CR 151338

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF HOUSTON HOUSTON, TEXAS

(NASA-CR-151338) SUFFICIENT STATISTICS FOR MIXTURES OF MEASURES IN A HOMOGENEOUS FAMILY (Houston Univ.) 18 p HC A02/MF A01 CSCL 12A

N77-24887

Unclas
G3/65 26140

SUFFICIENT STATISTICS
FOR MIXTURES OF MEASURES
IN A HOMOGENEOUS FAMILY

BY CHARLES PETERS

REPORT #64 MARCH 1977

PREPARED FOR
EARTH OBSERVATION DIVISION, JSC
UNDER
CONTRACT NAS-9-15000

HOUSTON, TEXAS 77004

Sufficient Statistics for Mixtures of Measures in a Homogeneous Family

By

Charles Peters
Department of Mathematics
University of Houston

March, 1977
Report 64

Sufficient Statistics for Mixtures of Measures in a Homogeneous Family

by

Charles Peters

Department of Mathematics

University of Houston

1. Introduction:

Let $(X,\mathcal{A})$ and $(Y,\mathcal{B})$ be measureable spaces and let $T : X \to Y$ be surjective and measureable. Let $\mathcal{M}$ be a set of finite positive measures on $(X,\mathcal{A})$. For each $\mu \in \mathcal{M}$ there corresponds a measure μT^{-1} on $(Y,\mathcal{B})$ defined for $F \in \mathcal{B}$ by

$$\mu T^{-1}(F) = \mu(T^{-1}(F)).$$

If f is a μ-integrable real valued function on X, then as a consequence of the Radon Nikodym Theorem, there is a μT^{-1}- integrable function $e_\mu(f)$ on Y satisfying

$$\int_F e_\mu(f)\,d\mu T^{-1} = \int_{T^{-1}(F)} f\,d\mu$$

for each $F \in \mathcal{B}$. Clearly $e_\mu(f)$ is defined only up to sets in Y of μT^{-1} measure 0 and $f = g$ a.e. (μ) implies $e_\mu(f) = e_\mu(g)$ a.e. (μT^{-1}). The linear operator e_μ defined as above maps the space $\mathcal{L}^1(X,\mathcal{A},\mu)$ to the space $\mathcal{L}^1(Y,\mathcal{B},\mu T^{-1})$ and is called the conditional expectation operator. Its value

$e_\mu(f)$ at $f \in \mathcal{L}'(X, \mathcal{A}, \mu)$ is called the <u>conditional expectation of f given T</u>.

The conditional probability of an event $E \in \mathcal{A}$ is defined as

$$P_\mu(E) = e_\mu(\chi_E)$$

where χ_E is the indicator function of E. The conditional probability functions satisfy

(a) $$P_\mu : \mathcal{A} \to \mathcal{J}(Y, \mathcal{B}, \mu T^{-1}).$$

where $\mathcal{J}(Y, \mathcal{B}, \mu T^{-1})$ is the set of all real valued $\mathcal{B}$-measureable functions on Y, with equality defined as equality a.e. (μT^{-1}).

(b) For each $F \in \mathcal{B}, E \in \mathcal{A}$,

$$\mu(E \cap T^{-1}(F)) = \int_F P_\mu(E) d\mu T^{-1}$$

(c) $0 \leq P_\mu(E) \leq 1$ for each $E \in \mathcal{A}$ and $P_\mu(X) = 1$.

(d) If $\{E_n\}_{n=1}^\infty$ is a disjoint sequence of events in $\mathcal{A}$,

$$P_\mu\left(\bigcup_{n=1}^\infty E_n\right) = \sum_{n=1}^\infty P_\mu(E_n) \quad \text{a.e. } (\mu T^{-1}).$$

It should be noted that P_μ satisfies property (c) even when μ is not a probability measure.

The transformation T is called a <u>sufficient statistic</u> for $\mathcal{M}$ if for each $E \in \mathcal{A}$ there is a $\mathcal{B}$-measureable function P(E) on Y such that for each $\mu \in \mathcal{M}$, $P_\mu(E) = P(E)$ a.e., (μT^{-1}). The set $\mathcal{M}$ is <u>dominated</u> by a measure λ (perhaps not in $\mathcal{M}$) if for each $\mu \in \mathcal{M}$, μ is absolutely continuous with respect to λ, (written $\mu << \lambda$.) $\mathcal{M}$ is <u>homogeneous</u> if it is dominated by each of its members. A measure λ is <u>equivalent</u> to $\mathcal{M}$ if λ dominates $\mathcal{M}$ and $\mu(E) = 0$ for each $\mu \in \mathcal{M}$ implies $\lambda(E) = 0$.

The notation and terminology used in this paper are taken from (Halmos and Savage; 1949), as are the following three theorems. The notation $\frac{d\mu}{d\lambda}(\epsilon)T^{-1}(\mathcal{B})$ means that there is an element of the equivalence class $\frac{d\mu}{d\lambda}$ of Radon-Nikodym derivatives which is $T^{-1}(\mathcal{B})$ measureable.

<u>Theorem</u> 1: If $\mathcal{M}$ is dominated, then a statistic T is sufficient for $\mathcal{M}$ if and only if there exists a measure λ equivalent to $\mathcal{M}$ such that for each $\mu \in \mathcal{M}$, $\frac{d\mu}{d\lambda}(\epsilon)T^{-1}(\mathcal{B})$.

<u>Theorem 2</u>: If $\mathcal{M}$ is dominated, then a statistic T is sufficient for $\mathcal{M}$ if and only if T is sufficient for each pair $\{\mu,\nu\}$ of elements of $\mathcal{M}$.

<u>Theorem 3</u>: If $\mathcal{M}$ is homogeneous, then a statistic T is sufficient for $\mathcal{M}$ if and only if $\frac{d\mu}{d\nu}(\epsilon)T^{-1}(\mathcal{B})$ for each $\mu,\nu \in \mathcal{M}$.

2. <u>Homogeneous Families</u>:

Henceforth, we will assume that $\mathcal{M}$ is homogeneous. Let $C(\mathcal{M})$ denote the cone generated by $\mathcal{M}$, excluding the zero measure. That is, $C(\mathcal{M})$ is the set of all finite linear combinations, with strictly positive coefficients, of elements of $\mathcal{M}$. Elements of $C(\mathcal{M})$ are termed <u>mixtures</u> of elements of $\mathcal{M}$. Clearly, $C(\mathcal{M})$ is also homogeneous; hence, the spaces $\mathcal{F}(Y,\mathcal{B},\mu T^{-1})$ are all the same for $\mu \in C(\mathcal{M})$ and may be denoted simply by $\mathcal{F}$. For $\mu \in C(\mathcal{M})$, P_μ maps $\mathcal{A}$ to $\mathcal{F}$ and it is clear from the definition of a sufficient statistic that T is sufficient for a subset $\mathcal{N}$ of $C(\mathcal{M})$ if and only if the conditional probability functions P_μ for $\mu \varepsilon \mathcal{N}$ are all equal.

<u>Lemma 4:</u> If $\mathcal{M}$ is dominated, $\mathcal{N} \subset C(\mathcal{M})$, and T is sufficient for $\mathcal{M}$, then T is sufficient for $\mathcal{N}$.

<u>Proof:</u> Let λ be that measure equivalent to $\mathcal{M}$ whose existence is assured by Theorem 1. If $\mu \varepsilon C(\mathcal{M})$, then μ can be written

$$\mu = \sum_{i=1}^{k} \beta_i \nu_i$$

with $\beta_i > 0$, $\nu_i \varepsilon \mathcal{M}$ for $i = 1,\ldots,k$. Hence,

$$\frac{d\mu}{d\lambda} = \sum_{i=1}^{k} \beta_i \frac{d\nu i}{d\lambda} \quad (\epsilon)\ T^{-1}(\mathcal{B}).$$

Thus T is sufficient for $C(\mathcal{M})$ and hence is sufficient for $\mathcal{N}$.

In order to characterize sufficient statistics for $\mathcal{N} \subset C(\mathcal{M})$, it suffices, by Theorem 2, to consider a pair

$$\mu_I = \sum_{i\varepsilon I} \beta_i \mu_i$$

and

$$\mu_J = \sum_{j\varepsilon J} \beta_j \mu_j$$

in $\mathcal{M}$, where I and J are finite sets; $\beta_k > 0$ for $k \varepsilon I\cup J$; and the measures $\{\mu_i\}_{i\varepsilon I}$ are distinct members of $\mathcal{M}$, as are the measures $\{\mu_j\}_{j\varepsilon J}$.

The set C( ) of all finite mixtures of elements of $\mathcal{M}$ is said to be <u>identifiable</u> (Teicher, 1960, 1961; Yakowitz 1969) if each element of $C(\mathcal{M})$ can be expressed in only one way as a linear combination with positive coefficients of elements of $\mathcal{M}$, except for the order of the summands. Equivalently, $C(\mathcal{M})$ is identifiable if the set $\mathcal{M}$ is linearly independent over the real numbers.

The concept of identifiability is very important in establishing the uniqueness and consistency of various estimators of the so called <u>mixing parameters</u> $\{\beta_i : i\varepsilon I\}$ in a mixture μ_I (Yakowitz, 1969).

Given a mixture μ_I in $C(\mathcal{M})$ we have for each $E \,\varepsilon\, \mathcal{A}$, $F \,\varepsilon\, \mathcal{B}$,

$$\begin{aligned}\int_F P_{\mu_I}(E)\,d\mu_I T^{-1} &= \mu_I\,(E \cap T^{-1}(F)) \\ &= \sum_{i\varepsilon I} \beta_i \mu_i\,(E \cap T^{-1}(F)) \\ &= \sum_{i\varepsilon I} \beta_i \int_F P_{\mu_i}(E)\,d\mu_i\,T^{-1} \\ &= \sum_{i\varepsilon I} \beta_i \int_F P_{\mu_i}(E)\,\frac{d\mu_i\,T^{-1}}{d\mu_I\,T^{-1}}\,d\mu_I\,T^{-1}.\end{aligned}$$

Let $I_1,\ldots,I_r$ be the equivalence classes in I modulo the relation $i \equiv k$ if and only if $P_{\mu_i} = P_{\mu_k}$; that is, if and only if T is sufficient for the pair $\{\mu_i, \mu_k\}$. Then we have

$$\begin{aligned}&\sum_{i\varepsilon I} \beta_i \int_F P_{\mu_i}(E)\,\frac{d\mu_i T^{-1}}{d\mu_I T^{-1}}\,d\mu_I T^{-1} \\ &= \int_F \sum_{\ell=1}^{r} \sum_{i\varepsilon I_\ell} \beta_i\,\frac{d\mu_i T^{-1}}{d\mu_I T^{-1}}\,P_{\mu_{I_\ell}}(E)\,\,d\mu_I\,T^{-1},\end{aligned}$$

where $P_{\mu_{I_\ell}}(E)$ is the common value of the $P_{\mu_i}(E)$ for $i \,\varepsilon\, I_\ell$. Thus,

$$P_{\mu_I} = \sum_{\ell=1}^{r} \frac{d\mu_{I_\ell} T^{-1}}{d\mu_I T^{-1}}\,P_{\mu_{I_\ell}}$$

where μ_{I_ℓ} is the mixture

$$\mu_{I_\ell} = \sum_{i \in I_\ell} \beta_i \mu_i$$

Whenever the conditional probability function P_{μ_I} of a mixture μ_I is written in this fashion with $I_1, \ldots, I_r$ being equivalence classes modulo the relation $\equiv$, we will say that P_{μ_I} is written in <u>normal</u> form.

<u>Definition 5</u>: The set $C(\mathcal{M})$ is <u>conditionally identifiable</u> with respect to the statistic T if for each pair $\{\mu_I, \mu_J\}$ in $C(\mathcal{M})$, whenever $P_{\mu_I} = P_{\mu_J}$ and P_{μ_I}, P_{μ_J} are expressed in normal form

$$P_{\mu_I} = \sum_{\ell=1}^{r} \frac{d\mu_{I_\ell} T^{-1}}{d\mu_I T^{-1}} P_{\mu_{I_\ell}}$$

$$P_{\mu_J} = \sum_{k=1}^{s} \frac{d\mu_{J_k} T^{-1}}{d\mu_J T^{-1}} P_{\mu_{J_k}} ,$$

then $r = s$ and for each $\ell = 1, \ldots, r$ there exists exactly one $k = 1, \ldots, r$ such that $\frac{d\mu_{I_\ell} T^{-1}}{d\mu_I T^{-1}} = \frac{d\mu_{J_k} T^{-1}}{d\mu_J T^{-1}}$ and $P_{\mu_{I_\ell}} = P_{\mu_{J_k}}$. The set $C(\mathcal{M})$ is <u>marginally identifiable</u> with respect to T if the set $\{\mu T^{-1} | \mu \in \mathcal{M}\}$ is linearly independent over the real numbers.

<u>Theorem 6</u>: If $C(\mathcal{M})$ is both marginally identifiable and conditionally identifiable with respect to a statistic T, then $C(\mathcal{M})$ is identifiable.

<u>Proof</u>: Suppose $\mu_I = \sum_{i \in I} \beta_i \mu_i = \sum_{j \in J} \beta_j \mu_j = \mu_J$, where the measures in each sum are distinct members of $\mathcal{M}$. Then, expressed in normal form,

$$P_{\mu_I} = \sum_{\ell=1}^{r} \frac{d\mu_{I_\ell}T^{-1}}{d\mu_I T^{-1}} P_{\mu_{I_\ell}} = \sum_{\ell=1}^{r} \frac{d\mu_{J_\ell}T^{-1}}{d\mu_J T^{-1}} P_{\mu_{J_\ell}} = P_{\mu_J},$$

and we may assume without loss of generality that

$$\frac{d\mu_{I_\ell}T^{-1}}{d\mu_I T^{-1}} = \frac{d\mu_{J_\ell}T^{-1}}{d\mu_J T^{-1}}$$

and $$P\mu_{I_\ell} = P\mu_{J_\ell} \quad \text{for} \quad \ell = 1,\ldots,r.$$

Since $\mu_I T^{-1} = \mu_J T^{-1}$, it follows that $\mu_{I_\ell} T^{-1} = \mu_{J_\ell} T^{-1}$. For $i,k \in I_\ell$, $\mu_i T^{-1} \neq \mu_k T^{-1}$, for otherwise, since $P\mu_i = P\mu_k$, we would have $\mu_i = \mu_k$, contradicting the assumption that $\{\mu_i : i \in I\}$ are distinct. Similarly, the $\mu_j T^{-1}$ for $j \in J_\ell$ are all distinct. Since $C(\mathcal{M})$ is marginally identifiable, I_ℓ and J_ℓ have the same number of elements and for each $i \in I_\ell$ there is a unique $j(i) \in J_\ell$ such that $\beta_i = \beta_{j(i)}$ and $\mu_i T^{-1} = \mu_{j(i)} T^{-1}$. Since $P_{\mu_i} = P_{\mu_{j(i)}}$, it follows that $\mu_i = \mu_{j(i)}$ for each $i \in I_\ell$. Therefore, there is one to one map j from I onto J such that $\beta_{j(i)} = \beta_i$ and $\mu_{j(i)} = \mu_i$ for each $i \in I$. Hence, $C(\mathcal{M})$ is identifiable, and the proof is complete.

For conditionally identifiable sets of measures, the following theorem and its corollary provide some characterizations of sufficient statistics.

<u>Theorem 7</u>: If $\mathcal{M}$ is homogeneous, $C(\mathcal{M})$ is conditionally identifiable with respect to a statistic T, and μ_I, μ_J are in $C(\mathcal{M})$, then T is sufficient for the pair μ_I, μ_J if and only if there exist partitions $I = I_1 \cup \ldots \cup I_r$ and $J = J_1 \cup \ldots \cup J_r$ such that for each $\ell = 1,\ldots,r$:

(a) $$d\Big(\sum_{i\varepsilon I_\ell} \beta_i \mu_i\Big) \Big/ d\Big(\sum_{j\varepsilon J_\ell} \beta_j \mu_j\Big) = \frac{d\mu_{I_\ell}}{d\mu_{J_\ell}} = \frac{d\mu_I}{d\mu_J}$$

and

(b) T is sufficient for the set $N_\ell = \{\mu_k : k \varepsilon I_\ell \cup J_\ell\}$.

Proof: First suppose such partitions exist. By (b) T is sufficient for the set N_1 and hence, by lemma 4, it is sufficient for the pair $\{\mu_{I_1}, \mu_{J_1}\}$. It follows from (a) and Theorem 3 that T is sufficient for the pair $\{\mu_I, \mu_J\}$.

Suppose that T is sufficient for the pair $\{\mu_I, \mu_J\}$. Then, expressed in normal form,

$$\sum_{\ell=1}^{r} \frac{d\mu_{I_\ell} T^{-1}}{d\mu_I T^{-1}} P\mu_{I_\ell} = \sum_{\ell=1}^{r} \frac{d\mu_{J_\ell} T^{-1}}{d\mu_J T^{-1}} P_{\mu_{J_\ell}},$$

and we may assume without loss of generality that

$$\frac{d\mu_{I_\ell} T^{-1}}{d\mu_I T^{-1}} = \frac{d\mu_{J_\ell} T^{-1}}{d\mu_J T^{-1}} \text{ and } P_{\mu_{I_\ell}} = P_{\mu_{J_\ell}} \quad \text{for each } \ell.$$

The condition $P_{\mu_{I_\ell}} = P_{\mu_{J_\ell}}$ is equivalent to (b). By Theorem 3, there exists a representative $f \,\varepsilon\, \frac{d\mu_I}{d\mu_J}$ which is $T^{-1}(\mathcal{B})$ measureable. If $g \,\varepsilon\, \frac{d\mu_I T^{-1}}{d\mu_J T^{-1}}$, then $g \cdot T$ is $T^{-1}(\mathcal{B})$ measureable and for each $F \,\varepsilon\, \mathcal{B}$,

$$\int_{T^{-1}(F)} g \cdot T \; d\mu_J = \int_F g \; d\mu_J T^{-1} = \mu_I T^{-1}(F)$$

$$= \int_{T^{-1}(F)} f \; d\mu_J$$

It follows that $g \circ T = f$ a.e.(μ_J). Thus,

$$\frac{d\mu_I T^{-1}}{d\mu_J T^{-1}} \circ T = \{g \circ T \mid g \in \frac{d\mu_I T^{-1}}{d\mu_J T^{-1}}\} \subset \frac{d\mu_I}{d\mu_J} .$$

Since T is also sufficient for the pair $\{\mu_{I_\ell}, \mu_{J_\ell}\}$, a similar argument gives

$$\frac{d\mu_{I_\ell} T^{-1}}{d\mu_{J_\ell} T^{-1}} \circ T \subset \frac{d\mu_{I_\ell}}{d\mu_{J_\ell}}$$

for each ℓ. Since $\frac{d\mu_{I_\ell} T^{-1}}{d\mu_J T^{-1}} = \frac{d\mu_I T^{-1}}{d\mu_J T^{-1}}$ for each ℓ, it follows that (a) holds for each ℓ and the proof is complete.

<u>Corollary 8</u>: If $\mathcal{M}$ is homogeneous and $C(\mathcal{M})$ is conditionally identifiable with respect to a statistic T, then T is sufficient for a pair $\{\mu_I, \mu_J\}$ in $C(\mathcal{M})$ if and only if there exist subsets $I_1 \subset I$ and $J_1 \subset J$ such that:

(a) $$\frac{d\mu_{I_1}}{d\mu_{J_1}} = \frac{d\mu_I}{d\mu_J}$$

and

(b) T is sufficient for $N = \{\mu_k : k \in I_1 \cup J_1\}$.

<u>Proof:</u> That T sufficient implies the existence of I_1 and J_1 satisfying (a) and (b) is immediate from Theorem 7. Conversely if I_1 and J_1 satisfy

(a) and (b), then T is sufficient for μ_{I_1}, μ_{J_1} by (b) and hence, by (a), T is sufficient for μ_I, μ_J.

Given a pair of mixtures μ_I, μ_J in $C(\mathcal{M})$, we will call their likelihood ratio $\frac{d\mu_I}{d\mu_J}$ <u>indecomposable</u> if $I_1 \subset I$, $J_1 \subset J$ and $\frac{d\mu_{I_1}}{d\mu_{J_1}} = \frac{d\mu_I}{d\mu_J}$ imply $I_1 = I$ and $J_1 = J$. It is clear from Theorem 7 that if $C(\mathcal{M})$ is conditionally identifiable with respect to T and a pair of mixtures μ_I, μ_J in $C(\mathcal{M})$ have an indecomposable likelihood ratio, then T is sufficient for $\{\mu_I, \mu_J\}$ if and only if it is sufficient for $\{\mu_k : k \in I \cup J\}$. Also, it is not difficult to see that for each pair μ_I, μ_J in $C(\mathcal{M})$ there exist nonempty subsets $I_1 \subset I$ and $J_1 \subset J$ such that

$$\frac{d\mu_{I_1}}{d\mu_{J_1}} = \frac{d\mu_I}{d\mu_J}$$

and the likelihood ratio $\frac{d\mu_{I_1}}{d\mu_{J_1}}$ is indecomposable. If μ_I and μ_J represent the probability laws for two alternative hypotheses, then there would be two advantages in being able to identify subsets I_1 and J_1 satisfying these two criteria. First, the maximum likelihood decision procedure would be simplified, and second, the search for a statistic sufficient for deciding between the two hypotheses and having the property that $C(\mathcal{M})$ is conditionally identifiable could be restricted to those statistics sufficient for $\{\mu_k : I_1 \cup J_1\}$.

3. <u>Sufficient Linear Statistics for Mixtures of Normals</u>:

If $\mathcal{R}$ is a subring of the ring $\mathcal{J}$ introduced in Section 2, then with the usual definition of addition and multiplication by elements of $\mathcal{K}$ the set of all functions $\phi : \mathcal{A} \to \mathcal{J}$ is a module over $\mathcal{R}$. Thus, it is natural to consider $\mathcal{R}$-independence of a set $\mathcal{S}$ of such functions. To be precise, $\mathcal{S}$ is $\mathcal{R}$-independent if whenever $\phi_1,\ldots,\phi_m$ is a finite set of distinct elements of $\mathcal{S}$ and $\gamma_1,\ldots,\gamma_m$ are elements of $\mathcal{R}$ such that

$$\gamma_1\phi_1(E) + \ldots + \gamma_m\phi_m(E) = 0 \quad \text{for each } E \in \mathcal{A},$$

then $\gamma_1 = \ldots = \gamma_m = 0$. If $\mathcal{K}$ is a subring of $\mathcal{J}$ which contains all the bounded Radon-Nikodym derivatives $\frac{d\mu T^{-1}}{d\nu T^{-1}}$ for $\mu, \nu \in C(\mathcal{M})$, then it is clear that $\mathcal{K}$-independence of the set $\{P_\mu : \mu \in \mathcal{M}\}$ implies that $C(\mathcal{M})$ is conditionally identifiable with respect to T.

For the remainder of this section we will assume that X is $\mathbb{R}^n$, Y is $\mathbb{R}^k$ $(k \le n)$ and $T : X \to Y$ is linear and full rank. $\mathcal{A}$ and $\mathcal{B}$ are respectively, the Borel fields on $\mathbb{R}^n$ and $\mathbb{R}^k$. We also assume that each $\mu \in \mathcal{M}$ is described by a normal density function f_μ with mean m_μ and covariance Ω_μ. That is, for each $E \in \mathcal{A}$,

$$\mu(E) = \int_E f_\mu \, d\lambda_n,$$

where λ_n is Lebesgue measure on $\mathbb{R}^n$.

By a suitable choice of the coordinate system, we may represent the densities f_μ as joint density functions $f_\mu(y,z)$ on $\mathbb{R}^k \times \mathbb{R}^{n-k}$ while representing T as the projection $T(y,z) = y$. Then the marginal densities

$$g_\mu(y) = \int_{\mathbb{R}^{n-k}} f_\mu(y,z)dz$$

are normal with means Tm_μ and covariance matrices $T\Omega_\mu T^1$ (Anderson, 1958).

The conditional density functions

$$h_\mu(z \mid y) = \frac{f_\mu(y,z)}{g_\mu(y)}$$

are normal as functions of $z \in \mathbb{R}^{n-k}$ with means

(1) $$Sm_\mu + S\Omega_\mu T^1(T\Omega_\mu T^1)^{-1}(y - Tm_\mu)$$

and covariances

(2) $$S\Omega_\mu S^1 - S\Omega_\mu T^1(T\Omega_\mu T^1)^{-1}\, T\Omega_\mu S^1.$$

where S is the linear operator $S(y,z) = z$. The conditional probabilities $P_\mu(E)$ are represented by

$$P_\mu(E \mid y) = \int_{S_y(E)} h_\mu(z|y)dz \quad .$$

where $S_y(E) = \{z \in \mathbb{R}^{n-k} \mid (y,z) \in E\}$.

Theorem 9: If $\mathcal{M}$ is a family of Borel measures on $\mathbb{R}^n$ given by n-variate normal density functions and $T : \mathbb{R}^n \to \mathbb{R}^k$ is linear of rank k, then $C(\mathcal{M})$ is conditionally identifiable with respect to T.

Proof: It can readily be verified that conditional identifiability of $C(\mathcal{M})$ is not affected by the change of variables just described. If μ_I and μ_J are in $C(\mathcal{M})$, then the Radon-Nikodym derivative $\frac{d\mu_I T^{-1}}{d\mu_J T^{-1}}$ is represented by a function of the form

$$\frac{g_I(y)}{g_J(y)} = \sum_{i\varepsilon I} \beta_i g_{\mu_i}(y) \Big/ \sum_{j\varepsilon J} \beta_j g_{\mu_j}(y);$$

i.e., a ratio of mixtures of k-variate normal density functions, which is continuous. Hence, by the remarks in the first paragraph of this section, it suffices to show that the set $\{P_\mu : \mu \in \mathcal{M}\}$ of conditional density functions is $\mathcal{R}$-independent, where $\mathcal{R}$ is the subring of $\mathcal{J}$ consisting of those elements of $\mathcal{J}$ which have a continuous representative. To this end, let $P_{\mu_1}, \ldots, P_{\mu_r}$ be distinct and let $\gamma_1, \ldots, \gamma_r$ be continuous real valued functions on $\mathbb{R}^k$ such that for each $E \in \mathcal{A}$,

$$\gamma_1(y)P_{\mu_1}(E|y) + \ldots + \gamma_r(y)P_{\mu_r}(E|y) = 0$$

for almost all y. In particular, choosing for E sets of the form $\mathbb{R}^k \times K$, where K is a borel set in $\mathbb{R}^{n-k}$, we have

$$\gamma_1(y) \int_K h_{\mu_1}(z|y)dz + \ldots + \gamma_r(y) \int_K h_{\mu_r}(z|y)dz = 0$$

for almost all y. For each K, $\int_K h_{\mu_i}(z|y)dz$ is a continuous function of y. Hence,

$$\int_K (\gamma_1(y)h_{\mu_1}(z|y) + \ldots + \gamma_r(y)h_{\mu_r}(z|y|)dz = 0$$

for each $y \in \mathbb{R}^k$. It follows that

$$\gamma_1(y)h_{\mu_1}(z|y) + \ldots + \gamma_r(y)h_{\mu_r}(z|y) = 0$$

for each $y \in \mathbb{R}^k$, $z \in \mathbb{R}^{n-k}$. Let F be the set of $y \in \mathbb{R}^k$ where two or more of the conditional density functions $h_{\mu_i}(z|y)$ are equal as functions of z. It is easily seen from (1) and (2) that the Lebesque measure of F is zero. For $y \notin F$, $\{h_{\mu_1}(\cdot|y),\ldots,h_{\mu_r}(\cdot|y)\}$ is a set of distinct normal density functions of z. Hence, (Yakowitz and Spragins; 1968), they are linearly independent over the real numbers. Therefore, for $y \notin F$, $\gamma_1(y) = \ldots = \gamma_r(y) = 0$. That is, $\gamma_1 = \ldots = \gamma_r = 0$ as elements of $\mathcal{F}$. Thus, $C(\mathcal{M})$ is conditionally identifiable.

If $\mu_I = \sum_{i\varepsilon I} \beta_i \mu_i$ is in $C(\mathcal{M})$, then μ_I has a density function

$$f_{\mu_I} = \sum_{i\varepsilon I} \beta_i f_{\mu_i}$$

which is a mixture of normal density functions. The following theorem is an immediate consequence of Theorems 7 and 9.

Theorem 10: Given the assumptions of Theorem 9, the statistic T is sufficient for a pair $\{\mu_I, \mu_J\}$ in $C(\mathcal{M})$ if and only if there exist partitions $I = I_1 \cup\ldots\cup I_r$ and $J = J_1 \cup\ldots\cup J_r$ such that for each $\ell = 1,\ldots,r$,

(a) $$\sum_{i\varepsilon I_\ell} \beta_i f_{\mu_i}(x) / \sum_{j\varepsilon J_\ell} \beta_j f_{\mu_j}(x)$$

$$= \sum_{i\varepsilon I} \beta_i f_{\mu_i}(x) / \sum_{j\varepsilon J} \beta_j f_{\mu_j}(x) \quad \text{for each } x \in \mathbb{R}^n,$$

and

(b) T is sufficient for the family $\{f_{\mu_k} : k \in I_\ell \cup J_\ell\}$ of normal density functions.

There is set of purely algebraic conditions which are equivalent to (b);

namely, that the expressions

$$\Omega_{\mu_k} - \Omega_{\mu_k} T^1 (T \Omega_{\mu_k} T^1)^{-1} T \Omega_{\mu_k}$$

$$m_{\mu_k} - \Omega_{\mu_k} T^1 (T \Omega_{\mu_k} T^1)^{-1} T m_{\mu_k}$$

$$\Omega_{\mu_k} T^1 (T \Omega_{\mu_k} T^1)^{-1}$$

are all independent of $k \in I_\ell \cup J_\ell$ (Peters, Redner, and Decell; 1976).

REFERENCES

1. Anderson, T.W. (1958). An Introduction to Multivariate Statistical Analysis. John Wiley and Sons. New York.

2. Bahadur, R.R. (1954). Sufficiency and statistical decision functions. Ann. Math. Statist. 25, 423-463.

3. Halmos, P.R. and Savage L.J. (1949). Application of the Radon-Nikodym theorem to the theory of sufficient statistics. Ann. Math. Statist. 20. 225-241.

4. Peters, B.C., Redner R., and Decell, H.P. (1976). Characterizations of linear sufficient statistics. Tech. Report no. 59, Department of Mathematics , University of Houston.

5. Teicher, H. (1961). Identifiability of mixtures. Ann. Math. Statist. 32, 244-248.

6. Teicher, H. (1963). Identifiability of finite mixtures. Ann. Math. Statist. 34, 1265-1269.

7. Yakowitz, S. and Spragins, J. (1968). On the identifiability of finite mixtures. Ann. Math. Statist. 39, 209-214.

8. Yakowitz, S. (1969). A consistent estimator for the identification of finite mixtures. Ann. Math. Statist. 4D, 1728-1735.