luxembourg)

SMTDA 2020
June 4, 2020

Outline

Outline

Outline

Outline

1. Introduction
2. Characterization of Identifiability
3. Identifiability of finite mixture models

Identifiability

Identifiability

Definition:

A finite mixture model is identifiable if a given dataset leads to a uniquely determined set of model parameter estimations up to a permutation of the clusters.

Identifiability of the parameters is a necessary condition for the existence of consistent estimators for any statistical model.

Without identifiability, there might be several solutions for the parameter estimation problem.

Literature Review

- Teicher (1963): The class of all mixtures of one-dimensional normal distributions is identifiable.

Literature Review

- Teicher (1963): The class of all mixtures of one-dimensional normal distributions is identifiable.
- Yakowitz and Spragins (1968): Extension to the class of all Gaussian mixtures.

Literature Review

- Teicher (1963): The class of all mixtures of one-dimensional normal distributions is identifiable.
- Yakowitz and Spragins (1968): Extension to the class of all Gaussian mixtures.
- Henning (2000): Identifiability for linear regression mixtures with Gaussian errors under certain conditions.

Finite Mixture Models

Data:

Variable of interest $Y_i = y_{i_1}, y_{i_2}, ..., y_{i_T}$

Covariants $x_1, ..., x_M$ and $z_{i_1}, ..., z_{i_T}$

a_{i_t} age of subject i at time t

Finite Mixture Models

Data:

Variable of interest $Y_i = y_{i_1}, y_{i_2}, ..., y_{i_T}$

Covariants $x_1, ..., x_M$ and $z_{i_1}, ..., z_{i_T}$

a_{i_t} age of subject i at time t

Model:

K groups of size π_k with trajectories

$$y_{i_t} = \sum_{j=0}^{s_k} \left(\beta_j^k + \sum_{m=1}^{M} \alpha_m^k x_m + \gamma_j^k z_{i_t} \right) a_{it}^j + \varepsilon_{it}^k, \qquad (1)$$

where $\varepsilon_{it}^k \sim \mathcal{N}(0, \sigma^k)$.

Outline

Notations

Distribution f of a finite mixture model:

$$f(y_i; \Omega) = \sum_{k=1}^{K} \pi_k g_k(y_i; \theta^k).$$

Cumulative distribution function F of a finite mixture model:

$$F(y_i; \Omega) = \sum_{k=1}^{K} \pi_k G_k(y_i; \theta^k).$$

Mixtures and mixing distributions

Let $\mathcal{F} = \left\{F(y;\omega),\ y \in \mathbb{R}^T,\ \omega \in \mathbb{R}_K^{s+2}\right\}$ be a family of T-dimensional cdf's indexed by a parameter set ω, such that $F(y;\omega)$ is measurable in $\mathbb{R}^T \times \mathbb{R}_K^{s+2}$.

The the $s+2$-dimensional cdf $H(x) = \int_{\mathbb{R}_K^{s+2}} F(y;\omega)dG(\omega)$ is the image of the above mapping, of the $s+2$-dimensional cdf G.

The distribution H is called the mixture of $\mathcal{F}$ and G its mixing distribution.

Let $\mathcal{G}$ denote the class of all $s+2$-dimensional cdf's G and $\mathcal{H}$ the induced class of mixtures H.

Then $\mathcal{H}$ is identifiable if Q is a one-to-one map from $\mathcal{G}$ onto $\mathcal{H}$.

Characterization of identifiability

The set $\mathcal{H}$ of all finite mixtures of class $\mathcal{F}$ of distributions is the convex hull of $\mathcal{F}$.

$$\mathcal{H} = \left\{ H(y) : H(y) = \sum_i c_i F(y, \omega_i),\ c_i > 0, \sum_i c_i = 1,\ F(y, \omega_i) \in \mathcal{F} \right\}. \tag{2}$$

Theorem

A necessary and sufficient condition for the class $\mathcal{H}$ of all finite mixtures of the family $\mathcal{F}$ to be identifiable is that $\mathcal{F}$ is a linearly independent family over the field of real numbers.

Outline

The Model

$$Y_{it} = f(a_{it}; \beta^k, \delta^k) + \varepsilon_{it}^k = \beta^k A_{it} + \delta^k W_{it} + \varepsilon_{it}^k. \quad (3)$$

We can write

$$\mathcal{L}\left((Y_i)_{i \in I}\right) = \bigotimes_{i \in I} F_{A_i, W_i, J}. \quad (4)$$

Identifiability of a model means that knowing the data distribution $\mathcal{L}(Y_i), i \in I$, one can uniquely identify the mixing distribution J.

That is, no two distinct sets of parameters lead to the same data distribution.

Nagin's base model

$$\mathcal{C}_1 = \left(F_{A,J} \ : \ F_{A,J} = \bigotimes_{i \in I} F_{A_i,J} \right)_{J \in \Omega_1}$$

Theorem

Let $h_j = \min \left\{ q \ : \ \{A_{ij}, i \in I\} \subseteq \cup_{i=1}^{q} H_i \ \ H_i \in \mathcal{H}_{n-1} \right\}$.
If there exist j *such that* $|S(J)| < h_j, \ \forall J$ *then* $\mathcal{C}_1$ *is identifiable.*

Adding covariates independent of cluster membership

$$\mathcal{C}_2 = \left(F_{A,J} \ : \ F_{A,J} = \bigotimes_{i \in I} F_{A_i, W_i, J} \right)_{J \in \Omega_1}, \quad (5)$$

$$\mathcal{C}_{2A} = \left(F_{A,J} \ : \ F_{A,J} = \bigotimes_{i \in I} F_{A_i, J} \right)_{J \in \Omega_1}, \quad (6)$$

$$\mathcal{C}_{2W} = \left(F_{A,J} \ : \ F_{A,J} = \bigotimes_{i \in I} F_{W_i, J} \right)_{J \in \Omega_1}. \quad (7)$$

Theorem

If $\mathcal{C}_{2A}$ and $\mathcal{C}_{2W}$ are identifiable and W_{ij} is not a multiple of A_{ij}, for all i, j, then $\mathcal{C}_2$ is identifiable.

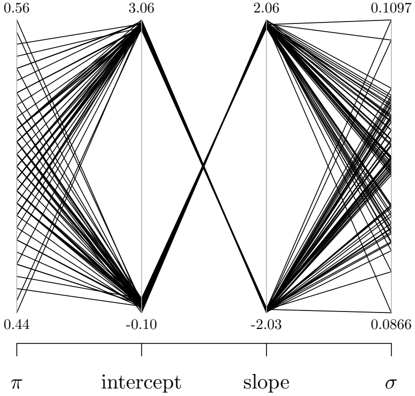
Numerical Example

- Two clusters with sizes $\pi_1 = \pi_2 = \frac{1}{2}$.
- Two time-points 1 and 2.
- Same variability in both clusters $\sigma = 0.1$

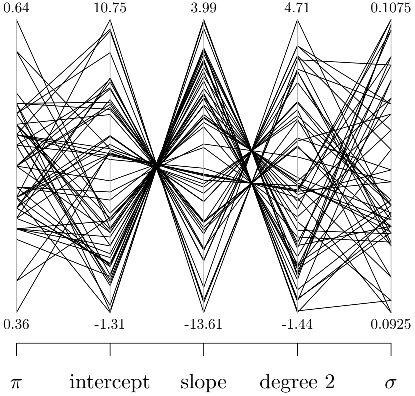
We simulate 50 samples of 100 trajectories with parameters

- $\beta^1 = (3, -2)$ and $\beta^2 = (0, 2)$ (linear model)
- $\beta^1 = (10, -12.5, 3.5)$ and $\beta^2 = (-2, 5, -1)$ (polynomial model).

Parallel coordinate plots of the estimated parameter



Linear Model



Parabolic Model

The generalized model

Theorem

The model is identifiable if

- $d_k < T$ *for all* $1 \leq k \leq K$ *and all* a_{it} *are distinct, for all* i_t.
- W_k *has full rank for all* $1 \leq k \leq K$.
- $rk(A_k, W_k) = rk(A_k) + rk(W_k)$, *for all* $1 \leq k \leq K$.

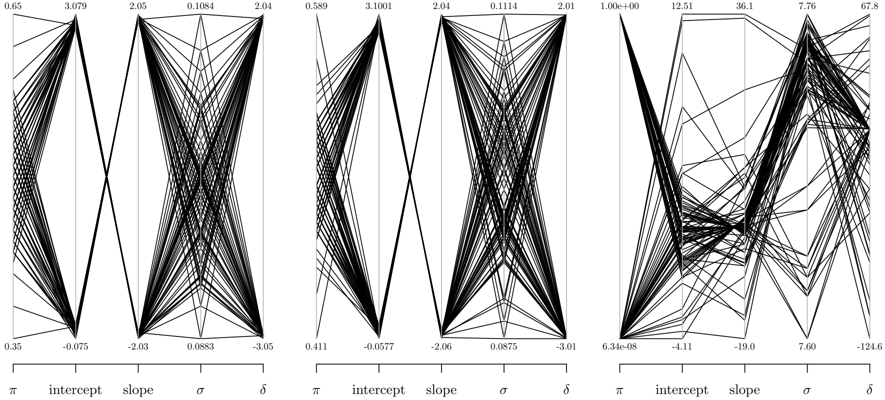
Numerical Example

- Two clusters with sizes $\pi_1 = \pi_2 = \frac{1}{2}$.
- Two time-points 1 and 2.
- Same variability in both clusters $\sigma = 0.1$
- Shape description parameters $\beta_1 = (3, -2)$, $\beta_2 = (0, 2)$, $\delta_1 = 2$ and $\delta_2 = -3$.

We simulate 50 samples of 100 trajectories for 3 types of models:

- The covariate is independent of time and only takes values 0 or 1
- The covariate is time dependent but in a nonlinear way
- The covariate is time dependent in a linear way

Parallel coordinate plots of the estimated parameter



Model 1 Model 2 Model 3

Bibliography

- Nagin, D.S. 2005: *Group-based Modeling of Development*. Cambridge, MA.: Harvard University Press.
- Schiltz, J. 2015: A generalization of Nagin's finite mixture model. In: Dependent data in social sciences research: Forms, issues, and methods of analysis' Mark Stemmler, Alexander von Eye & Wolfgang Wiedermann (Eds.). Springer 2015.
- Teicher, H. 1963: Identifiability of Finite Mixtures. *Annals of Mathematical Statistics*, 34,4,1265–1269.
- Yakowitz, S.J. and Spragins, J.D. 1968: On the identifiability of finite mixtures. *Annals of Mathematical Statistics*, 39, 209–214.
- Hennig, C. 2000: Identifiability of Models for Clusterwise Linear Regression. *Journal of Classification*, 17, 273–296.