The Annals of Statistics
1983, Vol. 11, No. 3, 984–988

IDENTIFIABILITY OF FINITE MIXTURES FOR DIRECTIONAL DATA

BY JOHN T. KENT

University of Leeds

In some problems of directional data, finite mixtures of simple distributions have been proposed as statistical models. In this paper we show that for a wide class of directional distributions, all such mixtures are identifiable.

1. Introduction. The most common distribution used to model unimodal data on the circle is the von Mises distribution. However, when modeling more complicated features such as multimodality there are two frequently used approaches. First, one can use a mixture of von Mises distributions,

$$g_1(\theta) = \sum_{i=1}^{N} \lambda_i \exp\{\kappa_i \cos(\theta - \alpha_i)\}. \tag{1.1}$$

Alternatively, one can use an exponential family obtained by adding higher order trigonometric terms to the exponent of the density,

$$g_2(\theta) = C \exp\{\sum_{j=1}^{m} \gamma_j \cos(j\theta - \beta_j)\}. \tag{1.2}$$

For an example with both (1.1) and (1.2) applied to bimodal circular data, see Mardia and Spurr (1973).

Two questions of identifiability arise here.

(i) Are finite mixtures of von Mises densities identifiable?

(ii) Are densities of the form (1.1) always distinct from densities of the form (1.2)?

One purpose of this paper is to show in a wide variety of situations that the answer to both these questions is "yes". More generally, the theorem proved below shows the identifiability of finite mixtures for a class of distributions which is wide enough to include both of the above choices. The identifiability of von Mises mixtures was first proved in Fraser, Hsu and Walker (1981), who give several references to applications of von Mises mixtures. This paper extends their results to more general distributions and more general manifolds.

The classes of probability distributions covered by this paper include the following examples: the multivariate normal distribution on R^p; the von Mises-Fisher, the Bingham and more general distributions on the unit sphere (Beran, 1979); a joint normal-von Mises distribution on the cylinder (Mardia and Sutton, 1978, and Johnson and Wehrly, 1978); a bivariate von Mises distribution on the torus (Mardia, 1975, and Jupp and Mardia, 1980); and more generally, a joint Bingham-von Mises-Fisher matrix distribution on the product of two Stiefel manifolds (author's reply in Mardia, 1975). The identifiability of finite mixtures of multivariate normals has been shown by Yakowitz and Spragins (1968) but the other examples seem to be new.

For a general survey of finite mixtures of distributions, see for example Everitt and Hand (1981) and Behboodian (1975).

The key argument in the proof of identifiability in this paper is contained in Section 3, where we show that a suitable limit in the complex plane of the ratio of two densities on the circle equals zero. The argument is similar in style to but more general than that used by Teicher (1963), Theorem 2).

2. The framework of the paper. Let M be a connected manifold which can be

Received May 1982; revised March 1983.
AMS 1980 *subject classifications.* Primary, 62E10.
Key words and phrases. Identifiability, finite mixture, directional data, Stiefel manifold.

naturally embedded in some Euclidean space, $M \subset R^p$. Let $\boldsymbol{E}(M)$ denote the family of functions on M of the form

$$\exp\{P(x)\}$$

as $P(x)$ ranges through the space of polynomials on R^p of arbitrary finite degree.

Suppose the manifold M can be described as a finite direct product of Stiefel manifolds and copies of the real line. Such manifolds are of interest in problems with directional data as the above examples illustrate. The Stiefel manifold $O(p, k)$ can be embedded in R^{pk} as the set of $p \times k$ matrices X such that $X^TX = I_k$, the k-dimensional identity matrix. When $k = p$, we add the additional restriction $\det(X) = +1$ in order to make $O(p, p)$ connected. As a special case for $k = 1$, we obtain the unit sphere in p dimensions.

Call two functions $f^{(1)}(x)$ and $f^{(2)}(x)$ "essentially distinct" if they differ by more than a constant factor. We make the following definition of identifiability for a family of functions (see Yakowitz and Spragins, 1968).

DEFINITION. A family of functions $\boldsymbol{F}$ on M is called *identifiable* if all finite sets of essentially distinct functions are linearly independent. That is, whenever $f^{(1)}(x), \cdots, f^{(N)}(x)$ are essentially distinct functions and $\lambda_1, \cdots, \lambda_N$ are real numbers such that

$$\lambda_1 f^{(1)}(x) + \cdots + \lambda_N f^{(N)}(x) = 0, \quad x \in M, \tag{2.1}$$

then necessarily $\lambda_1 = \cdots = \lambda_N = 0$.

The following theorem is the main result of this paper.

THEOREM. *Let M be a finite direct product of Stiefel manifolds and copies of the real line. Then the family of functions $\boldsymbol{E}(M)$ is identifiable.*

REMARKS.

1. Note that we have defined our concept of identifiability for *functions* in $\boldsymbol{E}(M)$. In any statistical application we will want to talk about the identifiability of *probability densities* on M with respect to some underlying measure $\mu(dx)$, where the densities are proportional to elements of $\boldsymbol{E}(M)$. Provided that the support of $\mu(dx)$ contains an open subset of M, then these two concepts of identifiability will coincide. For if a relation between densities of the form (2.1) holds almost surely with respect to $\mu(dx)$, then an analytic continuation argument will ensure that the relation (2.1) in fact holds for all $x \in M$. In particular, the exact form of the underlying measure $\mu(dx)$ is not important to us here.

2. Note that two distinct polynomials on R^p do not necessarily define distinct polynomials on M. If $P_1(x) - P_2(x) =$ constant for $x \in M$, then they define essentially the same function in $\boldsymbol{E}(M)$, but they are not necessarily equivalent on all of R^p. For example, the polynomials $P(x) \equiv 1$ and $P(x) = \{x_1^2 + x_2^2\}^3$, $x \in R^2$, are equal on the unit circle, but are not the same on all of R^2.

3. It is clear from standard linear independence arguments that the property of identifiability is closed under direct products. That is, if $\boldsymbol{F}$ and $\boldsymbol{G}$ are identifiable families of functions on two manifolds M_1 and M_2 respectively, then the space of product functions $\{f(x)g(y): f \in \boldsymbol{F}, g \in \boldsymbol{G}\}$ is identifiable on the product manifold $M_1 \times M_2$.

In view of this last remark, it is only necessary to prove the theorem on the real line and on all Stiefel manifolds. We shall start by proving it on a specific Stiefel manifold, the circle. The result then easily generalizes to the full class of Stiefel manifolds. An argument similar to that used on the circle can be used on the real line.

3. Proof on the circle $O(2,1)$. By using polar coordinates on the circle ($x_1 = \cos\theta$, $x_2 = \sin\theta$), we can get a unique representation of an element of $E(O(2,1))$ in the form

$$g(\theta) = C \exp\{\textstyle\sum_{j=1}^{m} \kappa_j \cos(j\theta - \alpha_j)\} \tag{3.1}$$

for some number $m \geq 0$. The parameters $\kappa_j \geq 0$ are uniquely determined, and if $\kappa_j > 0$ then $\alpha_j \in [0, 2\pi)$ is uniquely determined.

For any function $g(\theta)$ of the form (3.1) and a real number σ, define a *size vector* $v(\sigma) = (v_1(\sigma), \cdots, v_m(\sigma))^T$ by

$$v_j(\sigma) = \kappa_j \cos(j\sigma - \alpha_j), \quad j = 1, \cdots, m.$$

We can put a *total ordering* on size vectors as follows. For two size vectors of the same length m, say that $v^{(1)}(\sigma) > v^{(2)}(\sigma)$ if for some j with $1 \leq j \leq m$, we have

$$v_j^{(1)}(\sigma) > v_j^{(2)}(\sigma)$$

and

$$v_{j'}^{(1)}(\sigma) = v_{j'}^{(2)}(\sigma) \quad \text{for} \quad j' > j.$$

Further, by appending zeroes to the end of a shorter vector we can also compare two size vectors of unequal lengths.

On the circle the relation (2.1) takes the form

$$\lambda_1 g^{(1)}(\theta) + \cdots + \lambda_N g^{(N)}(\theta) = 0, \quad \theta \in [0, 2\pi), \tag{3.2}$$

with

$$g^{(l)}(\theta) = \exp\{\textstyle\sum_{j=1}^{m(l)} \kappa_j^{(l)} \cos(j\theta - \alpha_j)\}.$$

Since $g^{(l)}(\theta)$ is an analytic entire function of θ, we can extend (3.2) by analytic continuation to hold for all complex numbers $\theta = \sigma + i\tau$. Further, since

$$\begin{aligned}\cos(j\sigma - \alpha_j + ij\tau) &= \cos(j\sigma - \alpha_j)\cos(ij\tau) - \sin(j\sigma - \alpha_j)\sin(ij\tau) \\ &= \cos(j\sigma - \alpha_j)\cosh(j\tau) - i\sin(j\sigma - \alpha_j)\sinh(j\tau),\end{aligned}$$

it follows that

$$|g^{(l)}(\sigma + i\tau)| = \exp\{\textstyle\sum_{j=1}^{m(l)} v_j^{(l)}(\sigma)\cosh(j\tau)\}.$$

Hence it is easy to see that if for some σ, $v^{(l)}(\sigma) < v^{(l')}(\sigma)$, then

$$g^{(l)}(\sigma + i\tau)/g^{(l')}(\sigma + i\tau) \to 0 \quad \text{as} \quad \tau \to \infty.$$

Now given two components of the mixture $g^{(l)}$ and $g^{(l')}$, there exists at least one j for which $(\kappa_j^{(l)}, \alpha_j^{(l)}) \neq (\kappa_j^{(l')}, \alpha_j^{(l')})$. Hence for all but finitely many $\sigma \in [0, 2\pi)$, we have $v_j^{(l)}(\sigma) \neq v_j^{(l')}(\sigma)$, and so $v^{(l)}(\sigma) \neq v^{(l')}(\sigma)$.

Hence there is at least one σ which the $v^{(l)}(\sigma)$ are all distinct (in fact for all but finitely many σ). Choose such a σ and order the functions $g^{(l)}(\theta)$ so that $v^{(1)}(\sigma) > \cdots > v^{(N)}(\sigma)$. Dividing (3.2) by $g^{(1)}(\theta)$ with $\theta = \sigma + i\tau$ and letting $\tau \to \infty$ gives $\lambda_1 = 0$. Proceeding similarly with the remaining terms shows that all the coefficients vanish, $\lambda_1 = \cdots = \lambda_N = 0$, as required to prove identifiability.

4. Proof on a Stiefel manifold. Our objective here is to reduce the problem on a Stiefel manifold to the case of the circle and then to use the results of the last section. For notational convenience, we shall denote an element of the Stiefel manifold $O(p, k)$ as a $(p \times k)$ matrix X rather than as a vector x.

Without loss of generality we may suppose that $k = p$. For suppose first that $k < p$. Any polynomial $P_1(X_1)$ defined for $X_1 \in O(p, k)$ can be extended to $O(p, p)$ by the formula

$$P(X) = P_1(X_1), \quad X \in O(p, p),$$

where X_1 contains the first k columns of X. Thus any relation of the form (2.1) between essentially distinct functions on $O(p, k)$ gives rise to an analogous relation between essentially distinct functions on $O(p, p)$.

Hence let us take $k = p$ and suppose a relation of the sort (2.1) holds between N essentially distinct functions, $f^{(l)}(X) = \exp\{P^{(l)}(X)\}$, $l = 1, \cdots, N$, in $\boldsymbol{E}(O(p,p))$. We want to show that the coefficients vanish, $\lambda_1 = \cdots = \lambda_N = 0$.

The identity matrix I_p lies in $O(p, p)$. Without loss of generality we may choose the additive constants in the polynomials $P^{(l)}(X)$, $l = 1, \cdots, N$, so that $P^{(l)}(I_p) = 0$. Since $P^{(l)}(X)$ is an analytic function on the analytic manifold $O(p, p)$, it is determined on $O(p, p)$ by its values on any open subset in $O(p, p)$. Hence given any two distinct polynomials on $O(p, p)$, the points at which they differ must be dense in $O(p, p)$.

Therefore, the points at which $P^{(1)}(X), \cdots, P^{(N)}(X)$ take N distinct values must also be dense. Take such a point X^*.

Let $A(\theta)$ denote the 2-dimensional orthogonal matrix,

$$A(\theta) = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix},$$

for $\theta \in [0, 2\pi)$, and set $q = [p/2]$, where $[\cdot]$ denotes integer part. Define a p-dimensional block diagonal orthogonal matrix by

$$B(\theta_1, \cdots, \theta_q) = \operatorname{diag}(A(\theta_1), \cdots, A(\theta_q)) \quad (p \text{ even}),$$

$$B(\theta_1, \cdots \theta_q) = \operatorname{diag}(A(\theta_1), \cdots, A(\theta_q), 1) \quad (p \text{ odd}),$$

where if p is odd, there is also a last 1-dimensional block to be included. By a decomposition theorem for orthogonal matrices (see for example Herstein, 1964, page 306), there exists another orthogonal matrix H such that

$$X^* = HB(\theta_1^*, \cdots, \theta_q^*)H^T$$

where $\theta_i^* \in [0, 2\pi)$, i = 1, $\cdots$, q.

Consider the submanifold $M_0 = \{HB(\theta_1, \cdots, \theta_q)H^T \colon \theta_i \in [0, 2\pi),\ i = 1, \cdots, q\} \subset O(p, p)$, which is a multidimensional torus containing both I_p and X^*. Any polynomial in X can be regarded as a polynomial in $(\cos\theta_i, \sin\theta_i)$, $i = 1, \cdots, q$, on M_0. Hence the functions $f^{(l)}(X)$, $l = 1, \cdots, N$, can be regarded as essentially distinct functions in $\boldsymbol{E}(M_0)$. In view of Remark 3 in Section 2 and the results proved in Section 3, these functions must be linearly independent on M_0, and so $\lambda_1 = \cdots = \lambda_N = 0$, as required.

5. Proof on the line. Identifiability of the function space on the line, $\boldsymbol{E}(R)$, can be proved in a manner similar to Section 3. In fact the argument is simpler because there is no need to move into the complex plane. A ratio of any two distinct functions in $\boldsymbol{E}(R)$ will be either 0 or ∞ as $x \to \infty$.

REFERENCES

BEHBOODIAN, J. (1975). Structural properties and statistics of finite mixtures., in *Statistical Distributions in Scientific Work, Vol. I*, ed. by G. P. Patil et al., Riedel, Boston 103–112.

BERAN, R. (1979). Exponential models for directional data. *Ann. Statist.* **7** 1162–1178.

EVERITT, B. S., and HAND, D. J. (1981). *Finite Mixture Distributions.* Chapman and Hall, London.

FRASER, M. D., YU-SHENG HSU, and WALKER, J. J. (1981). Identifiability of finite mixtures of von Mises distributions. *Ann. Statist.* **9** 1130–1131.

HERSTEIN, I. N. (1964). *Topics in Algebra.* Blaisdell, New York.

JOHNSON, R. A., and WEHRLY, T. (1978). Some angular-linear distributions and related regression models. *J. Amer. Statist. Assoc.* **73** 602–606.

JUPP, P. E., and MARDIA, K. V. (1980). A general correlation coefficient for directional data and related regression problems. *Biometrika* **67** 163–173.

MARDIA, K. V. (1975). Statistics of directional data (with discussion). *J. Royal Statist. Soc.* **37** 349–393.

MARDIA, K. V., and SPURR, B. D. (1973). Multisample tests for multimodal and axial data. *J. Royal Statist. Soc. B* **35** 422–436.

MARDIA, K. V., and SUTTON, T. W. (1978). A model for cylindrical variables with applications. *J. Royal Statist. Soc. B* **40** 229–233.

TEICHER, H. (1963). Identifiability of finite mixtures. *Ann. Math. Statist.* **34** 1265–1269.

YAKOWITZ, S. J., and SPRAGINS, J. D. (1968). On the identifiability of finite mixtures. *Ann. Math. Statist.* **39** 209–214.

DEPARTMENT OF STATISTICS
UNIVERSITY OF LEEDS
LEEDS LS2 9JT
ENGLAND