IDENTIFIABILITY OF FINITE MIXTURES[1]

By Henry Teicher

Purdue University

1. Summary. In general, the class of mixtures of the family of normal distributions or of Gamma (Type III) distributions or binomial distributions is not identifiable (see [3], [4] or Section 2 below for the meaning of this statement). In [4] it was shown that the class of all mixtures of a one-parameter additively-closed family of distributions *is* identifiable. Here, attention will be confined to *finite mixtures* and a theorem will be proved yielding the identifiability of *all* finite mixtures of Gamma (or of normal[2]) distributions. Thus, estimation of the mixing distribution on the basis of observations from the mixture is feasible in these cases. Some separate results on identifiability of finite mixtures of binomial distributions also appear.

2. Introduction. Let $\mathfrak{F} = \{F(x; \alpha), \alpha \in R_1^m\}$ constitute a family of one-dimensional cumulative distribution functions (c.d.f.'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 measurable[3] in $R^1 \times R_1^m$. Then, the one-dimensional c.d.f.

$$H(x) = \int_{R_1^m} F(x; \alpha)\, dG(\alpha)$$

is the image under the above mapping, say $\tilde{\mathfrak{F}}$, of the m-dimensional c.d.f. G (where the measure μ_G induced by G assigns measure one to R_1^m). The distribution H is called a mixture[4] (or G-mixture of $\mathfrak{F}$) while G is referred to as the mixing c.d.f. Let $\mathcal{G}$ denote the class of all such m-dimensional c.d.f.'s G and $\mathcal{H}$ the induced class of mixtures H. Then $\mathcal{H}$ will be said to be identifiable if $\tilde{\mathfrak{F}}$ is a one-to-one map of $\mathcal{G}$ onto $\mathcal{H}$. Similarly, any subclass $\mathcal{H}_0$ of $\mathcal{H}$ will be called identifiable if $\tilde{\mathfrak{F}}$ is a one-to-one map of $\mathcal{G}_0 = \tilde{\mathfrak{F}}^{-1}(\mathcal{H}_0)$ onto $\mathcal{H}_0$. If $\mathcal{H}_0 \subset \mathcal{H}_1 \subset \mathcal{H}$, identifiability of $\mathcal{H}_1$ clearly implies that of $\mathcal{H}_0$. H is called a finite mixture if its mixing distribution G 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. Let $\mathcal{G}_k$, $k = 1, 2, \cdots$ denote the class of all discrete distributions assigning positive mass to exactly k points of R_1^m and set $\mathcal{G}' = \bigcup_{k=1}^{\infty} \mathcal{G}_k$; denote the corresponding induced classes of finite mixtures by $\mathcal{H}_k$ and $\mathcal{H}'$. Since $\mathcal{H}_k \subset \mathcal{H}'$, $k = 1, 2 \cdots$ it is the identifiability of the larger class $\mathcal{H}'$ that will be sought.

Received May 17, 1963.

[1] Research under NSF Grant G-14719.

[2] Proposition 1 couched in different language appears in [4].

[3] As noted in [4], this is unduly strong.

[4] In [3] and [4], an H corresponding to degenerate G, i.e., an element of $\mathfrak{F}$, has not merited the name mixture. There is, however, some virtue in including such an H in the class of mixtures (as is done here) and then singling it out with the appellation "degenerate mixture."

When $\mathfrak{F}$ has only a finite number of elements the situation borders on the classical and it is easy to give a necessary and sufficient condition for the identifiability of $\mathcal{H}'$ (see Theorem 1). Nonetheless, the sufficient condition embodied in Theorem 2, applicable to (certain) non-denumerable families $\mathfrak{F}$, is far more useful.

3. Identifiability. In the following, $F_i(x)$, $1 \leqq i \leqq k$ signify fixed one-dimensional c.d.f.'s. Results analogous to Theorems 1 and 2 can be formulated and proved in the multi-dimensional case.

THEOREM 1. *A necessary and sufficient condition that the class* $\mathcal{H} = \{\sum_{i=1}^{k} c_i F_i(x): c_i \geqq 0, \sum_{i=1}^{k} c_i = 1\}$ *of all finite mixtures of the finite family* $\mathfrak{F} = \{F_1(x), F_2(x), \cdots F_k(x)\}$ *be identifiable is that there exist k real values* x_1, $x_2 \cdots x_k$ *for which the determinant of* $F_i(x_j)$, $1 \leqq i, j \leqq k$ *does not vanish.*

NECESSITY. Suppose that the condition of the theorem does not obtain, that is,

$$(1)\qquad \begin{vmatrix} F_1(x) & F_2(x) \cdots\cdots & F_k(x) \\ F_1(y_1) & F_2(y_1) \cdots & F_k(y_1) \\ \vdots & & \\ F_1(y_{k-1}) & F_2(y_{k-1}) \cdots & F_k(y_{k-1}) \end{vmatrix} = 0 \quad \text{for all } x \text{ and all } y_1 \cdots y_{k-1}.$$

Then

$$(2)\qquad \sum_{i=1}^{k} a_i F_i(x) \equiv_x 0 \quad \text{where } a_i = a_i(y_1 \cdots y_{k-1}).$$

If $a_i = 0$, $1 \leqq i \leqq k$ for all choices of $y_1 \cdots y_{k-1}$, then (1) and hence (2) holds with k replaced by $k - 1$. Let r be the largest integer such that $a_i \not\equiv_i 0$ for some choice of $y_1 \cdots y_r$. Since $F_i(y) \not\equiv F_j(y)$ for $i \neq j$, necessarily $2 \leqq r \leqq k - 1$. Thus, for this choice of $y_1 \cdots y_r$,

$$(3)\qquad \sum_{i=1}^{r+1} a_i F_i(x) \equiv_x 0, \qquad a_i \not\equiv_i 0, \sum_{i=1}^{r+1} a_i = 0.$$

Let $S_1 = \{i: 1 \leqq i \leqq r + 1, a_i > 0\}$, $S_2 = \{i: 1 \leqq i \leqq r + 1, a_i < 0\}$, $S_3 = \{1, 2 \cdots k\} - S_1 - S_2$. In view of (3), S_1 and S_2 are non-empty whence, defining $b = \max_{i \varepsilon S_2} |a_i|$,

$$(4)\qquad \begin{aligned} &\sum_{i \varepsilon S_1} 2a_i F_i(x) + \sum_{i \varepsilon S_2} (2b + a_i)F_i(x) + \sum_{i \varepsilon S_3} F_i(x) \\ &\qquad \equiv_x \sum_{i \varepsilon S_1} a_i F_i(x) + \sum_{i \varepsilon S_2} 2bF_i(x) + \sum_{i \varepsilon S_3} F_i(x). \end{aligned}$$

All the coefficients in (4) are positive and the sum of those on the left-hand side equals that of the right-hand side. Dividing through by this common value, (4) asserts the non-identifiability of $\mathcal{H}$, contrary to the hypothesis of necessity.

SUFFICIENCY. Consider the relation $\sum_{i=1}^{k} c_i F_i(x) \equiv \sum_{i=1}^{k} c_i' F_i(x)$ where $\sum_{i=1}^{k} c_i = \sum_{i=1}^{k} c_i' = 1$, that is, $\sum_{i=1}^{k} d_i F_i(x) \equiv_x 0$, $d_i = c_i - c_i'$. In particular, $\sum_{i=1}^{k} d_i F_i(x_j) = 0$, $1 \leqq j \leqq k$. According to the sufficiency hypothesis, the above determinant is non-vanishing, whence $d_i \equiv_i 0$, that is, $c_i \equiv c_i'$. Thus, identifiability prevails.

THEOREM 2. *Let* $\mathfrak{F} = \{F\}$ *be a family of* c.d.f.'s *with transforms* $\phi(t)$ *defined for* $t \varepsilon S_\phi$ *(the domain of definition of* φ*) such that the mapping* $M:F \to \phi$ *is linear and one-to-one. Suppose that there exists a total ordering* $(\preccurlyeq)$ *of* $\mathfrak{F}$ *such that* $F_1 \prec F_2$ *implies* (i) $S_{\phi_1} \subseteq S_{\phi_2}$, (ii) *the existence of some* $t_1 \varepsilon \bar{S}_{\phi_1}$ *(*t_1 *being independent of* ϕ_2*) such that* $\lim_{t \to t_1} \phi_2(t)/\phi_1(t) = 0$. *Then the class* $\mathcal{H}'$ *of all finite mixtures of* $\mathfrak{F}$ *is identifiable.*

PROOF. Suppose there are two finite sets of elements of $\mathfrak{F}$, say $\mathfrak{F}_1 = \{F_i, 1 \leqq i \leqq k\}$ and $\mathfrak{F}_2 = \{\hat{F}_j, 1 \leqq j \leqq \hat{k}\}$ such that

$$(5) \qquad \sum_{i=1}^{k} c_i F_i(x) \equiv_x \sum_{j=1}^{\hat{k}} \hat{c}_j \hat{F}_j(x), \qquad 0 < c_i, \hat{c}_j \leqq 1, \qquad \sum_{i=1}^{k} c_i = \sum_{j=1}^{\hat{k}} \hat{c}_j = 1.$$

Without loss of generality, index the c.d.f.'s so that $F_i \prec F_j$, $\hat{F}_i \prec \hat{F}_j$ for $i < j$. If $F_1 \neq \hat{F}_1$, suppose also without loss of generality that $F_1 \prec \hat{F}_1$. Then, $F_1 \prec \hat{F}_j$, $1 \leqq j \leqq \hat{k}$ and from the transform(ed) version of (5), it follows that for $t \varepsilon T_1 = S_{\phi_1} \cdot [t:\phi_1(t) \neq 0]$,

$$c_1 + \sum_{i=2}^{k} c_i[\phi_i(t)/\phi_1(t)] \equiv_t \sum_{j=1}^{\hat{k}} \hat{c}_j[\hat{\phi}_j(t)/\phi_1(t)].$$

Letting $t \to t_1$ through values in T_1 (this is possible), $c_1 = 0$ contradicting the supposition of (5) that $c_1 > 0$. Thus, $F_1 = \hat{F}_1$ and for any $t \varepsilon T_1$

$$(c_1 - \hat{c}_1) + \sum_{i=2}^{k} [\phi_i(t)/\phi_1(t)] \equiv_t \sum_{j=2}^{\hat{k}} \hat{c}_j[\hat{\phi}_j(t)/\phi_1(t)].$$

Again letting $t \to t_1$ through values in T_1, $c_1 = \hat{c}_1$ whence

$$(6) \qquad \sum_{i=2}^{k} c_i F_i(x) \equiv_x \sum_{j=2}^{\hat{k}} \hat{c}_j \hat{F}_j(x).$$

Repeating the prior argument a finite number of times, we conclude that $F_i = \hat{F}_i$ and $c_i = \hat{c}_i$ for $i = 1, 2 \cdots \min(k, \hat{k})$. Further, if $k \neq \hat{k}$, say $k > \hat{k}$, then $\sum_{i=\hat{k}+1}^{k} c_i F_i(x) \equiv 0$ implying $c_i = 0$, $\hat{k} + 1 \leqq i \leqq k$ in contradiction to (5). Thus, $k = \hat{k}$, $c_i = \hat{c}_i$ and $F_i = \hat{F}_i$ $1 \leqq i \leqq k$, implying $\mathfrak{F}_1 = \mathfrak{F}_2$ and identifiability of $\mathcal{H}'$.

PROPOSITION 1. *The class of all finite mixtures of normal distributions is identifiable.*[2]

PROOF. Let $N = N(x; \theta, \sigma^2)$ denote the normal c.d.f. with mean θ and variance $\sigma^2 > 0$. Its bilateral Laplace transform is given by $\phi(t; \theta, \sigma^2) = \exp\{\sigma^2 t^2/2 - \theta t\}$. Order the family lexicographically by: $N_1 = N(x; \theta_1, \sigma_1^2) \prec N(x; \theta_2, \sigma_2^2) = N_2$ if $\sigma_1 > \sigma_2$ or if $\sigma_1 = \sigma_2$ but $\theta_1 < \theta_2$. Then Theorem 2 applies with $S_\phi = (-\infty, \infty)$ and $t_1 = +\infty$.

PROPOSITION 2. *The class of all finite mixtures of Gamma distributions is identifiable.*

PROOF. The Type III c.d.f.'s form a 2-parameter family with $F(x; \theta, \alpha) = \theta^\alpha[\Gamma(\alpha)]^{-1}\int_0^x y^{\alpha-1}e^{-\theta y}\, dy$, $\alpha > 0$, $\theta > 0$ and their Laplace transforms are given by

$\phi(t; \theta, \alpha) = (1 + t/\theta)^{-\alpha}$ for $t > -\theta$. Order these lexicographically by: $F(x; \theta_1, \alpha_1) \prec F(x; \theta_2, \alpha_2)$ if $\theta_1 < \theta_2$ or if $\theta_1 = \theta_2$ but $\alpha_1 > \alpha_2$. Then $S_{\phi_1} = (-\theta_1, \infty) \subseteq (-\theta_2, \infty) = S_{\phi_2}$ and $\lim_{t \to -\theta_1} \phi(t; \theta_2, \alpha_2)/\phi(t; \theta_1, \alpha_1) = 0$ with $-\theta_1 \varepsilon \bar{S}_{\phi_1}$. Thus, Theorem 2 is once more applicable.

Denote by $F(x; n, p)$ the binomial distribution with parameters n (a positive integer) and $p (0 < p < 1)$. Although this two-parameter family may be lexicographically ordered as in the prior examples, the condition of Theorem 2 is not satisfied. In fact, it is clear from [4] that the class of all finite mixtures of binomial distributions is not identifiable. Information as to the nature of those subfamilies of binomial distributions that are identifiable under finite mixture may be gleaned from

PROPOSITION 3. *Let* $\mathfrak{F}_1 = \{F(x; n_i', p_i'), 1 \leqq i \leqq k'\}$ *and* $\mathfrak{F}_2 = \{F(x; n_i'', p_i''), 1 \leqq i \leqq k''\}$ *denote 2 finite families of binomial distributions; let* k = *number of elements in* $\mathfrak{F}_1 \cup \mathfrak{F}_2$ *and* $\bar{n}_1 > \bar{n}_2 > \cdots > \bar{n}_h$ *be the distinct integral parameters of the members of* $\mathfrak{F}_1 \cup \mathfrak{F}_2$. *A necessary but, in general, insufficient condition for*

$$(7) \quad \sum_{i=1}^{k'} c_i' F(x; n_i', p_i') \equiv_x \sum_{i=1}^{k''} c_i'' F(x; n_i'', p_i''), \; \sum_{i=1}^{k'} c_i' = \sum_{i=1}^{k''} c_i'' = 1, \quad 0 < c_i, c_i''$$

to imply

$$(8) \qquad k' = k'', \; (n_i', p_i') = (n_{j_i}'', p_{j_i}'') \text{ for some permutation}$$

$(j_1, \cdots j_k)$ of $(1, 2 \cdots k)$ *is that*

$$(9) \qquad \bar{n}_h \geqq r_h - 1$$

where r_i = *number of occurrences of* $\bar{n}_i$ *among the elements of* $\mathfrak{F}_1 \cup \mathfrak{F}_2$, $1 \leqq i \leqq h$. *A sufficient condition that* (7) *imply* (8) *is that* (9) *and*

$$(10) \qquad \bar{n}_i - \bar{n}_{i+1} \geqq r_i, \qquad 1 \leqq i \leqq h - 1$$

hold.

PROOF. (7) is equivalent to

$$(11) \qquad \sum_{i=1}^{k} d_i F(x; n_i, p_i) \equiv_x 0, \qquad \sum_{i=1}^{k} d_i = 0$$

where $F(x; n_i, p_i)$, $1 \leqq i \leqq k$ are the elements of $\mathfrak{F}_1 \cup \mathfrak{F}_2$. The probability generating function of $F(x; n, p)$ is $(pz + 1 - p)^n = (1 + pw)^n$ where $w = z - 1$. Thus, (11) is tantamount to

$$(12) \qquad \sum_{i=1}^{k} d_i (1 + wp_i)^{n_i} \equiv_w 0.$$

Setting $s_j = \sum_{i=1}^{j} r_i$, $s_0 = 0$ we have from (12)

$$\sum_{i=1}^{s_1} d_i p_i^j = 0, \qquad \bar{n}_2 + 1 \leqq j \leqq \bar{n}_1$$

$$(13)\qquad \binom{\bar{n}_1}{j} \cdot \sum_{i=1}^{s_1} d_i\, p_i^j + \binom{\bar{n}_2}{j} \sum_{i=s_1+1}^{s_2} d_i\, p_i^j = 0, \qquad \bar{n}_3 + 1 \leqq j \leqq \bar{n}_2$$

$$\vdots$$

$$\binom{\bar{n}_1}{j} \cdot \sum_{i=1}^{s_1} d_i\, p_i^j + \cdots + \binom{\bar{n}_h}{j} \sum_{i=s_{h-1}+1}^{s_h} d_i\, p_i^j = 0, \qquad 0 \leqq j \leqq \bar{n}_h .$$

If now, in violation of (9), $\bar{n}_h < r_h - 1$, we may choose $d_i = 0, 1 \leqq i \leqq s_{h-1}$ and then a non-zero solution of $\sum_{i=s_{h-1}+1}^{s_h} d_i p_i^j = 0, 0 \leqq j \leqq \bar{n}_h$ thus satisfying (13) and consequently (7) but contradicting (8). A specific counter-example can then be constructed along the lines of Theorem 1.

On the other hand, if (9) and (10) obtain, then for each $\alpha = 1, 2 \cdots, h$ the only solution of the equations $\sum_{i=s_{\alpha-1}+1}^{s_\alpha} d_i p_i^j = 0$, $\bar{n}_{\alpha+1} + 1 \leqq j \leqq \bar{n}_\alpha$ (and hence also of (13)) is the zero solution. Thus, $d_i = 0, 1 \leqq i \leqq k$, implying (8) since $c_i', c_i'' > 0$. It is easy to adduce an example showing that (9) in and of itself, is not, in general, sufficient.

The preceding proposition leads directly to:

PROPOSITION 4.

(i) *Let* $\mathfrak{F} = \{F(x; n, p), 0 < p < 1\}$ *constitute a one-parameter family of binomial distributions, n being fixed. A necessary and sufficient condition that the class* $\bigcup_{j=1}^k \mathcal{H}_j$ *of all finite mixtures of at most k elements of* $\mathfrak{F}$ *be identifiable is that* $n \geqq 2k - 1$.

(ii) *Let* $\mathfrak{F} = \{F(x; n_i, p_i), 0 < p_i < 1, i = 1, 2, \cdots\}$ *be a countable family of binomial distributions with distinct integral parameters, that is,* $n_i \neq n_j$ for $i \neq j$. *Then the class* $\mathcal{H}'$ *of all finite mixtures of* $\mathfrak{F}$ *is identifiable.*

The necessity of the condition of Proposition 4 (i) for the case $k = 2$ is noted in [1].

REFERENCES

[1] BLISCHKE, W. R. (1962). Moment estimators for the parameters of a mixture of two binomial distributions. *Ann Math. Statist.* **33** 444–454.

[2] MEDGYESSY, P. (1961). *Decomposition of Superpositions of Distribution Functions.* Hungarian Acad. of Sciences, Budapest.

[3] TEICHER, H. (1960). On the mixture of distributions. *Ann. Math. Statist.* **31** 55–73.

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