A SURVEY ON THE IDENTIFIABILITY OF FINITE MIXTURES

BERLIAN SETIAWATY

Department of Mathematics,
Faculty of Mathematics and Natural Sciences,
Bogor Agricultural University
Jl. Raya Pajajaran, Kampus IPB Baranang Siang, Bogor, Indonesia

Abstract. This paper is a survey on the identifiability of finite mixtures. We collect all the results regarding sufficient conditions for finite mixtures to be identifiable and what kind of distributions family which is identifiable.

Key words: Finite mixture, identifiable.

1. Introduction

The purposes of this paper are to collect results concerning identifiability of finite mixtures which are scattered in several journals and books, and to present them as coherent as possible.

A good review of this subject can be found in [?], [?] and [?]. In particular, [?] provides extensive references.

We will begin with definition of finite mixtures and identifiability of finite mixtures. After that we discuss the identifiability properties of finite mixtures in the next section

2. Identifiability of Finite Mixtures

A formal definition of mixture distribution cited from [?] is as follows.

Definition 2.1. Let $\boldsymbol{\mathcal{F}} = \{F(\cdot,\theta) : \theta \in \mathcal{B}\}$ constitute a family of one dimensional distribution functions taking values in $\mathcal{Y}$ indexed by a point θ in a Borel subset $\mathcal{B}$ of Euclidean m-space $\boldsymbol{R}^m$, such that $F(\cdot,\cdot)$ is measurable in $\mathcal{Y} \times \mathcal{B}$. Let G be any distribution function such that the measure μ_G induced by G assigns measure 1 to $\mathcal{B}$. Then

$$H(y) = \int_{\mathcal{B}} F(y,\theta)\, d\mu_G(\theta) = \int_{\mathcal{B}} F(y,\theta)\, dG(\theta) \tag{2.1}$$

is called a ***mixture distribution*** and G is called the ***mixing distribution***.

Reference [?] gives a corresponding definition for a mixture density,

Definition 2.2. Let $\boldsymbol{\mathcal{F}} = \{f(\cdot,\theta) : \theta \in \mathcal{B}\}$ constitute a family of one dimensional density functions indexed by a point θ in a Borel subset $\mathcal{B}$ of Euclidean m-space $\boldsymbol{R}^m$, such that $f(\cdot,\cdot)$ is measurable in $\mathcal{Y} \times \mathcal{B}$. Let G be any distribution function such that the measure μ_G induced by G assigns measure 1 to $\mathcal{B}$. Then

$$h(y) = \int_{\mathcal{B}} f(y,\theta)\, d\mu_G(\theta) = \int_{\mathcal{B}} f(y,\theta)\, dG(\theta) \tag{2.2}$$

is called a ***mixture density*** and G is called the ***mixing distribution***.

Example 2.3. Let $\boldsymbol{\mathcal{F}}$ be the family of uniform distribution functions $U(a,b)$ with range (a,b), where $a \le b$. Let

$$G(a,b) = \tfrac{1}{2}\delta_{(-2,1)} + \tfrac{1}{2}\delta_{(-1,2)}, \quad a,b \in \boldsymbol{R}, \quad a \le b,$$

where $\delta_{(a,b)}$ is Dirac distribution of a point mass at (a,b). Then,

$$H(y) = \int_{\{(a,b): a,b\in \mathbf{R}, a\le b\}} U(a,b)\, dG(a,b) = \tfrac{1}{2}U(-2,1) + \tfrac{1}{2}U(-1,2),\ y \in \boldsymbol{R}$$

is a mixture distribution.

Example 2.4. Let $\boldsymbol{\mathcal{F}}$ be the family of Poisson density function $f(\cdot,\theta)$, where

$$f(y,\theta) = \frac{e^{-\theta}\theta^y}{y!}, \qquad y = 0,1,\dots \quad \text{and} \quad \theta \in (0,\infty).$$

Let

$$G(\theta) = e^{-\theta}, \quad \theta \in (0,\infty).$$

Then from the simple recurence $h(y) = \frac{1}{2}h(y-1)$, $h(0) = \frac{1}{2}$,

$$h(y) = \int_0^\infty \frac{e^{-\theta}\theta^y}{y!} \cdot e^{-\theta}\, d\theta = 2^{-(y+1)}, \quad y = 0,1,\dots$$

is a mixture density

From Definitions **??** and **??**, it can be seen that a mixture distribution and a mixture density can be derived from one another. Hence, it is enough to consider mixture distributions only. However, the results that we have for mixture distributions will also apply for mixture densities.

Let $\boldsymbol{\mathcal{G}}$ denote the class of all such m-dimensional distribution functions G and $\boldsymbol{\mathcal{H}}$ the induced class of mixtures H on $\boldsymbol{\mathcal{F}}$. Teicher [**?**] defines the identifiability of $\boldsymbol{\mathcal{H}}$ as follows.

Definition 2.5. A class of mixture distributions $\boldsymbol{\mathcal{H}}$ is said to be ***identifiable*** if and only if the equality of two representations

$$\int_{\mathcal{B}} F(y,\theta)\, dG(\theta) = \int_{\mathcal{B}} F(y,\theta)\, d\widehat{G}(\theta), \qquad \forall y \in \mathcal{Y}$$

implies $G = \widehat{G}$.

Definition **??** given above is the general one, but most of our applications are concerned with a special type of mixture. This type is generated by the special case when G is discrete and assigns positive probability to only finite number of points, as in the Example **??**.

Definition 2.6. H is called a ***finite mixture*** if its mixing distribution G or rather the corresponding measure μ_G is discrete and assigns positive mass to only a finite number of points in $\mathcal{B}$. Thus the class $\widetilde{\boldsymbol{\mathcal{H}}}$ of finite mixtures on $\boldsymbol{\mathcal{F}}$ is defined by

$$\begin{aligned}\widetilde{\boldsymbol{\mathcal{H}}} \;=\; & \Big\{H(\cdot) : H(\cdot) = \sum_{i=1}^{N} c_i F(\cdot, \theta_i),\ c_i > 0, \sum_{i=1}^{N} c_i = 1,\ F(\cdot,\theta_i) \in \boldsymbol{\mathcal{F}}, \\ & N = 1, 2 \ldots \Big\}\end{aligned}$$

that is, $\widetilde{\boldsymbol{\mathcal{H}}}$ is the convex hull of $\boldsymbol{\mathcal{F}}$.

Remarks 2.7. In every expression of finite mixture

$$H(\cdot) = \sum_{i=1}^{N} c_i F(\cdot, \theta_i),$$

$\theta_1, \ldots, \theta_N$ are assumed to be *distinct* members of Θ. The c_i and $\theta_i,\ i = 1, \ldots, N$ will be called respectively the *coeffients* and *support points* of the finite mixture.

Applying Definition **??** to the class of finite mixtures, we have the identifiability criteria for finite mixtures. The following formal definition states that the class of finite mixtures $\widetilde{\boldsymbol{\mathcal{H}}}$ is identifiable if and only if all members of $\widetilde{\boldsymbol{\mathcal{H}}}$ are distinct.

Definition 2.8. Let $\widetilde{\boldsymbol{\mathcal{H}}}$ be the class of finite mixtures on $\boldsymbol{\mathcal{F}}$. $\widetilde{\boldsymbol{\mathcal{H}}}$ is ***identifiable*** if and only if

$$\sum_{i=1}^{N} c_i F(\cdot, \theta_i) = \sum_{i=1}^{\widehat{N}} \widehat{c}_i F(\cdot, \widehat{\theta}_i)$$

implies $N = \widehat{N}$ and for each i, $1 \le i \le N$, there is j, $1 \le j \le N$, such that $c_i = \widehat{c}_j$ and $\theta_i = \widehat{\theta}_j$.

Definition **??** can be stated in different way using Dirac distributions. To show this, the following lemma is needed.

Lemma 2.9. *Suppose that*

$$\sum_{i=1}^{N} c_i \delta_{\theta_i} = \sum_{i=1}^{\widehat{N}} \widehat{c}_i \delta_{\widehat{\theta}_i}, \tag{2.3}$$

for

$$\begin{array}{lll} c_i \geq 0, & i = 1, \dots, N, & \sum_{i=1}^{N} c_i = 1 \\ \widehat{c}_i \geq 0, & i = 1, \dots, \widehat{N}, & \sum_{i=1}^{\widehat{N}} \widehat{c}_i = 1 \\ \theta_i, \theta_j \in \Theta, & , i = 1, \dots, N, & j = 1, \dots, \widehat{N} \end{array}$$

where δ_θ denotes the Dirac distribution of a point mass at θ.

1. *Suppose there are i, j, where $1 \leq i \leq N$ and $1 \leq j \leq \widehat{N}$, such that $\theta_i = \widehat{\theta}_j$. Let $D = \{k : \theta_k = \theta_i\}$ and $\widehat{D} = \{k : \widehat{\theta}_k = \widehat{\theta}_j\}$, then $\sum_{k \in D} c_k = \sum_{k \in \widehat{D}} \widehat{c}_k$.*
2. *If $c_i > 0$, for some i, $1 \leq i \leq N$, then there is j, $1 \leq j \widehat{N}$, such that $\theta_i = \widehat{\theta}_j$.*
3. *If $c_i > 0$ and θ_i are distinct, for $i = 1, \dots, N$, then $\widehat{N} \geq N$ and for every $i = 1, \dots N$, there is j, $1 \leq j \leq \widehat{N}$ such that $\theta_i = \widehat{\theta}_j$.*
4. *If $c_i > 0$ and θ_i are distinct for $i = 1, \dots, N$ and $N = \widehat{N}$, then there is a permutation σ on $\{1, \dots, N\}$ such that $c_i = \widehat{c}_{\sigma(i)}$ and $\theta_i = \widehat{\theta}_{\sigma(i)}$, for $i = 1, \dots, N$*

Proof :
To prove (a), suppose that there are i, j, where $1 \leq i \leq N$ and $1 \leq j \leq \widehat{N}$, such that $\theta_i = \widehat{\theta}_j$. Let $D = \{k : \theta_k = \theta_i\}$ and $\widehat{D} = \{k : \widehat{\theta}_k = \widehat{\theta}_j\}$. Let ζ be a smooth real function defined on Θ such that :

$$\zeta(\theta) = \begin{cases} 1, & \text{if} \quad \theta = \theta_i \\ 0, & \text{if} \quad \theta \in \Big(\{\theta_1, \dots, \theta_N\} \cup \{\widehat{\theta}_1, \dots, \widehat{\theta}_{\widehat{N}}\} \Big) \setminus \{\theta_i\}. \end{cases} \tag{2.4}$$

By (**??**) and (**??**),

$$\begin{aligned} \sum_{k=1}^{N} \int_\Theta c_k \zeta(\theta) \delta_{\theta_k}(\theta) &= \sum_{k=1}^{\widehat{N}} \int_\Theta \widehat{c}_k \zeta(\theta) \delta_{\widehat{\theta}_k}(\theta) \\ \sum_{k=1}^{N} c_k \zeta(\theta_k) &= \sum_{k=1}^{\widehat{N}} \widehat{c}_k \zeta(\widehat{\theta}_k) \\ \sum_{k \in D} c_k &= \sum_{k \in \widehat{D}} \widehat{c}_k. \end{aligned} \tag{2.5}$$

For (b), let $c_i > 0$, for some i, $1 \leq i \leq N$. Suppose that $\theta_i \neq \widehat{\theta}_j$, for every $j = 1, \dots, \widehat{N}$, then by (**??**)

$$\sum_{k \in D} c_k = 0,$$

implying $c_k = 0$ for all $k \in D$. Since $i \in D$, then $c_i = 0$, contradicting with $c_i > 0$. Thus there must be j, $1 \le j \le \widehat{N}$, such that $\theta_i = \widehat{\theta}_j$

For (c) and (d), suppose that $c_i > 0$ and θ_i are distinct, for $i = 1, \dots, N$. By part (b), for every $i = 1, \dots, N$, there is j, $1 \le j \le \widehat{N}$, such that

$$\theta_i = \widehat{\theta}_j. \tag{2.6}$$

Since θ_i are all distinct for $i = 1, \dots, N$, then it must be $\widehat{N} \ge N$. If $N = \widehat{N}$, the mapping $i \mapsto j$ is bijective. Let σ be the mapping and by (**??**),

$$\theta_i = \theta_{\sigma(i)}, \qquad \text{for} \quad i = 1, \dots, N. \tag{2.7}$$

By (**??**) and (**??**),

$$c_i = c_{\sigma(i)}, \qquad \text{for} \quad i = 1, \dots, N.$$

■

Lemma 2.10. *Let $\widetilde{\mathcal{H}}$ be the class of finite mixtures on $\mathcal{F}$. $\widetilde{\mathcal{H}}$ is identifiable if and only if*

$$\sum_{i=1}^{N} c_i F(\cdot, \theta_i) = \sum_{i=1}^{\widehat{N}} \widehat{c}_i F(\cdot, \widehat{\theta}_i) \quad \Longrightarrow \quad N = \widehat{N}, \quad \sum_{i=1}^{N} c_i \delta_{\theta_i} = \sum_{i=1}^{N} \widehat{c}_i \delta_{\widehat{\theta}_i}.$$

Proof :
To prove the lemma is enough to show that the necessary and sufficient condition for

$$\sum_{i=1}^{N} c_i \delta_{\theta_i} = \sum_{i=1}^{N} \widehat{c}_i \delta_{\widehat{\theta}_i}$$

where:

$$\begin{array}{ll} c_i, \widehat{c}_i > 0, \qquad i = 1, \dots, N, \qquad \sum_{i=1}^{N} c_i = 1, \qquad \sum_{i=1}^{N} \widehat{c}_i = 1 \\ \theta_i \quad \text{are distinct for} \quad i = 1, \dots, N \\ \widehat{\theta}_i \quad \text{are distinct for} \quad i = 1, \dots, \widehat{N}, \end{array}$$

is for each i, $i = 1, \dots, N$, there is j, $1 \le j \le N$ such that $c_i = \widehat{c}_j$ and $\theta_i = \widehat{\theta}_j$. The sufficient condition is obvious and the necessity follows from part (d) of Lemma **??**. ■

The following theorem is the first result concerning the sufficient conditions of the identifiability of finite mixtures.

Theorem 2.11 (Teicher [**?**])**.** *Let $\mathcal{F} = \{F(\cdot, \theta) : \theta \in \mathcal{B}\}$ be a family of one dimensional distribution functions indexed by a point θ in a Borel subset $\mathcal{B}$ of Euclidean m-space $\boldsymbol{R}^m$ such that $F(\cdot, \cdot)$ is measurable in $\boldsymbol{R} \times \mathcal{B}$. Suppose there exists a transform*

$$M : F \mapsto \phi,$$

where ϕ is a real valued functions defined on some S_ϕ, such that M is linear and injective. If there is a total ordering $(\preceq)$ of $\boldsymbol{\mathcal{F}}$ such that $F_1 \prec F_2$ implies

1. $S_{\phi_1} \subset S_{\phi_2}$,
2. *The existence of some $t_1 \in \bar{S}_{\phi_1}$ (t_1 being independent of ϕ_2) such that :*

$$\lim_{t \to t_1} \frac{\phi_2(t)}{\phi_1(t)} = 0,$$

then the class $\widetilde{\boldsymbol{\mathcal{H}}}$ of all finite mixtures on $\boldsymbol{\mathcal{F}}$ is identifiable.

Proof :
Suppose there are two finite sets of elements of $\boldsymbol{\mathcal{F}}$, say $\boldsymbol{\mathcal{F}}_1 = \{F(\cdot, \theta_i) : i = 1, \dots, N\}$ and $\boldsymbol{\mathcal{F}}_2 = \{F(\cdot, \widehat{\theta}_j) : j = 1, \dots, \widehat{N}\}$ such that

$$\sum_{i=1}^{N} c_i F(y, \theta_i) = \sum_{j=1}^{\widehat{N}} \widehat{c}_j F(y, \widehat{\theta}_j), \qquad \forall y \in \boldsymbol{R}, \tag{2.8}$$

where

$$0 < c_i, \widehat{c}_j \leq 1, \quad \sum_{i=1}^{N} c_i = 1, \quad \sum_{j=1}^{\widehat{N}} \widehat{c}_j = 1.$$

Without loss of generality, index $\boldsymbol{\mathcal{F}}_1$ and $\boldsymbol{\mathcal{F}}_2$ such that for $i < j$,

$$F(\cdot, \theta_i) \prec F(\cdot, \theta_j) \quad \text{and} \quad F(\cdot, \widehat{\theta}_i) \prec F(\cdot, \widehat{\theta}_j).$$

If $F(\cdot, \theta_1) \neq F(\cdot, \widehat{\theta}_1)$, suppose without loss of generality that

$$F(\cdot, \theta_1) \prec F(\cdot, \widehat{\theta}_1),$$

then

$$F(\cdot, \theta_1) \prec F(\cdot, \widehat{\theta}_j), \qquad j = 1, \dots, \widehat{N}.$$

Apply the transform to (**??**). Then for $t \in T_1 = S_{\phi_1} \cap \{t : \phi_1(t) \neq 0\}$,

$$\begin{aligned} \sum_{i=1}^{N} c_i \phi_i(t) &= \sum_{j=1}^{\widehat{N}} \widehat{c}_j \widehat{\phi}_j(t) \\ c_1 + \sum_{i=2}^{N} c_i \frac{\phi_i(t)}{\phi_1(t)} &= \sum_{j=1}^{\widehat{N}} \widehat{c}_j \frac{\widehat{\phi}_j(t)}{\phi_1(t)}. \end{aligned}$$

Letting $t \to t_1$, through values in T_1, we have $c_1 = 0$, contradicting (**??**) that $c_1 > 0$. Thus

$$F(\cdot, \theta_1) = F(\cdot, \widehat{\theta}_1)$$

and for any $t \in T_1$,

$$(c_1 - \widehat{c}_1) + \sum_{i=2}^{N} c_i \frac{\phi_i(t)}{\phi_1(t)} = \sum_{j=2}^{\widehat{N}} \widehat{c}_j \widehat{\frac{\phi_j(t)}{\phi_1(t)}}.$$

Again, letting $t \to t_1$, through values in T_1, we have $c_1 = \widehat{c}_1$. So now,

$$\sum_{i=2}^{N} c_i F(y, \theta_i) = \sum_{j=2}^{\widehat{N}} \widehat{c}_j F(y, \widehat{\theta}_j), \qquad \forall y \in \boldsymbol{R}.$$

Repeating the same argument $\min(N, \widehat{N})$ times, we have

$$F(\cdot, \theta_i) = F(\cdot, \widehat{\theta}_i) \quad \text{and} \quad c_i = \widehat{c}_i,$$

for $i = 1, 2, \ldots, \min(N, \widehat{N})$.

If $N \neq \widehat{N}$, without loss of generality assume $N > \widehat{N}$. Then

$$\sum_{i=\widehat{N}+1}^{N} c_i F(y, \theta_i) = 0 \qquad \forall y \in \boldsymbol{R}.$$

Letting $y \to \infty$ in the above equation, implies $c_i = 0$ for $i = \widehat{N}+1, \widehat{N}+2, \ldots, N$, in contradiction to (**??**). Therefore

$$N = \widehat{N}, \quad c_i = \widehat{c}_i \quad \text{and} \quad F(\cdot, \theta_i) = F(\cdot, \widehat{\theta}_i),$$

for $i = 1, 2, \ldots, N$. But $F(\cdot, \theta_i) = F(\cdot, \widehat{\theta}_i)$ imply $\theta_i = \widehat{\theta}_i$, for all $i = 1, 2, \ldots, N$. Then by definition $\widetilde{\boldsymbol{\mathcal{H}}}$ is identifiable. ■

An important application of Theorem **??** is the identifiability of the class of finite mixtures of one-dimensional normal distributions, the class of finite mixtures of one-dimensional gamma distributions and the class of finite mixtures of one-dimensional Poisson distributions.

Lemma 2.12 (Teicher [**?**])**.** *The class of all finite mixtures of one dimensional normal distributions is identifiable.*

Proof :
Let $\mathcal{N} = \{N(\cdot; \theta, \sigma^2) : \theta \in \boldsymbol{R}, \sigma > 0\}$ be a family of normal distributions, where $N(\cdot; \theta, \sigma^2)$ denotes a normal distribution with mean θ and variance σ^2.

Let $N(\cdot; \theta, \sigma^2) \in \mathcal{N}$ and define its (Laplace) transform by

$$\begin{aligned} \phi(t; \theta, \sigma^2) &= \int_{-\infty}^{\infty} N(y; \theta, \sigma^2)\, e^{-ty}\, dy \\ &= \int_{-\infty}^{\infty} \tfrac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{y-\theta}{\sigma}\right)^2}\, e^{-ty}\, dy \\ &= e^{\frac{1}{2}\sigma^2 t^2 - \theta t}, \end{aligned}$$

where $t \in S_\phi = (-\infty, \infty)$.

Order $\mathcal{N}$ by

$$N_1 = N(\cdot;\theta_1,\sigma_1^2) \prec N(\cdot;\theta_2,\sigma_2^2) = N_2$$

if $\quad \sigma_1 > \sigma_2 \quad$ or if $\quad \sigma_1 = \sigma_2, \quad$ but $\quad \theta_1 < \theta_2$.

Let $N_1 = N(\cdot;\theta_1,\sigma_1^2)$ and $N_2 = N(\cdot;\theta_2,\sigma_2^2)$ in $\mathcal{N}$ such that $N_1 \prec N_2$ and let $\phi_1(\cdot;\theta_1,\sigma_1^2)$ and $\phi_2(\cdot;\theta_2,\sigma_2^2)$ be their transform respectively. $S_{\phi_1} = S_{\phi_2} = (-\infty,\infty)$. Take $t_1 = \infty$. If $\sigma_1 > \sigma_2$, then

$$\lim_{t\to\infty} \frac{\phi_2(t)}{\phi_1(t)} = \lim_{t\to\infty} e^{\left\{\frac{(\sigma_2^2-\sigma_1^2)t^2}{2} - (\theta_2-\theta_1)t\right\}} = 0,$$

since

$$\lim_{t\to\infty} e^{\left\{\frac{(\sigma_2^2-\sigma_1^2)t^2}{2}\right\}} = 0.$$

If $\sigma_1 = \sigma_2$ and $\theta_1 < \theta_2$, then

$$\lim_{t\to\infty} \frac{\phi_2(t)}{\phi_1(t)} = \lim_{t\to\infty} e^{\{-(\theta_2-\theta_1)t\}} = 0.$$

Then the identifiability of the class of finite mixtures of $\mathcal{N}$ follows from Theorem **??**. ■

Lemma 2.13 (Teicher [**?**]). *The class of all finite mixtures of gamma distributions is identifiable*

Proof :
Let $\boldsymbol{\mathcal{F}} = \{F(\cdot;\theta,\alpha) : \theta > 0, \alpha > 0\}$ be a family of gamma distrsibutions, where

$$F(y;\theta,\alpha) = \tfrac{\theta^\alpha}{\Gamma(\alpha)} \int_0^y x^{\alpha-1} e^{-\theta x}\, dx\,, \quad \alpha > 0, \quad \theta > 0.$$

Let $F(\cdot;\theta,\alpha) \in \boldsymbol{\mathcal{F}}$, define its (Laplace) transform as follows.

$$\begin{aligned} \phi(t;\theta,\alpha) &= \int_0^\infty \tfrac{\theta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\theta x} e^{-tx}\, dx \\ &= \tfrac{\theta^\alpha}{\Gamma(\alpha)} \cdot \tfrac{\Gamma(\alpha)}{(\theta+t)^\alpha} \\ &= \left(1 + \tfrac{t}{\theta}\right)^{-\alpha}, \qquad \text{for} \quad t > -\theta. \end{aligned}$$

Order $\boldsymbol{\mathcal{F}}$ by

$$F_1 = F(\cdot;\theta_1,\alpha_1) \prec F(\cdot;\theta_2,\alpha_2) = F_2$$

if $\theta_1 < \theta_2 \quad$ or $\quad \theta_1 = \theta_2 \quad$ but $\quad \alpha_1 > \alpha_2$.

Let $F_1 = F(\cdot;\theta_1,\alpha_1)$ and $F_2 = F(\cdot;\theta_2,\alpha_2)$ be any elements of $\mathcal{N}$, such that $F_1 \prec F_2$ and let $\phi_1(\cdot;\theta_1,\alpha_1)$ and $\phi_2(\cdot;\theta_2,\alpha_2)$ be their transform respectively. Then $S_{\phi_1} = (-\theta_1, \infty) \subset S_{\phi_2} = (-\theta_2, \infty)$. If $\theta_1 < \theta_2$, then

$$\lim_{t \to -\theta_1} \frac{\phi_2(t;\theta_2,\alpha_2)}{\phi_1(t;\theta_1,\alpha_1)} = \lim_{t \to -\theta_1} \frac{\left(1+\frac{t}{\theta_2}\right)^{-\alpha_2}}{\left(1+\frac{t}{\theta_1}\right)^{-\alpha_1}} = 0,$$

since $\lim_{t \to -\theta_1} \left(1+\frac{t}{\theta_1}\right)^{-\alpha_1} = \infty$. If $\theta_1 = \theta_2$, but $\alpha_1 > \alpha_2$, then

$$\lim_{t \to -\theta_1} \frac{\phi_2(t;\theta_2,\alpha_2)}{\phi_1(t;\theta_1,\alpha_1)} = \lim_{t \to -\theta_1} \frac{\left(1+\frac{t}{\theta_2}\right)^{-\alpha_2}}{\left(1+\frac{t}{\theta_1}\right)^{-\alpha_1}} = \lim_{t \to -\theta_1} \left(1+\frac{t}{\theta_1}\right)^{\alpha_1-\alpha_2} = 0,$$

since $\alpha_1 - \alpha_2 > 0$. Then by Theorem **??**, the class of all finite mixtures of gamma distributions is identifiable. ■

Corollary 2.14. *The class of all finite mixtures of negative exponential distribution is identifiable*

Proof :
Let $\boldsymbol{\mathcal{F}} = \{F(\cdot,\theta) : \theta > 0\}$ be the family of exponential distributions, where

$$F(y,\theta) = \int_0^y \theta e^{-\theta x}\, dx\,, \qquad \theta > 0.$$

It can be seen that $\boldsymbol{\mathcal{F}}$ is a special case of the family of gamma distributions in Lemma **??** with $\alpha = 1$, then the result follows. ■

Lemma 2.15. *The class of all finite mixtures of Poisson distributions is identifiable.*

Proof :
Let $\boldsymbol{\mathcal{F}} = \{f(\cdot,\theta) : \theta > 0\}$ be the family of Poisson distribution with mean θ, where

$$f(y,\theta) = \frac{e^{-\theta}\theta^y}{y!}\,, \qquad y = 0,1,2,\ldots \quad \text{and} \quad \theta > 0.$$

For $f(\cdot,\theta) \in \boldsymbol{\mathcal{F}}$, define its transform as follows

$$\begin{aligned}
\phi(t,\theta) &= \sum_{y=0}^{\infty} e^{ty} \cdot \frac{e^{-\theta}\theta^y}{y!} \\
&= e^{-\theta}\left(\sum_{y=0}^{\infty} \frac{\theta^y}{y!}\, e^{ty}\right) \\
&= e^{-\theta}\left(\sum_{y=0}^{\infty} \frac{(e^t\theta)^y}{y!}\right) \\
&= e^{e^t\theta} \cdot e^{-\theta} \\
&= e^{\theta(e^t-1)}\,, \qquad t \in \boldsymbol{R}.
\end{aligned}$$

Order $\boldsymbol{\mathcal{F}}$ by

$$f_1 = f(\cdot, \theta_1) \prec f(\cdot, \theta_2) = f_2 \qquad \text{if} \qquad \theta_1 > \theta_2.$$

Let $f_1 = f(\cdot, \theta_1)$ and $f_2 = f(\cdot, \theta_2)$ be in $\boldsymbol{\mathcal{F}}$, such that $f_1 \prec f_2$ and let $\phi_1(\cdot, \theta_1)$ and $\phi_2(\cdot, \theta_2)$ be their transforms respectively. Then $S_{\phi_1} = S_{\phi_2} = (-\infty, \infty)$ and

$$\begin{aligned} \lim_{t\to\infty} \frac{\phi_2(t)}{\phi_1(t)} &= \lim_{t\to\infty} \frac{e^{\theta_2(e^t-1)}}{e^{\theta_1(e^t-1)}} \\ &= \lim_{t\to\infty} e^{\theta_2(e^t-1)-\theta_1(e^t-1)} \\ &= \lim_{t\to\infty} e^{(\theta_2-\theta_1)(e^t-1)} \\ &= 0\,, \end{aligned}$$

since $\theta_2 - \theta_1 < 0$. Then the result follows from Theorem **??**. ■

Yakowitz and Spragins [**?**] extended Teicher's results of identifiability to include multidimensional distribution functions. Let

$$\boldsymbol{\mathcal{F}}_n = \{F(\cdot, \theta) : \theta \in \mathcal{B}\}$$

be a family of n-dimensional distribution functions taking values in $\boldsymbol{R}^n$ indexed by a point θ in a borel subset $\mathcal{B}$ of Euclidean m-space $\boldsymbol{R}^m$, such that $F(\cdot, \cdot)$ is measurable in $\boldsymbol{R}^n \times \mathcal{B}$.

Let $\widetilde{\boldsymbol{\mathcal{H}}}_n$ be the class of all finite mixtures on $\boldsymbol{\mathcal{F}}_n$ defined as in Definition **??**, that is,

$$\begin{aligned} \widetilde{\boldsymbol{\mathcal{H}}}_n = \Big\{ & H(\cdot) : H(\cdot) = \sum_{i=1}^{N} c_i F(\cdot, \theta_i), c_i > 0, \sum_{i=1}^{N} c_i = 1, F(\cdot, \theta_i) \in \boldsymbol{\mathcal{F}}_n, \\ & N = 1, 2 \ldots \Big\}. \end{aligned}$$

As in one dimensional case, for every finite mixture

$$H(\cdot) = \sum_{i=1}^{N} c_i F(\cdot, \theta_i),$$

$\theta_1, \ldots, \theta_N$ are assumed to be *distinct*.

Theorem 2.16 (Yakowitz and Spragins [**?**]). *A necessary and sufficient condition that the class $\widetilde{\boldsymbol{\mathcal{H}}}_n$ of all finite mixtures on $\boldsymbol{\mathcal{F}}_n$ be identifiable is that $\boldsymbol{\mathcal{F}}_n$ be a linearly independent set over the field of real numbers.*

Proof :
Necessity:
Suppose $\boldsymbol{\mathcal{F}}_n$ is not a linearly independent set over the field of real numbers. Let

$$\sum_{i=1}^{N} a_i F(y, \theta_i) = 0, \qquad \forall y \in \boldsymbol{R}^n,$$

where $a_i \in \boldsymbol{R}$, $i = 1, 2, \dots, N$, be a linear relation in $\boldsymbol{\mathcal{F}}_n$.
Assume the a_i's are subscripted so that

$$a_i < 0 \quad \Longleftrightarrow \quad i \leq M\,.$$

Then

$$\sum_{i=1}^{M} |a_i| F(y, \theta_i) = \sum_{i=M+1}^{N} |a_i| F(y, \theta_i) \qquad \forall y \in \boldsymbol{R}^n\,. \tag{2.9}$$

By letting $y \to \boldsymbol{\infty}$ in (**??**), where $\boldsymbol{\infty} = (\infty, \infty, \dots, \infty)$,

$$\sum_{i=1}^{M} |a_i| = \sum_{i=M+1}^{N} |a_i|. \tag{2.10}$$

Let

$$b = \sum_{i=1}^{M} |a_i| \qquad \text{and} \qquad c_i = \frac{|a_i|}{b}, \quad i = 1, \dots, N. \tag{2.11}$$

By (**??**) and (**??**),

$$\begin{aligned} &b > 0 \\ &c_i > 0, \quad i = 1, \dots, M, \quad \sum_{i=1}^{M} c_i = 1 \\ &c_i \geq 0, \quad i = M+1, \dots, N, \quad \sum_{i=M+1}^{N} c_i = 1. \end{aligned}$$

Then

$$\sum_{i=1}^{M} c_i F(\cdot, \theta_i) = \sum_{i=M+1}^{N} c_i F(\cdot, \theta_i)$$

are two distinct representations of the same finite mixture and therefore $\widetilde{\boldsymbol{\mathcal{H}}}_n$ can not be identifiable.

Sufficiency :
Let $\langle \boldsymbol{\mathcal{F}}_n \rangle$ be the span of $\boldsymbol{\mathcal{F}}_n$. If $\boldsymbol{\mathcal{F}}_n$ is linearly independent, then it is a bases for $\langle \boldsymbol{\mathcal{F}}_n \rangle$. Two distinct representations of the same mixture implied by the non-identifiability of $\widetilde{\boldsymbol{\mathcal{H}}}_n \subset \langle \boldsymbol{\mathcal{F}}_n \rangle$ would contradict the uniqueness of representation of bases. ∎

From the properties of isomorphisms, $\boldsymbol{\mathcal{F}}_n$ is linearly independent if and only if the image of the isomorphism is linearly independent in the image space, the corollary below follows.

Corollary 2.17. *The class $\widetilde{\boldsymbol{\mathcal{H}}}_n$ of all finite mixtures of the family $\boldsymbol{\mathcal{F}}_n$ is identifiable if and only if the image $\boldsymbol{\mathcal{F}}_n$ under any vector isomorphism of $\langle \boldsymbol{\mathcal{F}}_n \rangle$ be linearly independent in the image space.*

The most important result of the application of Theorem **??** is the identifiability of the family of finite mixtures of multidimensional normal distributions.

Lemma 2.18 (Yakowitz and Spragins [**?**]). *The family of n dimensional normal distribution functions generates identifiable finite mixtures.*

Proof :
Let

$\mathcal{N} = \{N(\cdot;\theta,\Lambda) : \theta \in \boldsymbol{R}^n \text{ and } \Lambda \text{ is an } n\times n \text{ positive definite matrix }\}$

be a family of n-dimensional normal distribution with mean vector θ and covariance matrix Λ.

For $N(\cdot;\theta,\Lambda) \in \mathcal{N}$, let $M(\cdot;\theta,\Lambda)$ be its moment generating function defined by

$$\begin{aligned} M(t;\theta,\Lambda) &= \int_{\boldsymbol{R}^n} \exp\{-t^T y\} N(y;\theta,\Lambda)\, dy \\ &= \exp\left\{\theta^T t + \tfrac{1}{2} t^T \Lambda t\right\}, \qquad t \in \boldsymbol{R}^n. \end{aligned}$$

Note that θ, t and y are n-dimensional column vectors. It is clear that the mapping $N \mapsto M$ is an isomorphism.

Suppose that $\mathcal{N}$ does not generate identifiable finite mixtures. Then by Corollary **??**, the set

$\mathcal{M} = \{M(\cdot;\theta,\Lambda) : \theta \in \boldsymbol{R}^n \text{ and } \Lambda \text{ is an } n\times n \text{ positive definite matrix }\}$

is a linearly dependent set over $\boldsymbol{R}$. There are $M \geq 1,\ d_i \in \boldsymbol{R},\ d_i \neq 0,\ i = 1,\ldots,M$ and distinct pairs $(\theta_i, \Lambda_i),\ i = 1,\ldots,M$ such that

$$\sum_{i=1}^{M} d_i \exp\left\{\theta_i^T t + \tfrac{1}{2} t^T \Lambda_i t\right\} = 0, \qquad t \in \boldsymbol{R}^n. \tag{2.12}$$

Consider a special case of (**??**), when $t = \alpha s$, for a fixed vector s and $\alpha \in \boldsymbol{R}$. Then (**??**) becomes

$$\sum_{i=1}^{M} d_i \exp\left\{\alpha(\theta_i^T s) + \tfrac{1}{2}\alpha^2 (s^T \Lambda_i\, s)\right\} = 0, \quad \alpha \in \boldsymbol{R}. \tag{2.13}$$

If all $\theta_i,\ i = 1,\ldots,M$ are identical, then all $\Lambda_i,\ i = 1,\ldots,M$ are distinct. For $i \neq j$, $1 \leq i,j \leq M$,

$$s^T \Lambda_i\, s = s^T \Lambda_j\, s \quad \Longleftrightarrow \quad s \in \left\{z : z^T(\Lambda_i - \Lambda_j) z = 0\right\}.$$

So if

$$s \notin \bigcup_{i\neq j,\ 1\leq i,j\leq M} \left\{z : z^T(\Lambda_i - \Lambda_j) z = 0\right\},$$

then $s^T\Lambda_i\, s$, for $i = 1,\dots,M$ are all distinct positive real numbers, implying the pairs of real numbers $(\theta_i^T s, s^t\Lambda_i\, s)$, $i = 1,\dots,M$ are distinct.

Otherwise, Suppose without loss of generality that $\theta_1,\dots,\theta_k$, for some k, $k < M$, are the only distinct vectors among $\theta_1,\dots,\theta_M$. Then for $i \neq j$, $1 \le i,j \le k$,

$$\theta_i^T s = \theta_j^T s \quad\iff\quad s \in \left\{z : (\theta_i^T - \theta_j^T)z = 0\right\}.$$

So if

$$s \notin \bigcup_{i\neq j,\ 1\le i,j\le k} \left\{z : (\theta_i^T - \theta_j^T)z = 0\right\},$$

then the real numbers $\theta_i^T s,\ i = 1,\dots,k$ are distinct. Since the $(\theta_i,\Lambda_i),\ i = 1,\dots,M$ are all distinct, then the $\Lambda_i,\ i = k+1,\dots,M$ with the same θ_i, are different. So if

$$s \notin \bigcup_{i\neq j,\ k+1\le i,j\le M} \left\{z : z^T(\Lambda_i - \Lambda_j)z = 0\right\},$$

then the real numbers $s^T\Lambda_i\, s,\ i = k+1,\dots,M$ are distinct. Consequently, for

$$\begin{aligned} s \ \notin & \bigcup_{i\neq j,\ 1\le i,j\le k} \left\{z : (\theta_i^T - \theta_j^T)z = 0\right\} \\ & \bigcup \bigcup_{i\neq j,\ k+1\le i,j\le M} \left\{z : z^T(\Lambda_i - \Lambda_j)z = 0\right\}, \end{aligned} \tag{2.14}$$

the pairs of real numbers $(\theta_i^T s, s^T\Lambda_i\, s),\ i = 1,\dots,M$ are distinct.

Therefore, for such a choice of s, the equation (**??**) asserts that there is $M \ge 1,\ d_i \in \boldsymbol{R},\ d_i \neq 0,\ i = 1,\dots,M$ and distinct pairs (μ_i,σ_i^2), where

$$\mu_i = \theta_i^T s \quad \text{and} \quad \sigma_i^2 = s^T\Lambda_i\, s, \quad i = 1,\dots,M,$$

such that

$$\sum_{i=1}^{M} d_i \exp\left\{\mu_i\alpha + \tfrac{1}{2}\sigma_i^2\alpha^2\right\} = 0, \quad \alpha \in \boldsymbol{R}.$$

Corollary **??** implies that the class of finite mixtures of one dimensional normal distributions is not identifiable, contrary to Lemma **??**. ■

Teicher's result which is concerned with mixtures of product measures will be presented next. Teicher [**?**] stated that the identifiability of mixture distributions can be carried over to mixtures of product distributions.

Recall that for any $k \in \boldsymbol{N}$, we have defined

$$\boldsymbol{\mathcal{F}}_k = \{F(\cdot,\alpha) : \alpha \in \mathcal{B}\},$$

as a family of k-dimensional distribution functions indexed by a point α in a Borel subset $\mathcal{B}$ of Euclidean m-space $\boldsymbol{R}^m$, such that $F(\cdot,\cdot)$ is measurable in $\boldsymbol{R}^k \times \mathcal{B}$.

Define for every $k, n \in \boldsymbol{N}$,

$$\boldsymbol{\mathcal{F}}^*_{k,n} = \left\{ F^*(\cdot,\alpha) : F^*(\cdot,\alpha) = \prod_{i=1}^{n} F(\cdot,\alpha_i),\ F(\cdot,\alpha_i) \in \boldsymbol{\mathcal{F}}_k, 1 \le i \le n \right\}. \tag{2.15}$$

Notice that in (**??**), $F(\cdot,\cdot)$ is defined on $\boldsymbol{R}^k \times \mathcal{B}$ and $F^*(\cdot,\cdot)$ is defined on $\boldsymbol{R}^{kn} \times \mathcal{B}^n$.

Theorem 2.19 (Teicher [**?**]). *If the class of all mixtures on $\boldsymbol{\mathcal{F}}_1$ is identifiable, then for every $n > 1$, the class of mixtures on $\boldsymbol{\mathcal{F}}^*_{1,n}$ is identifiable. Conversely, if for some $n > 1$, the class of all mixtures on $\boldsymbol{\mathcal{F}}^*_{1,n}$ is identifiable, then the class of mixtures on $\boldsymbol{\mathcal{F}}_1$ is identifiable.*

Proof :
To prove the second part, suppose that the class of all mixtures on $\boldsymbol{\mathcal{F}}^*_{1,n}$ is identifiable for some $n > 1$. Let $F(\cdot,\alpha) \in \boldsymbol{\mathcal{F}}_1$. If

$$\int_{\mathcal{B}} F(y,\alpha)\, dG(\alpha) = \int_{\mathcal{B}} F(y,\alpha)\, d\widehat{G}(\alpha),$$

then multiplying both sides by $\displaystyle\prod_{i=1}^{n-1} F(y_i,\alpha_o)$, $\alpha_o \in \mathcal{B}$, necessarily,

$$I_{\alpha_o} \times \cdots \times I_{\alpha_o} \times G = I_{\alpha_o} \times \cdots \times I_{\alpha_o} \times \widehat{G},$$

where I_{α_o} is a characteristic function $\chi_{[\alpha_o,\infty)}$. Hence by the hypothesis $G = \widehat{G}$.

To prove the first part of the theorem, the mathematical induction will be used. Suppose the class of mixtures on $\boldsymbol{\mathcal{F}}_1$ is identifiable and also suppose the class of mixtures on $\boldsymbol{\mathcal{F}}^*_{1,n}$ is identifiable for fixed but arbitrary n. It will be shown that the class of mixtures on $\boldsymbol{\mathcal{F}}^*_{1,(n+1)}$ is also identifiable.

Suppose that for $F^* \in \boldsymbol{\mathcal{F}}_{1,n}$ and $F \in \boldsymbol{\mathcal{F}}_1$,

$$\int F^*(x,\alpha) F(y,\beta)\, dG(\alpha,\beta) = \int F^*(x,\alpha) F(y,\beta)\, d\widehat{G}(\alpha,\beta)\,. \tag{2.16}$$

Let $G_2(\beta)$ and $\widehat{G}_2(\beta)$ denote the marginal distribution of β corresponding to G and $\widehat{G}$. Let $G(\alpha|\beta)$, $\widehat{G}(\alpha|\beta)$ denote versions of the conditional probabilities, such that, for each β, $G(\alpha|\beta)$ and $\widehat{G}(\alpha|\beta)$ are distribution functions in the variable α and for each α, $G(\alpha|\beta)$ and $\widehat{G}(\alpha|\beta)$ are equal almost every where to measurable functions of β. Then (**??**) may be rewritten as,

$$\int F(y,\beta)H(x,\beta)\,dG_2(\beta)=\int F(y,\beta)\widehat{H}(x,\beta)\,d\widehat{G}_2(\beta)\,, \tag{2.17}$$

where

$$H(x,\beta) = \int F^*(x,\alpha)\,d_\alpha G(\alpha|\beta) \tag{2.18}$$

$$\widehat{H}(x,\beta) = \int F^*(x,\alpha)\,d_\alpha \widehat{G}(\alpha|\beta). \tag{2.19}$$

In turn, (**??**) may be expressed as

$$\int F(y,\beta)\,dJ_x(\beta)=\int F(y,\beta)\,d\widehat{J}_x(\beta), \tag{2.20}$$

where

$$J_x(\beta) = \int_{-\infty}^{\beta} H(x,\gamma)\,dG_2(\gamma)\le G_2(\beta) \tag{2.21}$$

$$\widehat{J}_x(\beta) = \int_{-\infty}^{\beta} \widehat{H}(x,\gamma)\,d\widehat{G}_2(\gamma)\le \widehat{G}_2(\beta), \tag{2.22}$$

as $H(x,\gamma)\le 1$ and $\widehat{H}(x,\gamma)\le 1$. Dominated convergence applied to (**??**), to ensure that

$$J_x(\infty)=\widehat{J}_x(\infty),$$

since this common value is finite by (**??**) and (**??**). Thus from (**??**) and since the class of mixture on $\boldsymbol{\mathcal{F}}_1$ is identifiable by the hypothesis, then

$$J_x=\widehat{J}_x.$$

Or equivalently from (**??**) and (**??**),

$$\int_{-\infty}^{\beta} H(x,\gamma)\,dG_2(\gamma)=\int_{-\infty}^{\beta}\widehat{H}(x,\gamma)\,d\widehat{G}_2(\gamma). \tag{2.23}$$

On the other hand, letting $x\to\infty$ in (**??**) and since

$$\lim_{x\to\infty} H(x,\beta) = \lim_{x\to\infty}\int F^*(x,\alpha)\,d_\alpha G(\alpha|\beta)=1 \tag{2.24}$$

$$\lim_{x\to\infty} \widehat{H}(x,\beta) = \lim_{x\to\infty}\int F^*(x,\alpha)\,d_\alpha \widehat{G}(\alpha|\beta)=1 \tag{2.25}$$

by monotone convergence theorem, then (**??**) gives

$$\int F(y,\beta)\,dG_2(\beta)=\int F(y,\beta)\,d\widehat{G}_2(\beta). \tag{2.26}$$

By the hypothesis,

$$G_2(\beta)=\widehat{G}_2(\beta). \tag{2.27}$$

However, (**??**) in conjunction with (**??**) necessitates

$$H(x,\beta)=\widehat{H}(x,\beta)\,, \tag{2.28}$$

for almost all β. Equation (**??**) together with (**??**) and (**??**), gives

$$\int F^*(x,\alpha)\, d_\alpha G(\alpha|\beta) = \int F^*(x,\alpha)\, d_\alpha \widehat{G}(\alpha|\beta). \tag{2.29}$$

By the induction hypothesis, that is, the class of mixtures on $\boldsymbol{\mathcal{F}}^*_{1,n}$ is identifiable and (**??**) imply

$$G(\alpha|\beta) = \widehat{G}(\alpha|\beta) \tag{2.30}$$

Finally, combining (**??**) and (**??**) we have

$$G(\alpha,\beta) = G(\alpha|\beta)\, G(\beta) = \widehat{G}(\alpha|\beta)\, \widehat{G}(\beta) = \widehat{G}(\alpha,\beta).$$

So that the class of mixtures on $\boldsymbol{\mathcal{F}}^*_{1,(n+1)}$ is identifiable. ■

Since Theorem **??** applies for general mixtures, then we have the following theorem for finite mixtures.

Theorem 2.20. *If the class of all finite mixtures on $\boldsymbol{\mathcal{F}}_1$ is identifiable, then for every $n > 1$, the class of finite mixtures on $\boldsymbol{\mathcal{F}}^*_{1,n}$ is identifiable. Conversely, if for some $n > 1$, the class of all finite mixtures on $\boldsymbol{\mathcal{F}}^*_{1,n}$ is identifiable, then the class of finite mixtures on $\boldsymbol{\mathcal{F}}_1$ is identifiable.*

Analogous result hold with $\boldsymbol{\mathcal{F}}_1$ and $\boldsymbol{\mathcal{F}}^*_{1,n}$ is replaced by $\boldsymbol{\mathcal{F}}_k$ and $\boldsymbol{\mathcal{F}}^*_{k,n}$, where $k > 1$.

References

[1] Everitt B.S and Hand D.J. 1981. *Finite Mixture Distributions*. Chapman and Hall, London.

[2] Mc. Lachlan G.J. and Basford K.E. 1988. *Mixture Models*. Marcell Dekker, New York.

[3] Teicher, H. 1963. Identifiability of finite mixtures. *Ann. Math. Statist.*, 34: 1265-1269.

[4] Teicher, H. 1967. Identifiability of mixtures of product measures. *Ann. Math. Statist.*, 38: 1300-1302.

[5] Titterington D.M, Smith A.F.M. and Makov V.E. 1965 *Statistical Analysis of Finite Mixture Distributions*. John Wiley, New York.

[6] Yakowitz, S.J. and Spragins J.D. 1968. On the identifiability of finite mixtures *Ann. Math. Statist.*, 39(1): 209-214.