# IDENTIFIABILITY OF MIXTURES[1]

By Henry Teicher

*New York University and Purdue University*

**1. Summary.** The class of mixtures of a one-parameter additively-closed family of distributions is proved identifiable. A condition for a class of scale parameter mixtures to be identifiable is indicated and applications to Type III and uniform distributions are made.

**2. Introduction.** Let $R_1^m$ be a measurable subset of Euclidean $m$-space $R^m$ and $\mathfrak{F} = \{F(x; \alpha), \alpha \varepsilon R_1^m\}$, where $F(x; \alpha)$ is a cumulative distribution function in the variable $x$ for each $\alpha \varepsilon R_1^m$ and also measurable[2] on the product space of $x$ ana $\alpha$. Then [1], for any non-degenerate[3] $m$-dimensional c.d.f. $G$ whose induced Lebesgue-Stieltjes measure $\mu_G$ assigns measure one to $R_1^m$, the c.d.f.

$$H(x) = \int_{R_1^m} F(x; \alpha) dG(\alpha) \tag{1}$$

is called a $G$-mixture of $\mathfrak{F}$ or, more briefly, a mixture.

Let $\mathcal{G}$ denote a class of such c.d.f.'s $[G]$, $\mathcal{H}$ the induced class of mixtures $[H]$ and $\mathcal{J}$ the class of degenerate[3] distributions in $R^m$. Then $\mathcal{H}$ will be called identifiable in $\mathcal{G}$ (with respect to $\mathfrak{F}$) if (1) effects a one-to-one correspondence between $\mathcal{H} \cup \mathfrak{F}$ and $\mathcal{G} \cup \mathcal{J}$; equivalently, if the relationship

$$H = \int F(x; \alpha)\, dG(\alpha) = \int F(x; \alpha)\, dG^*(\alpha)$$

implies $G = G^*$ for all $G^* \varepsilon \mathcal{G} \cup \mathcal{J}$. If $\mathcal{H}$ is identifiable in the class of all $G \not\varepsilon \mathcal{J}$, it will simply be called identifiable. Clearly, the identifiability question must be settled before one can meaningfully discuss the problem of estimating the mixing c.d.f. $G$ on the basis of observations from the mixture $H$. Here, the functional form of $\mathfrak{F}$ would be presumed known and if $R_1^m$ were countable, its elements might also be supposed known. Now, the only positive identifiability results familiar to the author concern the cases (i) $\mathfrak{F}$ in the Poisson family [2], (ii) $\mathfrak{F}$ is the normal family, [1]. It is the purpose of this note to provide tools (the theorem and proposition) via which one may establish the identifiability of mixtures and to carry the latter out for a few of the more popular families.

**3. Mixtures of Additively closed families.** Let $D$ be the generic notation for an additive Abelian semi-group; $D(I)$, $D(r)$, $D(R)$ will denote the semi-groups

Received February 16, 1960; revised July 25, 1960.

[1] Research under Office of Naval Research Contract.

[2] For this it suffices, according to Bourbaki, *Integration des Mesures*, p. 105, or [6], to stipulate that $F(x; \alpha)$ be measurable in $\alpha$ for all $x$ in view of the fact that $F(x; \alpha)$ is a cumulative distribution function for each $\alpha$.

[3] Here, degenerate signifies that $G$ concentrates all its mass at a single point of $R^m$. Concomitantly, any family $\mathfrak{F} = \{F(x; \alpha), \alpha \varepsilon R_1^m \subset R^m\}$ being mixed is tacitly presumed to contain at least two elements.

of integers, rationals and real numbers respectively; $D(I+)$, $D(r+)$, $D(R+)$ signify the analogous semi-groups restricted to positive values.

A family of c.d.f.'s, $\mathfrak{F} = \{F(x; \alpha), \alpha \varepsilon D\}$ has been called [3] additively closed (a.c.) provided for each $\alpha, \beta \varepsilon D$

$$F(x; \alpha) * F(x; \beta) = F(x; \alpha + \beta).^4$$

If (i) $\mathfrak{F} = \{F(x; \alpha), \alpha \varepsilon D\}$ is a.c. (ii) $F(x; \alpha)$ is measurable (iii) $D$ is a measurable subset of $R^m$ with $\mu_G\{D\} = 1$, the mixture (1) with $R_1^m = D$ is dubbed a mixture of the additively closed family $\mathfrak{F}$.

It was shown in [3] that for $m = 1$ and $D = D(I+)$, $D(r+)$ an a.c. family $\{F(x; \alpha)\}$ possesses characteristic functions (c.f.'s) $\phi(t; \alpha)$ of the form

$$\phi(t; \alpha) = [\phi(t)]^\alpha, \tag{2}$$

where $\phi(t) = \phi(t; 1)$ is a c.f. independent of $\alpha$. An examination of the proofs of Theorems 1 and 2 of [3] reveals that (2) also obtains in the case $D = D(R+)$ provided only that $F(x; \alpha)$ (hence $\phi(t; \alpha)$) is measurable. Thus in the cases of major interest, viz., $D = D(I+)$, $D(r+)$, $D(R+)$, (also $D(I)$, $D(r)$, $D(R)$), the c.f., say $\psi_H(t)$, of a mixture $H$ of an a.c. family $\mathfrak{F}$, is of the form

$$\psi_H(t) = \int_D [\phi(t)]^\alpha \, dG(\alpha). \tag{3}$$

If, in (3), $|\phi(t)| \equiv 1$ then $\phi(t) = e^{it\theta}$ with $\theta$ real and non-zero since $\mathfrak{F}$ contains at least two elements. Hence, $G$ is uniquely determined by $\psi_H(t)$ which, in turn, is uniquely generated by $H$. (This shows that the ensuing theorem is also valid but trivial when $D = D(I)$, $D(r)$ or $D(R)$).

Alternatively, if $D$ contains only non-negative values, the transform

$$\psi(z; G) = \int_D z^\alpha \, dG(\alpha)$$

is analytic at least in the annulus $0 < |z| < 1$; if two different c.d.f.'s $G_1$ and $G_2$ engendered the same mixture $H$, then $\psi(z; G_1)$ and $\psi(z; G_2)$ would coincide for $z = \phi(t)$ and consequently throughout the annulus. This would entail $\psi(\rho e^{it}; G_1) = \psi(\rho e^{it}; G_2)$ for all $\rho$ in $(0, 1)$ and hence, by the dominated convergence theorem, for $\rho = 1$. But $\psi(e^{it}; G_1) = \psi(e^{it}; G_2)$ implies $G_1 = G_2$ by the identity theorem for Fourier transforms. This proves the

THEOREM: *If $m = 1$ and $D$ is $D(I+)$ or $D(r+)$ or $D(R+)$, the class of mixtures $\{\int_D F(x; \alpha) \, dG(\alpha)\}$ of an additively closed family $\{F(x; \alpha), \alpha \varepsilon D\}$ is identifiable.*

Zero could also be included in $D$ without altering the result. The same argument yields Theorem 4 of [1] without superfluous restrictions:

COROLLARY: *If $m = 1$, no mixture of an* a.c. *family $\mathfrak{F}$($D$ as in the theorem) is an element of $\mathfrak{F}$.*

When $m > 1$ the c.f. of an a.c. family is of the form $\prod_{j=1}^m [f_j(t)]^{\alpha_j}$ (at least for suitable $D$). But here some of the parameters $\alpha_j$ may assume both positive

[4] As usual, $*$ denotes convolution.

and negative values and the preceding argument no longer applies. Furthermore, even the usually pliable normal family (when mixed on both parameters) does not generate an identifiable family. This and later examples suggest that no comparable conclusion obtains when $m > 1$.

**4. Translation and Scale Parameter Mixtures and Applications.** Cases of special interest not necessarily subsumed by the theorem arise when a single distribution $F(x)$ generates the family $\mathfrak{F} = \{F(x; \alpha)\}$ via location and/or scale changes. Consider a scale parameter mixture ($m = 1$)

$$H(x) = \int_0^\infty F(x\alpha)\, dG(\alpha), \tag{4}$$

where the "generating" distribution[5] $F(x)$ satisfies $F(0+) = 0$. Let[6] $x = e^y$, $\alpha = e^{-\beta}$. Then

$$\bar{H}(y) = H(e^y) = \int_{-\infty}^\infty \bar{F}(y - \beta)\, d\bar{G}(\beta) = \bar{F} * \bar{G},$$

where $\bar{F}(w) = F(e^w)$, $\bar{G}(\beta) = 1 - G(e^{-\beta})$ and $-\infty < w, \beta, y < \infty$. Consequently, $\int_0^\infty F(x\alpha)\, dG_1(\alpha) = \int_0^\infty F(x\alpha)\, dG_2(\alpha)$ implies $\bar{F} * \bar{G}_1 = \bar{F} * \bar{G}_2$ and we have the rather obvious

PROPOSITION.

(i) If the Fourier transform of $\bar{F}(x) = F(e^x)$ is not identically zero in some non-degenerate real interval, the class of scale parameter mixtures (4) is identifiable.

(ii) If the Fourier transform of $F(x)$ is not identically zero in some non-degenerate real interval, the class of translation parameter mixtures

$$\{\int F(x - \alpha)\, dG(\alpha)\}, \qquad m = 1,$$

is identifiable.

EXAMPLE 1. Mixtures of Type III distributions.

$$\begin{aligned} F(x; \lambda, \gamma) &= \gamma^\lambda[\Gamma(\lambda)]^{-1} \int_0^x u^{\lambda-1} e^{-\gamma u}\, du, \qquad x > 0, \gamma > 0, \lambda > 0, \\ \phi(t; \lambda, \gamma) &= [1 - (it/\gamma)]^{-\lambda}. \end{aligned} \tag{5}$$

Since, for fixed $\gamma$, $\{F(x; y, \gamma\}$ is an a.c. family, the theorem insures that the class of $G(\lambda)$-mixtures is identifiable.

When $\lambda$ is fixed, $\{F(x; \lambda, \gamma)\}$ is a scale-parameter family generated by $F(x) = F(x; \lambda, 1)$. Since the c.f. of $\bar{F}(x)$ is $\Gamma(\lambda + it)/[\Gamma(\lambda)]$, it follows from the proposition that the class of $G(\gamma)$-mixtures of $\{F(x; \lambda, \gamma)\}$ is identifiable. On the other hand, the class of all $G(\lambda, \gamma)$-mixtures of $\{F(x; \lambda, \gamma)\}$ is not identifiable as shown by the example of [4].

---

[5] Although a generating distribution is not uniquely determined by the family in question, there is usually a "natural" generator. Thus, in (5) when $\lambda$ is fixed $F(x; \lambda, 1)$ seems the obvious candidate.

[6] This is an ancient device and is also used in [5].

EXAMPLE 2. Mixtures of Uniform distributions

$$(6)\qquad F(x;\,\theta,\,\sigma) = \begin{cases} 1, & x > \theta + \sigma \\ \dfrac{x + \sigma - \theta}{2\sigma}, & \theta - \sigma \leqq x \leqq \theta + \sigma \\ 0, & y < \theta - \sigma. \end{cases}$$

Of course, $\sigma > 0$, $\theta \varepsilon R^1$. The class $\mathcal{H}$ of mixtures of $\mathfrak{F} = \{F(x;\,\theta,\,\sigma)\}$ has as generic element

$$(7)\qquad H(x) = \int F(x;\,\theta,\,\sigma)\, dG(\theta,\,\sigma),$$

where $\mu_G$ assigns measure zero to $\{\theta,\,\sigma | \sigma \leqq 0\}$. If $f(x;\,\theta,\,\sigma) = (\partial/\partial x)F(x;\,\theta,\,\sigma)$, then

$$(8)\quad h(x) = H'(x) = \int f(x;\,\theta,\,\sigma)\, dG(\theta,\,\sigma) = \int_{-\infty}^{\infty} \int_{|x-\theta|}^{\infty} \frac{1}{2\sigma} G_\theta(\sigma)\, dG_2(\theta),$$

where $G_\theta(\sigma)$ is the conditional distribution of $\sigma$ for fixed $\theta$ and $G_2$ is the marginal c.d.f. of $\theta$.

If $\sigma$ is unvarying (say, $\sigma = 1$), the class of $G(\theta)$-mixtures of $\{F(x;\,\theta,\,1)\}$ is identifiable since it consists of translation parameter mixtures and the c.f. $[(\sin t)/t]$ of the generating distribution has only countably many real zeros.

If $\theta$ is fixed (say, $\theta = 0$), the class of $G(\sigma)$-mixtures of $\{F(x;\,0,\,\sigma)\}$ is likewise identifiable, a direct consequence of

$$(9)\qquad h(x) = \int_{|x|}^{\infty} (2\sigma)^{-1}\, dG(\sigma).$$

In fact, as revealed by (9), any symmetric density which is continuous on one side and non-increasing on $(0,\,\infty)$ is a scale parameter mixture of uniform distributions with $dG(\sigma) = -2\sigma\, dh(\sigma)$, $\sigma > 0$, and $G(0) = 0$.

However, the class $\mathcal{H}$ of all mixtures of $\mathfrak{F}$ is not identifiable. In fact, a uniform distribution is itself a mixture of uniform distributions, e.g.,

$$F(x;\,\tfrac{1}{2},\,\tfrac{1}{2}) = \tfrac{1}{2}F(x;\,\tfrac{1}{4},\,\tfrac{1}{4}) + \tfrac{1}{2}F(x;\,\tfrac{3}{4},\,\tfrac{1}{4}).$$

Lest it be thought that the class of mixtures of a one-parameter family of c.d.f.'s is always identifiable, consider finally

EXAMPLE 3. Mixtures of Binomial Distributions.

$$(10)\qquad F(x;\,n,\,p) = \sum_{j<x} \binom{n}{j} p^j (1 - p)^{n-j}.$$

By the theorem, for fixed $p$ the class of $G(n)$-mixtures of $\mathfrak{F} = (F(x;\,n,\,p)\}$ is identifiable.

However, when $n$ is fixed, the class of $G(p)$-mixtures of $\mathfrak{F}$ is not identifiable.[7]

[7] Comparable statements apply for other choices of $\mathfrak{F}$. The crucial points appear to be (i) all distributions having essentially the same finite spectrum (ii) the functional form in which the parameter enters.

For, any $G(p)$-mixture is a step function with a jump of

$$h_j = \int_0^1 \binom{n}{j} p^j(1-p)^{n-j}\, dG(p) = \binom{n}{j} \sum_{i=0}^{n-j} (-1)^i \binom{n-j}{i} \int_0^1 p^{i+j}\, dG(p)$$

at $j = 0, 1, 2 \cdots n$. Consequently, a $G_1$-mixture of $\mathfrak{F}$ and a $G_2$-mixture of $\mathfrak{F}$ will be identical if and only if

$$\sum_{i=j}^{n} (-1)^i \binom{n-j}{i-j} \nu_i^{(1)} = \sum_{i=j}^{n} (-1)^i \binom{n-i}{i-j} \nu_i^{(2)}, \qquad j = 0, 1, \cdots, n,$$

where $\nu_i^{(k)} = \int_0^1 p^i\, dG_k(p)$, $k = 1, 2$; hence, if and only if $G_1$ and $G_2$ have their first $n$ moments identical.[8] Thus, the class of $G(p)$-mixtures of $\mathfrak{F} = \{F(x; n, p)\}$ is not identifiable.

In conclusion, note that the analogue of a "basis" theorem does not hold for mixtures. That is, it does not follow from the non-identifiability of the class $\mathfrak{H}$ of mixtures of $\mathfrak{F} = \{F(x; \alpha)\}$ that

$$F(x; \beta) = \int F(x; \alpha)\, dG(\alpha) \tag{11}$$

for some element $F(x; \beta)$ of $\mathfrak{F}$ (and non-degenerate $G$). (Clearly, (11) implies the nonidentifiability of $\mathfrak{H}$). It suffices to take $\mathfrak{F}$ to be the binomial family with $n \geqq 2$ but fixed. Since $F(x; \beta)$ may be (temporarily) regarded as a $G_1$-mixture of $\mathfrak{F} = \{F(x; \alpha)\}$ with $G_1$ degenerate (and having unit saltus at $\alpha = \beta$) and the remarks of the preceding paragraph apply even for degenerate $G_1$, it would follow from (11) that $G$ had first and second moments equal to $\beta$ and $\beta^2$ respectively. But this in turn entails $G = G_1$.

## REFERENCES

[1] H. Teicher, "On the mixture of distributions," *Ann. Math. Stat.*, Vol. 31 (1960), pp. 55–73.

[2] W. Feller, "On a general class of contagious distributions," *Ann. Math. Stat.*, Vol. 14 (1943), pp. 389–399.

[3] H. Teicher, "On the convolution of distributions," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 775–778.

[4] Herbert E. Robbins and E. J. G. Pitman, "Application of the method of mixtures to quadratic forms in normal variates," *Ann. Math. Stat.*, Vol. 20 (1949), pp. 552–560.

[5] E. M. L. Beale and C. L. Mallows, "Scale mixing of symmetric distributions with zero means," *Ann. Math. Stat.*, Vol. 30 (1959), pp. 1145–1151.

[6] E. Marczewski and C. Ryll-Nardzewski, "Sur la measurabilité des fonctions de plusieurs variables," *Ann. Soc. Mathématique*, Vol. 25 (1952), pp. 145–154.

[8] This appears to be known in some quarters and was discovered independently by Prof. M. Skibinsky of Purdue University.