The Annals of Mathematical Statistics
1968, Vol. 39, No. 1, 209–214

ON THE IDENTIFIABILITY OF FINITE MIXTURES

BY SIDNEY J. YAKOWITZ[1] AND JOHN D. SPRAGINS[2]

Arizona State University

1. Summary. H. Teicher [5] has initiated a valuable study of the identifiability of finite mixtures (these terms to be defined in the next section), revealing a sufficiency condition that a class of finite mixtures be identifiable and from this, establishing the identifiability of all finite mixtures of one-dimensional Gaussian distributions and all finite mixtures of gamma distributions. From other considerations, he has generalized [4] a result of Feller [1] that arbitrary (and hence finite) mixtures of Poisson distributions are identifiable, and has also shown binomial and uniform families do not generate identifiable mixtures. In this paper it is proven that a family $\mathfrak{F}$ of cumulative distribution functions (cdf's) induces identifiable finite mixtures if and only if $\mathfrak{F}$ is linearly independent in its span over the field of real numbers. Also we demonstrate that finite mixtures of $\mathfrak{F}$ are identifiable if $\mathfrak{F}$ is any of the following: the family of n products of exponential distributions, the multivariate Gaussian family, the union of the last two families, the family of one-dimensional Cauchy distributions, and the non-degenerate members of the family of one-dimensional negative binomial distributions. Finally it is shown that the translation-parameter family generated by any one-dimensional cdf yields identifiable finite mixtures.

2. Introduction. We make an easy modification of Teicher's definition of identifiability to include multidimensional cdf's. Let

$$\mathfrak{F} = \{F(x; \alpha); \alpha \in R_1^m, x \in R^n\} \tag{1}$$

constitute a family of n-dimensional cdf's indexed by a point α in a Borel subset R_1^m of Euclidean m space R^m such that $F(x, \alpha)$ is measureable in $R^n \times R_1^m$. Then the n-dimensional cdf $H(x) = \int_{R_1^m} F(x, \alpha)\, dG(\alpha)$ is the image of the above mapping, say Q, of the m-dimensional cdf G (where the measure μ_G induced by G assigns measure 1 to R_1^m). The distribution H is called the mixture (or G-mixture) of $\mathfrak{F}$ and G the mixing distribution. Let $\mathcal{G}$ denote the class of all such m-dimensional cdf's G and $\mathcal{H}$ the induced class of mixtures H. Then $\mathcal{H}$ will be said to be identifiable if Q is a one-to-one map of $\mathcal{G}$ onto $\mathcal{H}$. H is called a finite mixture if its mixing distribution or rather the corresponding measure μ_G is discrete and doles out positive mass to only a finite number of points in R_1^m. Thus the set $\mathcal{H}$ of all finite mixtures of a class $\mathfrak{F}$ of distributions is the convex hull of $\mathfrak{F}$:

$$\mathcal{H} = \{H(x) : H(x) = \textstyle\sum_{i=1}^N c_i F(x, \alpha_i),\ c_i > 0,\ \sum_{i=1}^N c_i = 1,\ F(x, \alpha_i) \in \mathfrak{F},\ N = 1, 2, \cdots\}. \tag{2}$$

Received 11 August 1967.

[1] Now with the University of Arizona. While undertaking this study, he was recipient of an NSF traineeship.

[2] Now with General Electric, Phoenix.

The α_i's are presumed distinct and the subscripting is not meant to imply anything about the cardinality of $\mathfrak{F}$. In the context of finite mixtures, the definition of "identifiable" implies $\mathfrak{F}$ generates identifiable finite mixtures if and only if the convex hull of $\mathfrak{F}$ has the uniqueness of representation property:

$$\sum_{i=1}^{N} c_i F_i = \sum_{i=1}^{M} c_i' F_i' \tag{3}$$

implies $N = M$ and for each i, $1 \leqq i \leqq N$ there is some j, $1 \leqq j \leqq N$, such that $c_i = c_j'$ and $F_i = F_j'$. Here and hereafter, we write F_i, meaning $F(x; \alpha_i)$.

3. Characterizations of identifiability.

THEOREM. *A necessary and sufficient condition that the class $\mathfrak{H}$ of all finite mixtures of the family $\mathfrak{F}$ of (1) be identifiable is that $\mathfrak{F}$ be a linearly independent set over the field of real numbers.*

It is intended here that the vector operations should be the usual function addition and scalar multiplication. $\langle A \rangle$ will denote the span of A over the real numbers.

NECESSITY. Let $\sum_{i=1}^{N} a_i F_i = 0$, $a_i \varepsilon R$, be a linear relation in $\mathfrak{F}$. Assume the a_i's are subscripted so that $a_i < 0 \Leftrightarrow i \leqq M$. Then $\sum_{i=1}^{M} |a_i| F_i = \sum_{i=M+1}^{N} |a_i| F_i$. Since the F_i's are cdf's, if $\underset{\sim}{\infty}$ denotes the n-tuple $(\infty, \infty, \cdots \infty)$, $F(\underset{\sim}{\infty}) = 1$, and therefore $\sum_{i=1}^{M} |a_i| F_i(\underset{\sim}{\infty}) = \sum_{i=1}^{M} |a_i| = \sum_{i=M+1}^{N} |a_i| \equiv b > 0$. If $c_i \equiv |a_i|/b$, then $\sum_{i=1}^{M} c_i F_i = \sum_{i=M+1}^{N} c_i F_i$ are two distinct representations of the same finite mixture and therefore $\mathfrak{H}$ cannot be identifiable.

SUFFICIENCY. If $\mathfrak{F}$ is linearly independent, then it is a basis for $\langle \mathfrak{F} \rangle$. Two distinct representations of the same mixture, implied by the non-identifiability of $\mathfrak{H} \subset \langle \mathfrak{F} \rangle$, would contradict the uniqueness of representation property of bases.

COROLLARY. *A necessary and sufficient condition that the class $\mathfrak{H}$ of all finite mixtures of the family $\mathfrak{F}$ of (1) be identifiable is that the image of $\mathfrak{F}$ under any vector isomorphism on $\langle \mathfrak{F} \rangle$ be linearly independent in the image space.*

PROOF. By elementary properties of isomorphisms, $\mathfrak{F}$ is linearly independent if and only if the image is linearly independent in the image space.

Our theorem may be regarded as an extension of Theorem 1 of [5] from finite to arbitrary families $\mathfrak{F}$. The sufficiency condition of Theorem 2 of [5], which gives conditions assuring that $\mathfrak{F}$ is linearly independent, provides a useful method for establishing the identifiability of a proposed mixture class. Our results in Section 4 on families which yield identifiable mixtures can be proven independently of our characterization theorem from [5].

4. Some families which generate identifiable mixtures.

PROPOSITION 1. *If n is a positive integer and $\mathfrak{F}$ is the family of products of n exponential* cdf's, *then the class of all finite mixtures of $\mathfrak{F}$ is identifiable.*

PROOF. *Let* $\bar{a} = (a_1, a_2, \cdots a_n)$, $a_j > 0$, and $\bar{x} = (x_1, x_2, \cdots x_n)$, x_j, a real variable.

Let $K(\bar{a})$ be the product $\prod_{i=1}^{n} a_i$. A typical element of the densities of $\mathfrak{F}$ is

$$\begin{aligned} f(\bar{x}; \bar{a}) &= K(\bar{a}) \exp(-\bar{a} \circ \bar{x}), \quad \text{if} \quad x_i > 0, i = 1, 2, \cdots n \\ &= 0, \quad \text{otherwise}, \end{aligned} \tag{4}$$

$\bar{a} \circ \bar{x}$ is the inner product of $\bar{a}$ and $\bar{x}$. Assume

$$0 = \sum_{i=1}^{N} c_i f(\bar{x}; \overline{a(i)}), \qquad \overline{a(i)} \; \varepsilon \; R^n \tag{5}$$

is a linear relation in $\mathfrak{F}$. $\overline{a(i)} \circ (\bar{x})$ may be regarded as a linear functional on the variable $\bar{x}$.

$$[\overline{a(1) - a(j)}] \circ (\bar{x}) = \sum_{i=1}^{n} (a(1)_i - a(j)_i) x_i$$

is also a non-zero linear functional if $j \neq 1$. As the kernel of a non-zero linear functional is a hyperplane [4, p. 120], there is some point $\bar{u} \; \varepsilon \; R^n$, $u_j > 0$, $j = 1, \cdots n$, such that $0 < \overline{a(i)} \circ \bar{u} \equiv \xi_i$ and $\xi_i \neq \xi_1 > 0$, $i = 2, \cdots N$. Thus for all vectors $b\bar{u}$, $b > 0$ (5) gives

$$\sum_{i=1}^{N} K(\bar{a}(i)) c_i \exp(-b\xi_i) = 0 \tag{6}$$

where $\xi_i \neq \xi_1$ if $i \neq 1$.

A result in [5] is that the entire univariate gamma family generates identifiable finite mixtures. By our theorem, this implies that the gamma family is linearly independent. As $\xi_i \neq \xi_1$, $i > 1$, (6) would contradict the linear independence of the gamma family, which subsumes the exponential family, if $c_1 \neq 0$. Continuing in this fashion, the relation (5) is shown to be trivial. Thus $\mathfrak{F}$ is linearly independent and therefore generates identifiable mixtures.

PROPOSITION 2. *The family $\mathfrak{F}$ of n-dimensional Gaussian* cdf's *generates identifiable finite mixtures.*

PROOF. (This proof, supplied by our referee, is considerably shorter and more elegant than our own.)

Suppose that $\mathfrak{F}$ is not identifiable. Denoting mean vectors by θ and covariance matrices by Λ this implies, in terms of moment generating functions, that for some $M \geqq 1$,

$$\sum_{j=1}^{M} d_j \exp\{\tfrac{1}{2} T' \Lambda_j T + T' \theta_j\} \equiv 0 \tag{7}$$

where the pairs (Δ_j, θ_j) are all distinct. Setting $T = cu$ where c is a scalar and u is a vector, (7) becomes

$$\sum_{j=1}^{M} d_j \exp\{(c^2/2) u' \Lambda_j u + c u' \theta_j\} = 0. \tag{8}$$

If all Δ_j, $1 \leqq j \leqq M$ are identical, all θ_j, $1 \leqq j \leqq M$ are distinct whence for u outside a finite number of hyperplanes, the pairs of real numbers $(u'\Delta_j u, u'\theta_j)$, $1 \leqq j \leqq M$ are distinct. Otherwise suppose without loss of generality that $\Lambda_1, \cdots \Lambda_k$ are the only distinct matrices among $\Lambda_1, \cdots, \Lambda_M$. Then for u not lying on any of a finite number of conics, the real numbers $u'\Lambda_i u$, $1 \leqq i \leqq k$ are distinct. Since the (Λ_j, θ_j) are all distinct, the θ_j associated with the same Λ_i are different $(k \leqq j \leqq M)$ and so outside a finite number of hyperplanes the corresponding numbers $u'\theta_j$ are distinct. Consequently, for u not lying on a finite number of conics and hyperplanes the pairs of real numbers $(u'\Lambda_j u, u'\theta_j)$, $1 \leqq j \leqq M$ are distinct. For such a choice of u, (8) asserts that the class of finite mixtures of one dimensional normal distributions is not identifiable, contrary to [5].

The referee has requested we include the following remarks connecting Proposition 2 to related results in [3]: The "lexicographical ordering" of Proposition 3 of [3] as well as the restriction that $\mathfrak{F}$ be a finite family are thus superfluous. Moreover, Proposition 1 thereof, with its incorrect assertion about the class of all mixtures of one-dimensional normal distributions, can hardly be a special case of Theorem 1 whose conclusion concerns finite mixtures. It in no way diminishes the class of one-dimensional normal distributions to suppose it lexicographically ordered (the so-called constraint of Proposition 1).

PROPOSITION 3. *Let $\mathfrak{F}_1$ and $\mathfrak{F}_2$ be the families of Propositions* 1 *and* 2. *If $\mathfrak{F} = \mathfrak{F}_1 \cup \mathfrak{F}_2$, then the set of all finite mixtures induced by $\mathfrak{F}$ is identifiable.*

PROOF. *Let* $\sum_{i=1}^{N} c_i F_i(x) = 0$ be a relation in $\mathfrak{F}$ and suppose $i \leqq k$ if and only if $F_i \varepsilon \mathfrak{F}_2$. Then, as $i > k$ implies $F_i(x) = 0$ if $x \varepsilon R^n$ has any negative components, for any such x,

$$\sum_{i=1}^{k} c_i F_i(x) = 0. \tag{9}$$

However, as (1) any $F \varepsilon \mathfrak{F}_2$ has an n-dimensional Taylor's series expansion about any point valid for R^n, (2) sums of functions are representable by the sums of their Taylor's series, and (3) any Taylor's series agreeing with the 0 function on a non-degenerate n-cube agrees with it everywhere, (9) must hold for all $x \varepsilon R^n$. By Proposition 2, this means that $c_j = 0, j \leqq k$. But this leaves a linear relation in $\mathfrak{F}_1$ contrary to Proposition 1.

PROPOSITION 4. *The set of all finite mixtures on the family $\mathfrak{F}$ of Cauchy densities is identifiable.*

PROOF. An element of $\mathfrak{F}$ is

$$f(x; u, k) = k(\pi(k^2 + (x - u)^2))^{-1}, \quad k > 0, \quad u \varepsilon R. \tag{10}$$

M_1 is the characteristic function operator. Thus $M_1(f(x; u, k)) = \exp(iut - k|t|) = \phi(t; u, k)$ where t is a real variable. On the set of characteristic functions generated by $\mathfrak{F}$ we define a second linear operator, M_2 such that

$$M_2(\phi(t; u, k)) = \begin{cases} \phi(t; u, k), & \text{if } t \geqq 0 \\ 0, & t < 0 \end{cases} = \xi(t; u, k). \tag{11}$$

Finally, on the functions composing the range of M_2, we define M_3, the Laplace transform, which also is a linear operator.

$$M_3(\xi(t; u, k)) = M_3 M_2 M_1(f(x; u, k)) = k(s + b)^{-1} \tag{12}$$

where $b = k - iu$ and s is a variable on the set D_k of complex numbers whose real part is greater than $-k$. $M_3 M_2 M_1$ is extended so that it is an isomorphism on $\mathfrak{F}$. The image of a linear relation may be expressed

$$0 = \sum_{j=1}^{N} c_j (s + b_j)^{-1}, \qquad s \varepsilon D = \bigcap_{j=1}^{N} D_{k_j}. \tag{13}$$

Assume $k_1 \leqq k_j$, $j > 1$. Multiply (13) by $(s + b_1)$ and let s converge to $-b_1$, while remaining in D. Then

$$|c_1| \leqq \lim_{s \to -b_1} |(s + b_1)| \sum_{i=2}^{N} |c_i (s + b_i)^{-1}| = 0. \tag{14}$$

Similarly, the other scalars are shown to be 0 and we have that the image of $\mathfrak{F}$ is linearly independent. Reference to the corollary completes the proof.

PROPOSITION 5. *The family $\mathfrak{F}$ of all non-degenerate negative binomial distributions induces an identifiable set of finite mixtures.*

PROOF. $\mathfrak{F}$ is indexed by the variable (p, r), $0 < p < 1$, $r > 0$ so that

$$f(x; p, r) = \binom{r+x-1}{x} p^r q^x, \quad \text{where} \quad q = 1 - p. \tag{15}$$

Theorem 2 of [5] states that $\mathfrak{F}$ generates identifiable finite mixtures if there is a 1–1 linear transformation on $\mathfrak{F}$ which totally orders it so that $F_1 < F_2$ implies ($\phi_j(t)$ being the image of F_j, S_{ϕ_j} the domain of $\phi_j(t)$) that $S_{\phi_1} \subset S_{\phi_2}$ and there is some element t_1 in the closure of S_{ϕ_1}, t_1 being independent of the choice of F_2, such that $\lim_{t \to t_1} (\phi_2(t)/\phi_1(t)) = 0$. In our case, the generating function transformation, $M(f(x; p, r)) = (p/(1 - qt))^r$, gives the required transformation. For if $F_1 < F_2 \Leftrightarrow [p_2 > p_1$, or $p_2 = p_1$ and $r_2 < r_1]$ then

$$S_{\phi_1} = \{t:(1 - p_1)^{-1} > |t|\} \subset \{t:(1 - p_2)^{-1} > |t|\} = S_{\phi_2}. \tag{16}$$

If $t_1 \equiv (1 - p_1)^{-1} = q_1^{-1}$,

$$\lim_{t \to t_1} \phi_2(t)/\phi_1(t) = \lim_{t \to t_1} (1 - q_1 t)^{r_1} [p_2^{r_2}/((1 - q_2 t)^{r_2} p_1^{r_1})] = 0. \tag{17}$$

Henry Teicher [4] has proven that the set of arbitrary mixtures generated by a translation parameter family $\mathfrak{F}$ is identifiable provided the Fourier transformation of $F(x)$, the cdf inducing $\mathfrak{F}$, is not 0 on any non-degenerate interval. In Proposition 6 the provision is avoided at the price of restricting the conclusion to finite mixtures.

PROPOSITION 6. *Let F be any univariate* cdf *whatsoever and $\mathfrak{F}$ the translation parameter family induced by F. Then the set of finite mixtures on $\mathfrak{F}$ is identifiable.*

PROOF. $\mathfrak{F}$ is indexed by α, a real variable, so that $F(x; \alpha) = F(x + \alpha)$. M_1 is the characteristic function transformation on $\mathfrak{F}$.

$$M_1(F(x; \alpha)) = \phi_\alpha(t) = \exp(-i\alpha t)\phi_0(t) = \exp(-i\alpha t) M_1(F(x)). \tag{18}$$

Let

$$0 = \textstyle\sum_{j=1}^N c_j \phi_{\alpha_j}(t) = \left(\sum_{j=1}^N c_j \exp(-i\alpha_j t)\right)\phi_0(t) \tag{19}$$

be a relation in $M_1(\mathfrak{F})$. $\phi_0(t)$, being a characteristic function, is equal to 1 at $t = 0$, and is continuous. Thus for some $d > 0$,

$$0 = \textstyle\sum_{j=1}^N c_j \exp(-i\alpha_j t), \qquad t \,\varepsilon\, (-d, d). \tag{20}$$

The following facts are easily verified: The function $e^{-i\alpha t}$ has a Taylor's series expansion valid for all real numbers. The function which is the sum of finitely many functions is representable as the sum of their Taylor's series. The zero function is the only function representable by a Taylor's series which has uncountably many 0's. Consequently, (20) is valid for any real number t. M_2 is the unilateral Laplace transformation, which is defined

$$M_2(g(t)) = \int_0^\infty g(t) \exp(-st)\, dt. \tag{21}$$

Under M_2, the transformation of (20) gives

$$0 = \sum_{j=1}^{N} c_j(s + i\alpha_j)^{-1}, \qquad \mathrm{Re}\,(s) > 0. \tag{22}$$

Then for any k, $1 \leqq k \leqq N$,

$$|c_k| \leqq \lim_{s \to -i\alpha_k} \sum_{j=1}^{N} |c_j(s + i\alpha_j)^{-1}|\,|s + i\alpha_k| = 0. \tag{23}$$

The relation (20) must therefore be trivial. Since M_2M_1 is an isomorphism on $\langle \mathfrak{F} \rangle$, reference to our corollary completes the proof.

Acknowledgment. We are grateful to Professor Edward Patrick for showing us an early version of [3] (which called our attention to [5] and [4]) and otherwise encouraging us in these studies.

REFERENCES

[1] Feller, W. (1943). On a general class of contagious distributions. *Ann. Math. Statist.* **14** 389–399.

[2] Nering, E. (1963). *Linear Algebra and Matrix Theory.* Wiley, New York.

[3] Patrick, E. and J. Hancock. (1966). Nonsupervised sequential classification and recognition of patterns. *IEEE Transactions on Information Theory. IT-12* 362–372.

[4] Teicher, H. (1961). Identifiability of mixtures. *Ann. Math. Statist.* **32** 244–248.

[5] Teicher, H. (1963). Identifiability of finite mixtures. *Ann. Math. Statist.* **34** 1265–1269