The Annals of Statistics
1983, Vol. 11, No. 3, 783–792

THE GEOMETRY OF MIXTURE LIKELIHOODS, PART II[1]: THE EXPONENTIAL FAMILY

BY BRUCE G. LINDSAY

The Pennsylvania State University

Geometric analysis of the mixture likelihood set for univariate exponential family densities yields results which tie the number and location of support points for the nonparametric maximum likelihood estimator of the mixing distribution to sign changes in certain integrated polynomials. One corollary is a very general uniqueness theorem for the estimator.

1. Introduction. This paper continues from Part I (Lindsay, 1983) with a geometric approach to the problem of finding the nonparametric maximum likelihood estimator of a mixing distribution. The basic premise is that $\{x_1, \cdots, x_n\}$ is a random sample from the mixture density

$$f_Q(x) = \int f_\theta(x)\, dQ(\theta),$$

where $f_\theta(\cdot)$ is a density function for each value of θ and Q is an unknown probability measure on the parameter space of θ. In this paper attention will be directed to properties in the exponential family of densities.

As described in Part I the process of finding the maximum likelihood estimator $\hat{Q}$ of the unknown mixing distribution Q can be conceptually decomposed into three steps. The first step is to construct the curve Γ in K-dimensional Euclidean space consisting of all vectors of the form $(f_\theta(y_1), \cdots, f_\theta(y_K))$, where θ varies over the parameter space Ω and $\{y_1, \cdots, y_K\}$ are the distinct observations in the random sample $\{x_1, \cdots, x_n\}$. From this set we can form the mixture likelihood set; this is the set of all fitted density vectors $(f_Q(y_1), \cdots, f_Q(y_K))$ allowable under the mixture model. Under compactness of Γ this is the convex hull of the set Γ, conv(Γ).

The second step is to find the point $\hat{\mathbf{f}}$ in the mixture likelihood set conv(Γ) which maximizes the log product function $\phi(\mathbf{p}) = \Sigma\, n_k \log p_k$, where n_k is the number of times y_k appears in the sample. The last step is to identify from the fitted density vector $\hat{\mathbf{f}}$ the mixing distributions Q which satisfy $\hat{\mathbf{f}} = \mathbf{f}_Q$ and so are maximum likelihood estimators. In this part we concentrate upon a geometric understanding of the last step.

Some properties of the maximizing measures $\hat{Q}$ have been identified in Part I. The points of support in Ω of $\hat{Q}$ must be zeroes and maxima of the gradient function

$$D(\theta;\hat{Q}) = \textstyle\sum_{k=1}^{K} n_k \left\{\dfrac{f_\theta(y_k)}{f_{\hat{Q}}(y_k)} - 1\right\};$$

recall from Theorem 4.1 of Part I that it is necessary and sufficient for $\hat{Q}$ to be maximal that this function have supremum zero (0) over θ. The geometric interpretation of this statement is that the support points $\mathbf{f}_\theta$ in Γ of the maximal point $\hat{\mathbf{f}}$ lie in a particular support hyperplane of the mixture likelihood set. Addressing the statistical question as to the number of support points in the maximizing mixture is then equivalent to the geometric question of determining the number of contact points of the curve Γ with a support hyperplane of conv(Γ).

Received March 1981; revised January 1983.

[1] This research was partially supported by The National Science Foundation under Grant MCS-8003081.

AMS 1970 *subject classifications.* Primary 62A10, 62G05; secondary 52A40, 53A05.

Key words and phrases. Mixture, maximum likelihood, curvature.

It will be demonstrated in this paper that in the exponential family case there are polynomial-like functions involving the observations which describe the geometry of the curve Γ, and, in particular, provide information as to the number and location of the supports. One corollary will be that the estimator $\hat{Q}$ which satisfies $\mathbf{f}_{\hat{Q}} = \hat{\mathbf{f}}$ is unique for most exponential family cases; uniqueness issues are discussed further in Sections 5 and 6. We first approach the larger geometric questions.

2. The main theorem. A family of densities $\{f_\theta : \theta \in \Omega\}$ will be said to belong to the exponential class of densities if a typical member can be written in the form

$$f_\theta(x) = \exp(\theta x - \kappa(\theta))$$

with respect to a sigma-finite measure v. The natural parameter space Ω is the interval of $\mathbb{R}$ on which

$$\exp(\kappa(\theta)) = \int \exp(\theta x)\, dv(x)$$

is finite. The function $\mu(\theta) = \kappa'(\theta) = E_\theta[X]$ is strictly increasing and so defines a mean value reparameterization of the family of densities. Hereafter we asume that the observed values are ordered: $y_1 < y_2 < \cdots < y_K$.

In this section we present the main theorem of the paper. It presents certain key relationships between the number and location of support points to $\hat{Q}$ and certain integrated polynomials which describe the differential geometry of the curve $\Gamma = \{\mathbf{f}_\theta : \theta \in \Omega\}$. We first define the polynomials

$$p_k(x) = (x - y_1)(x - y_2) \cdots (x - y_k), \quad k = 1, 2, \cdots, K.$$

THEOREM 2.1. $(\alpha, \beta) \subset \Omega$. *Let* $\mathbf{w}$ *be a nonnegative vector.* (a) *An upper bound to the number of sign change zeroes to the vector inner product* $\langle \mathbf{w}, \mathbf{f}'_\theta \rangle$ *for* θ *in* (α, β) *is* $K - 1$ *plus the number of sign change zeroes on* (α, β) *of the mixture polynomial*

$$M(\theta) = E_\theta[p_K(X)],$$

provided $M(\theta)$ *is not identically zero.*

(b) *Moreover, if for some value of* k *both* $E_\theta[p_k(X)]$ *and* $E_\theta[p_{k-1}(X)(X - y_K)]$ *are of the same sign on* (α, β), *then an upper bound is* $k - 1$.

The complete proof of this theorem has been relegated to the appendix.

The idea of the proof is as follows. We wish to bound the number of zeroes of

$$D'(\theta; Q) = \Sigma\, w_k E_\theta(y_k - X)\exp(\theta y_k - \kappa(\theta)) \tag{2.1}$$

over all possible sets of nonnegative coefficients w_k. First divide through by $E_\theta(y_1 - X)\exp(\theta y_1 - \kappa(\theta))$ to obtain

$$\sum_{k=2}^{K} w_k \frac{E_\theta(y_k - X)}{E_\theta(y_1 - X)} \exp[\theta(y_k - y_1)] + w_1. \tag{2.2}$$

In an interval where the divisor is never zero (0), (2.2) has the same zero behavior as (2.1). Differentiating (2.2) with respect to θ yields the function

$$\sum_{k=2}^{K} w_k (y_k - y_1) \frac{E_\theta[(y_k - X)(y_1 - X)]}{\{E_\theta(y_1 - X)\}^2} \exp[\theta(y_k - y_1)]. \tag{2.3}$$

We have eliminated the coefficient w_1, and the function (2.1) can have at most one more zero in an open region where $y_1 - E_\theta(X) \neq 0$ than (2.3) has. This procedure is repeatable, the steps successively eliminating w's and introducing higher order polynomials. The chief difficulty in the proof is counting zeroes around the poles introduced by the divisors.

Before proceeding to the geometric interpretation of Theorem 2.1 in Section 4, we present in the next section some technical properties of the integrated polynomials $E_\theta[p_k(X)]$, hereafter called *mixture polynomials.*

3. Mixture polynomials. The mixture polynomials are true polynomials in the mean value parameter μ for six natural exponential families: the normal, Poisson, gamma, binomial, negative binomial, and hyperbolic secant (Morris, Theorem 3, 1982). However, they quite generally behave like polynomials in μ. In this section we aggregate some useful properties of this type. Several other results can be found as lemmas in the appendix.

LEMMA 3.1. *Let f_θ be an exponential class density with respect to measure v. Let $p_k(x) = (x - y_1) \cdots (x - y_k)$. If p_k is not the zero function almost everywhere with respect to v, then $E_\theta(p_k(X))$ satisfies the following on any open interval of Ω:*

(a) *It has no more than k zeroes counting multiplicities.*

(b) *If the measure v is discrete with support points $\mathscr{X}$ then the mixture polynomial can have no more zeroes, counting multiplicities, than the number of sign changes in the sequence $p_k(x)$ as one proceeds in order through the set $\mathscr{X}$.*

PROOF. See Karlin (1968), pages 230–240, for terminology and theorems used below.

Part(a). The function p_k can have no more than k relevant sign changes under any measure v. Since $\exp(\theta x)$ is a strictly totally positive kernel and since one can interchange integration and differentiation of all orders in the function

$$\exp(\kappa(\theta))M(\theta) = \int p_k(x)\exp(\theta x)\, d\nu(x),$$

this function falls in the domain of Karlin's Theorem 3.2, page 239.

Part (b). For a discrete measure ν the number of relevant sign changes of p_k is as given above. □

Of particular interest in the ensuing discussion is the mixture quadratic $M(\theta) = E_\theta[(X - y_1)(X - y_2)]$, which can be reexpressed in the mean value parameter as

$$M(\mu) = (y_1 - \mu)(y_2 - \mu) + \mathrm{Var}_\mu(X). \tag{3.1}$$

This function is strictly convex as a function of μ (unless identically zero). From (3.1) all zeroes must occur for μ in the range $[y_1, y_2]$. Since $M(y_1)$ and $M(y_2)$ are both nonnegative, there are two zeroes or no zeroes in this range.

The following lemma describes further quadratic-like behavior which will be used in Theorem 4.1.

LEMMA 3.2. *Suppose $y_1 \le y_2 \le y_3 \le y_4$. Then*

(a) *For $\mu \in [y_1, y_4]$,*

$$M_1(\mu) \equiv E_\mu[(X - y_2)(X - y_3)] \ge E_\mu[(X - y_1)(X - y_4)] \equiv M_2(\mu).$$

(b) *If $M_1(\mu)$ has two real roots $a_1 \le a_2$, then $M_2(\mu)$ has two real roots b_1 and b_2 which satisfy*

$$y_1 \le b_1 \le a_1 \le a_2 \le b_2 \le y_4$$

(c) *If $M_2(\mu)$ has no real roots, then neither does $M_1(\mu)$.*

PROOF. The function

$$E_\mu[(X - y_2)(X - y_3) - (X - y_1)(X - y_4)] = \mu(y_1 + y_4 - y_2 - y_3) + y_2y_3 - y_1y_4$$

is linear in μ and equals $(y_3 - y_1)(y_2 - y_1)$ and $(y_4 - y_3)(y_4 - y_2)$ at $\mu = y_1$ and $\mu = y_4$ respectively; hence it is positive for μ in the interval $[y_1, y_4]$. This gives (a). Parts (b) and (c) follow from the convexity in μ of the mixture quadratic and the nonnegativity of $M_2(\mu)$ at $\mu = y_1$ and y_4,where $M_2(\mu) = \mathrm{Var}_\mu X$. □

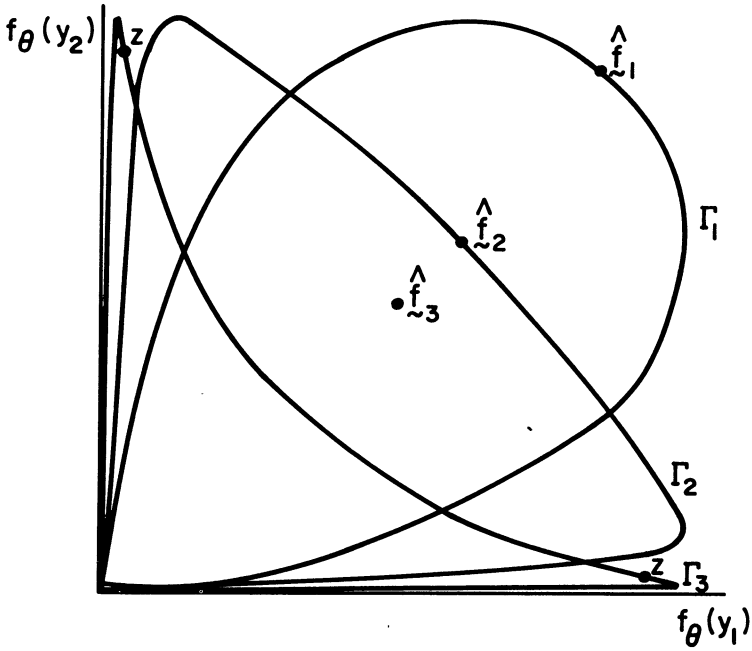

FIG. 1. *Two observations* (y_1, y_2) *from a normal density with mean* θ *and variance* 1. Γ_1: *curve for* $y_1 = -.5, y_2 = .5$; Γ_2: *curve for* $y_1 = -1, y_2 = +1$; Γ_3: *curve for* $y_1 = -1.5, y_2 = +1.5$.

4. Geometric interpretation. Our geometric interpretation of Theorem 2.1 starts with Figure 1, which represents the atomic likelihood curve Γ for a normal density with mean θ and variance 1 for three sets of values of (y_1, y_2): $(-.5, .5)$, $(-1, 1)$ and$(-1.5, 1.5)$. Theorem 2.1 embodies a simple but important characterization of the obvious and dramatic change in shape for the two dimensional ($K = 2$) atomic likelihood curve. Of critical importance in this exposition are the points indicated by the letter Z in the figure. They indicate a zero in the curvature of the curve; that is, points where the curve is neither bending left or right as one traces the curve through increasing values of θ. Between these points the curve bends monotonely left or right. If a simple closed curve $\mathbf{f}_\theta$ has no zeroes in curvature, then the boundary of conv(Γ) will be Γ itself, as in Γ_1 and Γ_2 of Figure 1. If there are zeroes in curvature, then a concavity or indentation can occur in the region bounded by Γ, as in Γ_3 of Figure 1.

We now formally introduce the notion of curvature. In the exponential family the *velocity vector* $\mathbf{f}'_\theta$ is never identically zero. At each point $\mathbf{f}_\theta$ on the curve there is a *principal normal vector* $\mathbf{n}_\theta$ which is orthogonal to $\mathbf{f}'_\theta$ and has signed length which is called the *curvature* of the curve. For $K = 2$ this curvature is

$$\det(\mathbf{f}'_\theta, \mathbf{f}''_\theta)/\langle \mathbf{f}'_\theta, \mathbf{f}'_\theta \rangle^{3/2} \tag{4.1}$$

The sign of the curvature clearly indicates the relative orientation of $(\mathbf{f}'_\theta, \mathbf{f}''_\theta)$: negative values correspond to right turns in the curve, positive values to left turns.

If we define

$$M(\theta) = \frac{f''_\theta(y_1)}{f_\theta(y_1)} \frac{f'_\theta(y_2)}{f_\theta(y_2)} - \frac{f'_\theta(y_1)}{f_\theta(y_1)} \frac{f''_\theta(y_2)}{f_\theta(y_2)}$$

then the numerator of the curvature (4.1) is equal to $f_\theta(y_1)f_\theta(y_2)M(\theta)$, so $M(\theta)$ is equivalent in sign to the curvature. Moreover, for $\{f_\theta\}$ in the exponential class we have the equivalent formulation in the mean value parameter $M(\mu) = E_\mu[(X - y_1)(X - y_2)]$. Thus the mixture quadratic contains the sign information of the curvature, and so provides an index to the convexity of the curve Γ. For values of K greater than 2 we have

$$E_\theta[(X - y_1)(X - y_2) \cdots (X - y_K)] = \det[\mathbf{f}'_\theta, \mathbf{f}''_\theta, \cdots, \mathbf{f}^{(K)}_\theta] f_\theta(y_1) \cdots f_\theta(y_K).$$

Thus zeroes in the highest order mixture polynomial correspond to linear dependence in the vectors $\{\mathbf{f}'_\theta, \mathbf{f}''_\theta, \cdots, \mathbf{f}^{(K)}_\theta\}$; at these points the curve can be embedded locally in a lower-dimensional space. For $K = 3$, these zeroes correspond to zeroes in torsion. Intuitively, sign changes again correspond to increasing complexity of the surface structure of conv(Γ).

From this geometric analysis one can describe the m.l.e. $\hat{Q}$ for $K = 2$ in the exponential family. If $M(\mu)$ has no zeroes, Γ is the boundary of conv(Γ), the mixture m.l.e. $\hat{\mathbf{f}}$ must therefore lie in Γ, and so $\hat{Q} = \delta(\hat{\mu})$, where $\hat{\mu} = \bar{x}$. If $M(\mu)$ has two zeroes, then it *may* be the case that there is a two point solution Q_2 with higher likelihood. This will certainly be true if $M(\hat{\mu})$ is negative, as then $\mathbf{f}_{\hat{\mu}}$ lies in the interior of conv(Γ), while the solution $\hat{\mathbf{f}}$ lies in the boundary. More generally, there can be no support points of $\hat{Q}$ in regions where $M(\mu)$ is negative (between its two zeroes).

Example 4.1. Let f_θ be the normal density with mean θ and variance σ^2. Here

$$M(\theta) = E_\theta[(X - y_1)(X - y_2)] = \theta^2 - (y_1 + y_2)\theta + y_1 y_2 + \sigma^2,$$

which has zeroes at

$$(z_1, z_2) = \bar{y} \pm \frac{1}{2}\sqrt{(y_2 - y_1)^2 - 4\sigma^2}, \tag{4.1}$$

provided the discriminant term in the radical is nonnegative. Thus there is a concavity in the curve if $y_2 - y_1 > 2\sigma$. Putting mass 1 at $\theta = \bar{x}$ must be the best one support-point estimator. One need only check to see if there is a superior two-point mixing distribution. If there is no concavity, then all two-point mixtures give likelihoods in the interior of conv(Γ), so the one-point estimator must be best. If there is a region of concavity then $\bar{y}$ is at its center by (4.1) and so $\mathbf{f}_{\bar{y}}$ must be in the interior of conv(Γ). If $n_1 = n_2$, then $\bar{x} = \bar{y}$, so that there must exist a superior two point estimator. In this latter case it is also geometrically clear that $\hat{\mathbf{f}}$ has one of its support points in each of the intervals (y_1, z_1) and (z_2, y_2).

Example 4.2 Let $\{f_\theta\}$ be the Poisson family of densities. In the mean value parameterization the mixture quadratic is

$$M(\mu) = \mu^2 - (y_1 + y_2 - 1)\mu + y_1 y_2,$$

which has roots

$$(z_1, z_2) = \bar{y} - \tfrac{1}{2} \pm \tfrac{1}{2}\sqrt{(y_2 - y_1)^2 - 2(y_1 + y_2) + 1}$$

provided the discriminant is nonnegative. To illustrate, suppose $y_1 = 1$. Then $\mathbf{f}_{\bar{x}}$ is the mixture maximum likelihood estimator $\hat{\mathbf{f}}$ whenever $y_2 \le 4$. If $y_1 = 1$ and $y_2 \ge 5$, then the number of support points to $\hat{\mathbf{f}}$ will depend on n_1 and n_2.

Example 4.3. For the gamma family of densities with fixed shape parameter α the mixture quadratic has two zeroes when the ratio of the two data points satisfies

$$y_2/y_1 \ge 1 + (2/\alpha) + 2(1 + \alpha)^{1/2}/\alpha,$$

the critical number for the exponential density ($\alpha = 1$) being $3 + 2\sqrt{2}$.

The following extension of Theorem 2.1 enriches the applications of the mixture quadratic to values of K greater than two.

Theorem 4.1. (a) *There can be no more than one point of support to $\hat{Q}$ in each interval where $E_\mu[(X - y_1)(X - y_K)]$ is strictly positive. If the mixture quadratic $E_\mu[(X - y_1)(X - y_K)]$ is strictly positive on $[y_1, y_K]$ then the mixture* m.l.e. *must be mass one at $\mu = \bar{x}$.*

(b) *If* $y_k < y_{k+1}$ *are adjacent data points, the mixture* m.l.e. *can have no support points between the real zeroes of the mixture quadratric* $E_\mu[(X - y_k)(X - y_{k+1})]$.

PROOF. Let $\mathbf{w} \equiv \hat{\mathbf{f}}^*$, where the $*$ operator denotes elementwise inversion, as in Part I. We can reduce the K-dimensional problem back to two dimensions by considering the trace Δ of the curve $\mathbf{g}_\theta = (g_1(\theta), g_2(\theta))$ defined by

$$g_1(\theta) = \textstyle\sum_{i=1}^{k} w_i f_\theta(y_i), \quad g_2(\theta) = \sum_{j=k+1}^{K} w_j f_\theta(y_j).$$

This curve will be shown similar in shape to the curves of Figure 1.

Since $\mathbf{w}$ defines a support hyperplane of conv(Γ) with $\langle \mathbf{w}, \mathbf{f}_\theta \rangle \le n$, $\mathbf{1} = (1, 1)^t$ defines a support hyperplane H of conv(Δ) with $\langle \mathbf{1}, \mathbf{g}_\theta \rangle \le n$. The θ-support points of $\hat{\mathbf{f}}$ satisfy $\mathbf{g}_\theta \in H$. The curvature of $\mathbf{g}_\theta$ has the same sign as

$$\begin{aligned} \det(\mathbf{g}'_\theta, \mathbf{g}''_\theta) &= \textstyle\sum_{i=1}^{k} \sum_{j=k+1}^{K} w_i w_j [f''_\theta(y_i) f'_\theta(y_j) - f'_\theta(y_i) f''_\theta(y_j)] \\ &= \textstyle\sum_{i=1}^{k} \sum_{j=k+1}^{K} w_i w_j f_\theta(y_i) f_\theta(y_j) E_\theta[(X - y_i)(X - y_j)]. \end{aligned} \tag{4.2}$$

Also

$$\det(\mathbf{g}_\theta, \mathbf{g}'_\theta) = \textstyle\sum_{i=1}^{k} \sum_{j=k+1}^{K} w_i w_j f_\theta(y_i) f_\theta(y_j)[y_j - y_i] > 0, \tag{4.3}$$

so that the vectors $(\mathbf{g}_\theta, \mathbf{g}'_\theta)$ always form a right-hand coordinate system and hence, as θ increases, the point $\mathbf{g}_\theta$ always traces to the left as viewed from the origin.

It is now clear geometrically because hyperplane H sits above the trace Δ that in any interval of negative curvature there can be no points $\mathbf{g}_\theta$ in H; also, in any interval of positive curvature there is at most one such point. Part (a) of the theorem then follows from Lemma 3.2 applied to (4.2). Part (b) follows from Lemma 3.2 as all terms in (4.2) are negative between the roots of $E_\mu[(X - y_k)(X - y_{k+1})]$. □

5. Uniqueness. The uniqueness of the maximal mixing distribution $\hat{Q}$ in the exponential family depends upon whether the observed values $\{y_1, y_2, \cdots, y_K\}$ saturate the support of the generating measure ν. If they do, then the likelihood curve Γ lies entirely in the hyperplane of $\mathbb{R}^K$ where the coordinates sum to 1. The mixture likelihood set has a nonempty interior within this hyperplane and if the multinomial maximum likelihood estimator $\hat{\mathbf{p}} = (n_1/n, n_2/n, \cdots, n_K/n)$ falls within this interior, the mixing distribution specification $\hat{Q}$ cannot be unique. This case was discussed for the binomial in Lindsay (1981) and will be treated further elsewhere. In the unsaturated case, the machinery has already been produced to verify that the mixture estimator $\hat{Q}$ is unique.

THEOREM 5.1. *Let* f_θ *be an exponential class density with respect to measure* ν. *The maximum likelihood estimator* $\hat{Q}$ *of the mixing distribution is unique provided the function* $p_K(x)$ *is not zero almost-everywhere* (ν) *and provided there are no support points of* $\hat{\mathbf{f}}$ *which come from the boundary of* Ω.

PROOF. Within the interior of the parameter space Ω we have at most $2K - 1$ zeroes to the function $D'(\theta; \hat{Q})$ by Lemma 3.1 and Theorem 2.1. Hence there are at most K maximal zeroes to $D(\theta; \hat{Q})$; which proves the theorem by the arguments of Section 8, Part I. □

Consideration of the limit points of $\{\mathbf{f}_\theta\}$ adds a new set of technical difficulties to the problem. To minimize these, we restrict attention to the regular case of the exponential family (Barndorff-Nielsen, 1980, page 114); that is, when Ω is open. Most familiar densities fit this description. In this case $f_\theta(y) \to 0$ as θ approaches either boundary of Ω unless y is itself a boundary point of the support $\mathscr{X}$ of measure ν. Thus if no $\mathscr{X}$-boundary points are sampled, $\mathbf{f}_\theta \to \mathbf{0}$ as θ approaches its boundaries, and these limit points $(\mathbf{0})$ cannot be support points of $\hat{\mathbf{f}}$.

Although this essentially disposes of continuous regular exponential families, in the discrete case the boundaries of $\mathscr{X}$ may be sampled with positive probability, as in the Poisson ($x = 0$) or the binomial ($x = 0$ or $x =$ parameter n). Proving uniqueness in such cases is beyond the scope of the methods developed in this paper, and so will be postponed until a more detailed treatment of the discrete exponential family mixture can be given.

6. Further problems. It is clear that for reasons of economy of interpretation and computation, it is often desirable to obtain an estimated mixing distribution Q_m with fixed support size m. This raises several issues related to the earlier sections.

First, is there an identifiability problem? That is, is there possibly another mixture Q with support size less than or equal to m such that $\mathbf{f}_{Q_m} = \mathbf{f}_Q$? The answer is no for mixtures of sufficiently low support size.

LEMMA 6.1. *Let f_θ be an exponential class density, with observation set* $\{y_1, y_2, \cdots, y_K\}$. *Then every element* $\mathbf{f}$ *of* $\operatorname{conv}(\Gamma)$ *which has a representation* $\mathbf{f} = \mathbf{f}_Q$, *where Q has no more than $K/2$ points of support, has a unique such representation.*

PROOF. If $\mathbf{f} = \mathbf{f}_{Q_1} = \mathbf{f}_{Q_2}$, where Q_1 and Q_2 are of support size $\le K/2$, then $\mathbf{f} = \mathbf{f}_{\pi Q_1 + (1-\pi)Q_2}$ for $\pi \in [0, 1]$. The measure $\pi Q_1 + (1 - \pi)Q_2$ has no more than K points of support, say $\theta_1, \theta_2, \cdots, \theta_t$, so by the linear independence of $\{\mathbf{f}_{\theta_1}, \cdots, \mathbf{f}_{\theta_t}\}$, there is a unique set of weights π such that $\Sigma \pi_j \mathbf{f}_{\theta_j} = \mathbf{f}$. Hence $Q_1 = Q_2$.

A second point in regard to m-point mixtures is that since the support points of any local maximum Q_m are all double zeroes of $D(\theta; Q_m)$ (Part I, Theorem 7.1) the results of Theorem 2.1 also apply to bounding support sizes of Q_m.

APPENDIX I

PROOF OF THEOREM 2.1. Given the ordered values $(y_1, \cdots, y_K)$ let

$$p_m(x) = (x - y_1) \cdots (x - y_m), \quad m = 1, 2, \cdots, K.$$

For a measurable function g define the integral operator

$$I_\theta[g(X)] = I_\theta(g) = \int g(x) \exp(\theta x)\, d\nu(x).$$

Notice that $E_\theta[g(X)] = I_\theta(g)/I_\theta(1)$.

Define the following functions:

$$G_m(\theta) = \textstyle\sum_{k=m}^K w_k I_\theta[p_{m-1}(X)(X - y_k)] \exp(\theta y_k),$$

$$G^*_{m+1}(\theta) = \textstyle\sum_{k=m+1}^K w^*_k I_\theta[p_m(X)(X - y_k)] \exp(\theta y_k).$$

where $w^*_k = w_k(y_k - y_m) \ge 0$.

It will be argued that G_m has at most one more sign change zero in (α, β) than G^*_{m+1} for any nonnegative numbers $w_m, \cdots, w_K$. Inductive application of this argument starting with the function

$$G_1(\theta) = -\exp(2\kappa(\theta)) \langle \mathbf{w}, \mathbf{f}'_\theta \rangle = \textstyle\sum_{k=1}^K w_k I_\theta(X - y_k) \exp(\theta y_k)$$

yields the desired result for part (a) of Theorem 2.1.

The following lemma accumulates some elementary facts about mixture polynomials. The proof is straightforward.

LEMMA A.

(1) $\dfrac{d}{d\theta} I_\theta[p_m] = I_\theta[X p_m]$;

(2) $\dfrac{d}{d\theta} \{I_\theta[p_m] \exp(-\theta y)\} = I_\theta[(X - y)p_m] \exp(-\theta y)$;

(3) *If Z is a zero of order $k \geq 2$ for $I_\theta[p_m]$ it is a zero of order $k - 1$ for $I_\theta[p_m(X - y)]$ for all y, and the first nonzero derivative of $I_\theta[p_m(X - y)]$ is*

$$\frac{d^{k-1}}{d\theta^{k-1}} I_\theta[p_m(X-y)]\,|_{\theta=Z} = I_Z[X^k p_m]$$

(4) *If Z is a zero of order $k \geq 2$ for $I_\theta[p_{m-1}]$ it is a zero of order $k - 1$ for G_m and $k - 2$ for G^*_{m+1}.*

Next we define the functions

$$g(\theta) \equiv w_m I_\theta[p_m] \exp(\theta y_m)$$

$$h(\theta) \equiv \textstyle\sum_{k=m+1}^{K} w_k I_\theta[p_{m-1}(X - y_k)] \exp(\theta y_k),$$

noting that $g + h = G_m$.

LEMMA B. *The functions g and h above satisfy the following*
(1) *At a zero Z of g or of h one has*

$$g(Z) < h(Z) \quad \text{if} \quad I_Z[p_{m-1}] < 0$$

$$g(Z) > h(Z) \quad \text{if} \quad I_Z[p_{m-1}] > 0$$

$$0 = g(Z) = h(Z) \quad \text{if} \quad I_Z[p_{m-1}] = 0$$

(2) *Simultaneous zeroes of g and h occur only when $I_\theta[p_{m-1}]$ has a zero of order $k \geq 2$ in which case $g + h$ has a zero of order $k - 1$. The functions g and h have the same sign in a neighborhood of such a point.*

PROOF. The ordering relationships

$$\text{(I.1)} \quad \begin{aligned} &I_\theta[p_{m-1}(X - y_m)] < \cdots < I_\theta[p_{m-1}(X - y_K)] \quad \text{if} \quad I_\theta[p_{m-1}] < 0 \\ &I_\theta[p_{m-1}(X - y_m)] > \cdots > I_\theta[p_{m-1}(X - y_K)] \quad \text{if} \quad I_\theta[p_{m-1}] > 0 \end{aligned}$$

plus the positivity of the w_r's imply part (1).

From part (1), the simultaneous zeroes of g and h are zeroes of $I_\theta[p_{m-1}]$ and also zeroes of its derivative $I_\theta[Xp_{m-1}]$, hence a zero of order two or more for $I_\theta[p_{m-1}]$. The rest of part (2) follows from Lemma A(3) and A(4). □

The next step is to relate the zeroes of $g + h$ to those of

$$\text{(I.2)} \quad d(\theta) \equiv g(\theta)h'(\theta) - g'(\theta)h(\theta) = w_m g^*_{m+1}(\theta) I_\theta[p_{m-1}] \exp(\theta y_m)$$

Label the open intervals where $gh > 0$ as *even* intervals; label the open intervals where $gh < 0$ as *odd* intervals. Label those points where g and h have simultaneous odd zeroes as degenerate odd intervals.

LEMMA C. *The above described intervals have the following properties*:
(1) *With the exclusion of points which are not zeroes of $g + h$, the region (α, β) is covered by alternating even and odd intervals.*
(2) *In every odd interval, $g + h$ has no more than one more sign change zero than d.*
(3) *In every even interval, $g + h$ has no zeroes.*

PROOF. The sign change zeroes of g and h create a partition of (α, β) such that between points of sign change, g and h are of constant sign. Lemma B(2) ensures that the odd intervals (possibly degenerate) separate even intervals. Any points not in the even or odd intervals are sign changes of g, but not h, or vice-versa. Hence $g + h$ is non-zero there.

In a degenerate odd interval, $g + h$ has an odd order zero, but d has a zero which is even order (see (I.2), Lemma B(2) and Lemma A(4).) Hence part (2) of the lemma is satisfied for the degenerate case.

Next, the open odd interval. With the exception of zeroes of g the following factorization holds

$$g + h = g[1 + h/g] \equiv g[r].$$

The function in brackets (r) has the same sign changes as $g + h$ between zeroes of g. There are no common zeroes to g and h in a nondegenerate odd interval; there may, however, be even order zeroes of one or the other. These even order zeroes are not zeroes of $g + h$, but they are odd order zeroes of d. For example, at a second order zero Z of g,

$$d(Z) = g(Z)h'(Z) - g'(Z)h(Z) = 0$$

$$d'(Z) = g(Z)h''(Z) - g''(Z)h(Z) = -g''(Z)h(Z) \neq 0$$

It follows that if we show that r and hence $g + h$ has no more than one more sign change zero than d on each of the open subintervals between zeroes of g and h, then there is a maximum excess of one zero for $g + h$ over d in the entire odd interval. The function r is analytic on the open subinterval and so it has at most one more sign change that its derivative.

$$r' = d/g^2,$$

as was to be shown.

Part (3) of the lemma follows because g and h have the same sign on an even interval. □

We now finish the proof of the theorem. In an open odd interval, $I_\theta[p_{m-1}]$ can have no zeroes, as at such a zero point g and h have the same sign. In a degenerate odd interval, $I_\theta[p_{m-1}]$ has an even zero. Hence the odd zeroes of d in an odd interval are also odd zeroes of G^*_{m+1} (Equation I.1). Hence Lemma C(2) indicates that G_m can have no more than one more sign change than G^*_{m+1} in each odd interval.

Next, we establish that G^*_{m+1} has at least one more sign change zero than G_m on every even interval except those with α or β as endpoints, where it has at least the same number. Since G_m has no sign change on even intervals, we need only count the zeroes of G^*_{m+1}.

We consider several cases. If both endpoints of the even interval are sign change zeroes of just one of g and h, then Lemma B(1) indicates $I_\theta[p_{m-1}]$ has an even number of sign changes. But $d = gh' - g'h$ must have an odd number, hence so must G^*_{m+1}. If the endpoints are sign changes, one from g alone and one from h alone, then $I_\theta[p_{m-1}]$ has an odd number of sign changes, but d must have an even number, so G^*_{m+1} has an odd number.

The argument must be executed with greater care if one or both endpoints are simultaneous sign change zeroes of g and h. The idea of the argument is that at such a zero, $I_\theta[p_{m-1}]$ has an even order zero, so locally $g > h$ or $h > g$. The function G^*_{m+1} has the same sign at (and near) such a zero as g'. One may conclude that treating the zero as if it were a zero of only one of the functions g and h, namely the one nearest the X-axis, yields a sign-counting argument as in the previous paragraph.

If we now proceed across the interval (α, β), the alternation of even and odd intervals and the relative numbers of zeroes of G_m and G^*_{m+1} in each give a maximum of one more zero for G_m than G^*_{m+1}. This gives part (a). Part (b) follows from the inductive argument plus the fact that the ordering relationship (I.1) ensures that G_k has no zeroes on an interval where the given mixture polynomials have the same sign.

REFERENCES

BARNDORFF-NIELSEN, O. (1978). *Information and Exponential Families in Statistical Theory.* Wiley, New York.

KARLIN, S. (1968). *Total Positivity.* Stanford University Press, Stanford, California.

LINDSAY, B. G. (1981). Properties of the maximum likelihood estimator of a mixing distribution. *Statistical Distributions in Scientific Work*, Vol. 5. C. Taillie *et al.* (eds.), 95–109.

LINDSAY, B. G. (1983). The geometry of mixture likelihoods, part I: a general theory. *Ann. Statist.* **11** 86–94.

MORRIS, C. N. (1983). Natural exponential families with quadratic variance functions. *Ann. Statist.* **11** 515–529.

SIMAR, L. (1976). Maximum likelihood estimation of a compound Poisson process. *Ann. Statist.* **4** 1200–1209.

DEPARTMENT OF STATISTICS
COLLEGE OF SCIENCE
THE PENNSYLVANIA STATE UNIVERSITY
219 POND AVENUE
UNIVERSITY PARK, PENNSYLVANIA 16802