Identifiability of Finite Mixtures of Elliptical Distributions

HAJO HOLZMANN and AXEL MUNK

Institut für Mathematische Stochastik, University of Göttingen

TILMANN GNEITING

Department of Statistics, University of Washington

ABSTRACT. We present general results on the identifiability of finite mixtures of elliptical distributions under conditions on the characteristic generators or density generators. Examples include the multivariate *t*-distribution, symmetric stable laws, exponential power and Kotz distributions. In each case, the shape parameter is allowed to vary in the mixture, in addition to the location vector and the scatter matrix. Furthermore, we discuss the identifiability of finite mixtures of elliptical densities with generators that correspond to scale mixtures of normal distributions.

Key words: characteristic function, elliptically symmetric, finite mixture, identifiability, Laplace transform, multivariate *t* distribution, normal scale mixture

1. Introduction

Finite mixture distributions have been widely used in the statistical and general scientific literature, both as tools for modelling population heterogeneity and as a flexible method for relaxing parametric distributional assumptions. Comprehensive surveys are available in Titterington *et al*. (1985), Lindsay (1995), Böhning (2000) and McLachlan & Peel (2000). The assumption of identifiability for a statistical model lies at the heart of most statistical theory and practice (Titterington *et al*., 1985, p. 35). In the context of finite mixture models, identifiability allows for the recovery of the mixing distribution from the mixture and is essential for consistent estimation (Leroux, 1992). Different variants of the identifiability problem arise in specific mixture models used in practice (Link, 2003; Holzmann *et al*., 2006).

Teicher (1961) pioneered the study of identifiability for finite mixture distributions, and identifiability has been proved in numerous special cases since. Nevertheless, identifiability often is tacitly assumed to hold while proofs remain unavailable (Titterington *et al*., 1985, section 3.1.3; Lindsay *et al*., 2004, p. 396), and the literature appears to be incomplete in two ways. First, the number of parameters allowed to vary in the mixture is limited. Typically, a single location parameter or univariate location and scale parameters are considered only (Holzmann *et al*., 2004). However, it is often desirable to allow additional shape parameters to vary with the mixture components as well. Secondly, results for multivariate distributions are generally unavailable, with a few but notable exceptions. Yakowitz & Spragins (1968) showed that finite mixtures of multivariate normal distributions with variable mean vectors and covariance matrices are identifiable, and Kent (1983) proved a general result on the identifiability of finite mixtures of distributions on Stiefel manifolds. Chandra (1977) studied identifiability of finite mixtures on general measurable spaces.

Our work addresses both limitations simultaneously, by considering finite mixtures of elliptical densities of the form

$$f_{\alpha,p}(x) = |\Sigma|^{-1/2} f_p\left[(x-\mu)^{\mathrm{T}} \Sigma^{-1}(x-\mu); \theta\right], \quad x \in \mathbb{R}^p, \quad \alpha = (\theta, \mu, \Sigma) \in \mathcal{A}^p. \tag{1}$$

Here, $f_p(\cdot; \theta)$ is a density generator, that is, a non-negative function on $[0, \infty)$ or $(0, \infty)$ such that the spherically symmetric function $f_p(x^T x; \theta)$, $x \in \mathbb{R}^p$, integrates to 1. The parameter vector $\alpha = (\theta, \mu, \Sigma)$ belongs to $\mathcal{A}^p \subset \mathbb{R}^{k \times p \times p(p+1)/2}$ where $\mu \in \mathbb{R}^p$ is a location parameter, Σ is a positive definite scatter matrix, $|\Sigma|$ denotes its determinant, and $\theta \in \Theta \subseteq \mathbb{R}^k$ may be an additional shape parameter which may also be allowed to vary in the mixture. Finite mixtures of elliptical densities play crucial roles in model-based clustering, discriminant analysis, density estimation and other applications (Fraley & Raftery, 2002).

Elliptically symmetric densities can be defined via characteristic generators or via density generators (Kelker, 1970; Cambanis *et al.*, 1981; Fang *et al.*, 1990). Specifically, a function $\phi(u)$, $u \geq 0$, is a characteristic generator in dimension $p \geq 1$ if $\phi(t^T t)$ is the characteristic function of a probability distribution on $\mathbb{R}^p$. A characteristic generator in dimension p is a characteristic generator in dimension p' if $p' \leq p$, but not necessarily if $p' > p$. If the probability distribution in $\mathbb{R}^p$, determined by $\phi(t^T t)$, has a density, it is of the form $f_p(x^T x)$. The density generator f_p depends on the dimension, p, and generates a location-scatter family of elliptically symmetric densities in $\mathbb{R}^p$.

The paper is organized as follows. In section 2, we present identifiability theorems for elliptically symmetric densities, defined via characteristic generators or via density generators. When expressed in terms of characteristic functions, identifiability questions for elliptical distributions reduce to univariate problems, and this allows us to work with tail conditions on the characteristic generators. In terms of density generators, we prove identifiability under tail conditions, or smoothness conditions at the origin. These results extend and complete previous work by Teicher (1961, 1963), Yakowitz & Spragins (1968), Kent (1983) and Holzmann *et al.* (2004).

Section 3 turns to examples. In particular, we show that finite location-scatter mixtures from the multivariate *t*-distribution, even with variable degree of freedom, are identifiable. Peel & McLachlan (2000) proposed the use of mixtures of this type to robustify cluster analysis and discussed maximum likelihood estimation via the expectation maximization (EM) algorithm in such models. We also prove identifiability of mixtures from multivariate stable laws, where the characteristic exponent takes the role of the variable shape parameter, and we consider finite mixtures of Kotz type distributions. These results are summarized in Table 1. Further examples for identifiable finite location-scatter mixtures include generalized hyperbolic secant and Bessel distributions.

Section 4 considers finite location-scatter mixtures of elliptical distributions with completely monotone characteristic generators, corresponding to scale mixtures of normal distributions. Here, a condition for identifiability can be stated in terms of the tail behaviour of the Laplace transform of the mixing measure in the scale mixture. We extend the respective Abelian theory which links the tail behaviour of the Laplace transform to the behaviour of the mixing distribution at the origin, and this result might be of independent interest. The paper closes with a discussion in section 5. All proofs are deferred to the appendix.

Table 1. *Examples of location-scatter families in $\mathbb{R}^p$ with identifiable finite mixtures*

Distribution	Additional shape parameters	Shape parameter in mixture
Multivariate *t*	Degrees of freedom $\theta > 0$	Variable
Symmetric stable	Characteristic exponent $0 < \alpha \leq 2$	Variable
Kotz	Shape parameter $r > -p/2$	Variable
Exponential power	Exponent $s > 0$	Variable with $s \geq 1$

2. Identifiability theorems

Finite mixtures are said to be identifiable if distinct mixing distributions with finite support correspond to distinct mixtures. In the present context, finite mixtures from the location-scatter family $\{f_{\alpha, p} : \alpha = (\theta, \mu, \Sigma) \in \mathcal{A}^p\}$ are identifiable if a relation of the form

$$\sum_{j=1}^{m} \lambda_j f_{\alpha_j, p}(x) = \sum_{j=1}^{m} \lambda_j' f_{\alpha_j', p}(x), \quad x \in \mathbb{R}^p, \tag{2}$$

where m is a positive integer, $\sum_{j=1}^{m} \lambda_j = \sum_{j=1}^{m} \lambda_j' = 1$ and $\lambda_j, \lambda_j' \geq 0$ for $j = 1, \ldots, m$, implies that there exists a permutation $\sigma \in S_m$ such that $(\lambda_j, \alpha_j) = (\lambda_{\sigma(j)}', \alpha_{\sigma(j)}')$ for all j. Evidently, finite mixtures are identifiable if the family $\{f_{\alpha, p} : \alpha = (\theta, \mu, \Sigma) \in \mathcal{A}^p\}$ is linearly independent. A classical result of Yakowitz & Spragins (1968) states that this linear independence is a necessary and sufficient condition for identifiability.

2.1. Identifiability via characteristic generators

Let $\phi(u; \theta)$, where $u \geq 0$ and $\theta \in \Theta$, be a parametric family of characteristic generators, giving rise to spherically symmetric characteristic functions $\phi(t^{\mathrm{T}} t; \theta)$, $t \in \mathbb{R}^p$, where $1 \leq p < q$ and $1 \leq q \leq \infty$. The positive integer q gives an upper bound on the permissible dimensions in which $\phi(u; \theta)$ is a characteristic generator, and might be infinite. We suppose that the associated distributions have spherically symmetric densities $f_p(x^{\mathrm{T}} x; \theta)$, $x \in \mathbb{R}^p$, and recall that the density generator, f_p, depends on the dimension, p. The respective family $\{f_{\alpha, p} : \alpha = (\theta, \mu, \Sigma) \in \mathcal{A}^p\}$ of elliptical densities is defined in (1).

   The following result shows that identifiability reduces to a univariate problem if the distributions are defined via characteristic generators.

Lemma 1
Suppose that a parametric family of characteristic generators gives rise to families of elliptical densities $\{f_{\alpha, p} : \alpha = (\theta, \mu, \Sigma) \in \mathcal{A}^p\}$ in dimension $1 \leq p < q$. Then the identifiability of finite mixtures from the univariate family $\{f_{\alpha, 1} : \alpha \in \mathcal{A}^1\}$ implies the identifiability of finite mixtures from the family $\{f_{\alpha, p} : \alpha \in \mathcal{A}^p\}$ for each $1 \leq p < q$.

Next we give a sufficient condition for identifiability. To this end, we introduce a scale parameter, $a > 0$, and consider the family

$$\phi_\beta(u) = \phi(au; \theta), \quad \beta = (a, \theta) \in (0, \infty) \times \Theta =: \mathcal{B}. \tag{3}$$

Theorem 1
Suppose that a parametric family of characteristic generators gives rise to families of elliptical densities $\{f_{\alpha, p} : \alpha = (\theta, \mu, \Sigma) \in \mathcal{A}^p\}$ in dimension $1 \leq p < q$. Suppose there exists a total ordering $\preceq$ on the set $\mathcal{B}$ such that $\beta_1 \prec \beta_2$ implies

$$\lim_{u \to \infty} \frac{\phi_{\beta_2}(u)}{\phi_{\beta_1}(u)} \to 0 \tag{4}$$

for the corresponding characteristic generators. Then finite mixtures from the class $\{f_{\alpha, p} : \alpha = (\theta, \mu, \Sigma) \in \mathcal{A}^p\}$ of elliptical distributions in $\mathbb{R}^p$ are identifiable for each $1 \leq p < q$.

Note that the shape parameter is allowed to vary in the mixture, in addition to the location parameter and the scatter matrix. Pure location-scatter mixtures are covered by the special case in which Θ is trivial.

2.2. Identifiability via density generators

Typically, identifiability of finite mixtures is proved via integral transforms such as the Fourier or Laplace transform (Teicher, 1961). However, it is often more convenient to argue in terms of densities, since the corresponding transforms may not be known in closed form or may even not exist. A prominent example was given by Kent (1983), who argued directly via the densities of certain distributions on Stiefel manifolds.

In this section, we present conditions on density generators that guarantee the identifiability of finite mixtures from the associated elliptical distributions. Let $f_p(u; \theta)$, where $u \in [0, \infty)$ or $(0, \infty)$ and $\theta \in \Theta$, be a parametric family of density generators on $\mathbb{R}^p$, that is, $f_p(x^{\mathrm{T}} x; \theta)$ is non-negative and has unit integral over $x \in \mathbb{R}^p$, for all $\theta \in \Theta$. The associated location-scatter family $\{f_{\alpha, p} : \alpha = (\theta, \mu, \Sigma) \in \mathcal{A}^p\}$ is given by (1).

Theorem 2

Let $f_p(\cdot; \theta)$, $\theta \in \Theta$, be a parametric family of density generators for spherically symmetric distributions in $\mathbb{R}^p$. Let $\mathcal{C} = \Theta \times (0, \infty) \times \mathbb{R}$ and let $\gamma_j = (\theta_j, a_j, b_j) \in \mathcal{C}$ for $j = 1, 2$. Suppose there exists a total ordering $\preceq$ on the set $\mathcal{C}$ such that $\gamma_1 \prec \gamma_2$ implies

$$\lim_{u \to \infty} \frac{f_p(a_2 u^2 + b_2 u + c_2; \theta_2)}{f_p(a_1 u^2 + b_1 u + c_1; \theta_1)} = 0 \quad \text{for} \quad c_1, c_2 \in \mathbb{R}. \tag{5}$$

Then finite mixtures from the family $\{f_{\alpha, p} : \alpha = (\theta, \mu, \Sigma) \in \mathcal{A}^p\}$ of elliptical distributions in $\mathbb{R}^p$ are identifiable.

The following result concerns location-scatter mixtures for density generators that lack smoothness at the origin.

Theorem 3

Let f_p be a density generator for a spherically symmetric distribution in $\mathbb{R}^p$. Suppose that f_p has continuous derivatives of any order everywhere on $(0, \infty)$, and suppose that there exists a positive integer k_0 such that the function $h(u) = f_p(u^2)$, $u \in \mathbb{R} \setminus \{0\}$, satisfies

$$h^{(2k-1)}(0+) \neq h^{(2k-1)}(0-) \quad \text{for} \quad k = k_0, k_0 + 1, \ldots \tag{6}$$

Then finite mixtures from the location-scatter family $\{f_{\alpha, p} : \alpha = (\mu, \Sigma)\}$ are identifiable.

3. Examples

Example 1 (Multivariate t-distribution). The multivariate t-distribution with location parameter μ, scatter matrix Σ and $\theta > 0$ degrees of freedom has density function

$$f_{\alpha, p}(x) = \frac{\Gamma(\frac{\theta + p}{2})}{\pi^{p/2} \Gamma(\theta/2) |\Sigma|^{1/2}} \left[1 + (x - \mu)^{\mathrm{T}} \Sigma^{-1} (x - \mu) \right]^{-(\theta + p)/2}, \quad x \in \mathbb{R}^p,$$

where $\alpha = (\theta, \mu, \Sigma) \in \mathcal{A}^p$. Finite mixtures from the multivariate t-distribution are identifiable, even with variable degree of freedom. Indeed, the respective characteristic generator (Kotz & Nadarajah, 2004, p. 39) is given by

$$\phi(u; \theta) = \frac{(u\theta)^{\theta/4}}{2^{\theta/2 - 1} \Gamma(\theta/2)} K_{\theta/2}\left(\sqrt{u\theta}\right), \quad u \geq 0, \tag{7}$$

where $K_{\theta/2}$ denotes the McDonald or modified Bessel function of the second kind (Watson, 1944, p. 78). A well-known estimate (Watson, 1944, p. 202) implies that

$$\frac{\phi_{\beta_2}(u)}{\phi_{\beta_1}(u)} \propto e^{-\sqrt{u}(\sqrt{a_2\theta_2} - \sqrt{a_1\theta_1})} u^{(\theta_2-\theta_1)/4} \quad \text{as} \quad u \to \infty, \tag{8}$$

where $\beta_j = (a_j, \theta_j)$ for $j = 1, 2$, and $\propto$ means that the quotient converges to a positive constant. We define a total ordering on the set $\mathcal{B}$, specified in (3), as follows: $\beta_1 \prec \beta_2$ if $a_1\theta_1 < a_2\theta_2$, or if $a_1\theta_1 = a_2\theta_2$ and $\theta_1 > \theta_2$. With this total ordering, the asymptotic relationship (8) implies (4), and theorem 1 applies. A straightforward extension of the argument shows that finite mixtures of elliptical distributions from the union of multivariate t and normal distributions are also identifiable.

Example 2 (Symmetric stable law). The characteristic generator of the multivariate symmetric stable law with characteristic exponent $0 < \theta \leq 2$ is given by

$$\phi(u; \theta) = \exp\left(-u^{\theta/2}\right), \quad u \geq 0.$$

To show that finite mixtures from the associated family of elliptically symmetric densities are identifiable, let $\beta_j = (a_j, \theta_j)$ for $j = 1, 2$ and define a total ordering on the set $\mathcal{B}$, specified in (3), as follows: $\beta_1 \prec \beta_2$ if $\theta_1 < \theta_2$, or if $\theta_1 = \theta_2$ and $a_1 < a_2$. With this total ordering, theorem 1 applies.

Example 3 (Kotz type distributions). Kotz type distributions in $\mathbb{R}^p$ (Fang *et al.*, 1990, p. 76) are defined via the density generators

$$f_p(u; \theta) = \frac{s\Gamma(p/2)}{\pi^{p/2}\Gamma\left[\frac{2r+p}{2s}\right]} u^r \exp\left(-u^s\right), \quad u > 0,$$

where $\theta = (r, s) \in (-p/2, \infty) \times (0, \infty)$. This flexible family includes the exponential power distribution (Box & Tiao, 1992, p. 157), the original Kotz distribution (Fang *et al.*, 1990, p. 76) and the multivariate normal law as the special cases in which $r = 0$, $s = 1$ and $(r, s) = (0, 1)$, respectively. When compared with Gaussian densities, Kotz type distributions allow for lighter as well as heavier tail behaviour.

We apply theorem 2 to show that finite location-scatter mixtures from the Kotz family with variable shape parameter $\theta = (r, s) \in (-p/2, \infty) \times (1/2, \infty)$ are identifiable. Let $\theta_j = (r_j, s_j)$ and $\gamma_j = (\theta_j, a_j, b_j)$ for $j = 1, 2$ and define a total ordering on the set $\mathcal{C}$, specified in the theorem, as follows: $\gamma_1 \prec \gamma_2$ if $s_1 < s_2$; or if $s_1 = s_2$ and $a_1 < a_2$; or if $s_1 = s_2$, $a_1 = a_2$ and $b_1 < b_2$; or if $s_1 = s_2$, $a_1 = a_2$, $b_1 = b_2$ and $r_1 > r_2$. With this total order, condition (5) is satisfied and theorem 2 applies.

For fixed parameter values condition (6) is satisfied, unless both r and s are integers. Hence, theorem 3 shows that the respective location-scatter families are identifiable. For $r = 0$ and $s = 1/2$ we obtain the identifiability of finite location-scatter mixtures from the multivariate Laplace distribution ($k_0 = 1$). Al-Hussaini & Ahmad (1981) considered mixtures of univariate Laplace distributions, but proved identifiability with fixed location or scale parameters only. A similar comment applies to many of their examples. The aforementioned results can be combined to prove the identifiability of location-scatter mixtures for the exponential power distribution with variable shape parameter $s \geq 1/2$. We omit the tedious proof.

Example 4 (Location-scatter mixtures). We give additional examples of elliptical distributions with identifiable finite location-scatter mixtures.

The multivariate generalized hyperbolic secant distribution (Harkness & Harkness, 1968), with shape parameter $\theta > 0$, has characteristic generator

$$\phi(u; \theta) = \operatorname{sech}^{2\theta}(\sqrt{u}/2) = \left[(e^{\sqrt{u}/2} + e^{-\sqrt{u}/2})/2 \right]^{-2\theta}, \quad u \geq 0.$$

For fixed $\theta > 0$ finite mixtures from the associated location-scatter family are identifiable. Indeed, $\phi(au; \theta) \sim e^{-\theta\sqrt{au}}$ as $u \to \infty$ for $a > 0$, and theorem 1 applies with the standard order on $(0, \infty)$. However, theorem 1 cannot be used to prove identifiability with variable shape parameter.

The symmetric Bessel distribution (Fang *et al.*, 1990, p. 92), with shape parameter $\theta > 0$, has a density generator $f_p(\cdot; \theta)$ that is proportional to the characteristic generator (7) of the multivariate t distribution. One can show that for fixed shape parameter θ, theorem 3 applies, yielding identifiability of the associated location-scatter mixtures. Indeed, f_p evidently has continuous derivatives of all orders on $(0, \infty)$ since this holds for the modified Bessel function K. Furthermore, (6) can be checked with $k_0 = 1$ using the series expansions in Abramowitz & Stegun (1965, p. 375).

Example 5 (Band-limited densities). Densities with compactly supported characteristic generators are called band-limited and play significant roles in signal processing (Kay, 1988). Let ϕ be a characteristic generator in $\mathbb{R}^p$ where $1 \leq p < q$. A slight variant of the arguments in the proof of theorem 1 shows that if there exists $d > 0$ such that $\phi(u) = 0$ for $u \geq d$ and $\phi(u) \neq 0$ for $u < d$, then finite mixtures from the respective location-scatter family $\{f_{\alpha, p} : \alpha = (\mu, \Sigma)\}$ are identifiable.

The classical example of a non-identifiable location-scatter family is based on the uniform distribution on $\mathbb{R}$ (Teicher, 1961). We give a similar counterexample.

Example 6 (Triangular distributions). Consider the univariate triangular location-scale family

$$f(x; \mu, a) = \frac{1}{a} \left(1 - \left| \frac{x - \mu}{a} \right| \right) 1_{|x-\mu| \leq a}, \quad x \in \mathbb{R},$$

where $\mu \in \mathbb{R}$ and $a > 0$. It is easily seen that

$$f(x; 0, 1) = \frac{4}{9} f(x; \tfrac{1}{3}, \tfrac{2}{3}) + \frac{4}{9} f(x; -\tfrac{1}{3}, \tfrac{2}{3}) + \frac{1}{9} f(x; 0, \tfrac{1}{3}), \quad x \in \mathbb{R}.$$

Hence, finite location-scale mixtures of the triangular density are not identifiable.

4. Normal scale mixtures

In this section, we study conditions on a particular class of characteristic generators ϕ, as to guarantee identifiability of the associated location-scatter families.

To characterize this class, we recall Schoenberg's theorem, which states that a function ϕ is a characteristic generator in any dimension if and only if it is of the form

$$\phi(u) =: \phi_G(u) = \int_0^\infty e^{-ur/2} \, dG(r), \quad u \geq 0, \tag{9}$$

where G is a probability measure on $(0, \infty)$. See, for example, Fang *et al.* (1990, p. 48). Equivalently, the associated elliptical distributions form scale mixtures of normal distributions. Many of the examples in section 3 are of this form (Gneiting, 1997).

We now fix the measure G and consider the respective location-scatter family $\{f_{\alpha,p} : \alpha = (\mu, \Sigma)\}$, defined in terms of the associated characteristic functions,

$$\phi_{\alpha,G}(t) = \exp(i\mu^{\mathrm{T}}t)\phi_G\left(t^{\mathrm{T}}\Sigma t\right), \quad t \in \mathbb{R}^p.$$

Note that $\phi_G(2u)$ equals the Laplace transform, $L_G(u)$, of the mixing measure, G, in the representation (9). Hence, if $L_G(a_2 u)/L_G(a_1 u) \to 0$ as $u \to \infty$ for $a_2 > a_1 > 0$ or, equivalently,

$$\lim_{u \to \infty} \frac{L_G(au)}{L_G(u)} = 0 \quad \text{for} \quad a > 1, \tag{10}$$

then theorem 1 implies that finite mixtures from the family $\{f_{\alpha,p}, \alpha = (\mu, \Sigma)\}$ generated by ϕ_G are identifiable. The following result provides a sufficient condition.

Theorem 4

Suppose that the density g of G exists and satisfies

$$\lim_{r \to 0} \frac{g(r)}{g(ar)} = 0 \quad for \quad a > 1. \tag{11}$$

Then the tail condition (10) holds and, consequently, finite mixtures from the family $\{f_\alpha : \alpha = (\mu, \Sigma)\}$ generated by ϕ_G are identifiable.

In particular, if the mixing density g satisfies

$$g(r) \propto r^\beta \exp(-c/r^\alpha) \quad \text{as} \quad r \to 0 \tag{12}$$

for some $\alpha, c > 0$ and $\beta \in \mathbb{R}$, or decays even faster as $r \to 0$, then (11) holds. On the other hand, if g satisfies

$$g(r) \propto r^\beta \quad \text{as} \quad r \to 0 \tag{13}$$

for some $\beta \geq 0$, or decays even slower as $r \to 0$, then L_G decays at most algebraically and f lacks smoothness at the origin; therefore, theorem 3 applies. Clearly, (12) and (13) span a comprehensive range of behaviours for the mixing distribution under which finite mixtures from the location-scatter family generated by ϕ_G are identifiable. It is tempting to conjecture that finite location-scatter mixtures based on a normal scale mixture are always identifiable, irrespectively of the mixing measure.

Note that theorem 4 links the tail behaviour of the Laplace transform L_G to the behaviour of the measure G at the origin. This result complements the Abelian theory in section XIII of Feller (1973) and might be of independent interest.

5. Discussion

There appears to be an understanding in the literature that finite location-scatter mixtures of continuous densities are generally identifiable, with few exceptions (McLachlan & Peel, 2000, p. 28). However, various questions remain open, particularly as to which densities are the exceptional ones, which parameters are allowed to vary in the mixture, and as to the effects of smoothness and dimensionality. We believe that our results are reassuring, in that identifiability for elliptical distributions generally persists in higher dimensions, and if in addition to the location vector and scatter matrix additional shape parameters are allowed to vary as well. Furthermore, our results apply to differentiable densities and also to densities that lack smoothness. Some of our findings, and in particular theorems 2 and 3, might well extend to

location-scatter families associated with less stringent or alternative notions of multivariate symmetry, such as those discussed by Fang *et al*. (1990).

Acknowledgements

References

Abramowitz, M. & Stegun, I. A. (1965). *Handbook of mathematical functions, with formulas, graphs, and mathematical tables*. 9th edn. Dover, New York.

Al-Hussaini, E. & Ahmad, K. E.-D. (1981). On the identifiability of finite mixtures of distributions. *IEEE Trans. Inform. Theory* **27**, 664–668.

Böhning, D. (2000). *Computer-assisted analysis of mixtures and applications: meta-analysis, disease mapping and others*. Chapman & Hall, London.

Box, G. E. P. & Tiao, G. C. (1992). *Bayesian inference in statistical analysis*. Wiley, New York.

Cambanis, S., Huang, S. & Simons, G. (1981). On the theory of elliptically contoured distributions. *J. Multivariate Anal.* **11**, 368–385.

Chandra, S. (1977). On the mixtures of probability distributions. *Scand. J. Statist.* **4**, 105–112.

Fang, K.-T., Kotz, S. & Ng, K. (1990). *Symmetric multivariate and related distributions*. Chapman & Hall, London.

Feller, W. (1971). *An introduction to probability theory and its applications*, Vol. 2, 2nd edn. Wiley, New York.

Fraley, C. & Raftery, A. E. (2002). Model-based clustering, discriminant analysis and density estimation. *J. Amer. Statist. Assoc.* **97**, 611–631.

Gneiting, T. (1997). Normal scale mixtures and dual probability densities. *J. Statist. Comput. Simul.* **59**, 375–384.

Harkness, W. L. & Harkness, M. L. (1968). Generalized hyperbolic secant distributions. *J. Amer. Statist. Assoc.* **63**, 329–337.

Holzmann, H., Munk, A. & Stratmann, B. (2004). Identifiability of finite mixtures – with applications to circular distributions. *Sankhya* **66**, 440–449.

Holzmann, H., Munk, A. & Zucchini, W. (2006). On identifiability in capture–recapture models. *Biometrics* (in press).

Kay, S. (1988). *Modern spectral estimation: theory and applications*. Prentice-Hall, Englewood Cliffs, NJ.

Kelker, D. (1970). Distribution theory of spherical distributions and a location-scale parameter generalization. *Sankhya Ser. A* **32**, 419–430.

Kent, J. T. (1983). Identifiability of finite mixtures for directional data. *Ann. Statist.* **11**, 984–988.

Kotz, S. & Nadarajah, S. (2004). *Multivariate t distributions and their applications*. Cambridge University Press, Cambridge.

Leroux, B. G. (1992). Consistent estimation of the mixing distribution. *Ann. Statist.* **20**, 1350–1360.

Lindsay, B. G. (1995). *Mixture models: theory, geometry, and applications*. Institute for Mathematical Statistics, Hayward, CA.

Lindsay, B. G., Kettenring, J. & Siegmund, D. O. (2004). A report on the future of statistics (with discussion). *Statist. Sci.* **19**, 387–413.

Link, W. A. (2003). Nonidentifiability of population size from capture-recapture data with heterogeneous detection probabilities. *Biometrics* **59**, 1123–1130.

McLachlan, G. J. & Peel, D. (2000). *Finite mixture models*. Wiley, New York.

Peel, D. & McLachlan, G. J. (2000). Robust mixture modeling using the *t*-distribution. *Statist. Comput.* **10**, 339–348.

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

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

Titterington, D. M., Smith, A. F. M. & Makov, U. E. (1985). *Statistical analysis of finite mixture models*. Wiley, Chichester.

Watson, G. N. (1944). *A treatise on the theory of bessel functions*, 2nd edn. Cambridge University Press, Cambridge.

Yakowitz, S. J. & Spragins, J. D. (1968). On the identifiability of finite mixtures. *Ann. Math. Statist.* **39**, 209–214.

Hajo Holzmann, Institute for Mathematical Stochastics, Georg-August-Universität Göttingen, Maschmühlenweg 8–10, 37073 Göttingen, Germany.
E-mail: holzmann@math.uni-goettingen.de

Appendix: Proofs

Proof of lemma 1

Let the dimension, p, be fixed and suppose that

$$\sum_{j=1}^{n} \lambda_j f_{\alpha_j, p}(x) = 0, \quad x \in \mathbb{R}^p, \tag{14}$$

where $\lambda_j \in \mathbb{R}$ and where the tuples $\alpha_j = (\theta_j, \mu_j, \Sigma_j) \in \mathcal{A}^p$ are pairwise distinct. Taking the p-dimensional Fourier transform in (14) gives

$$\sum_{j=1}^{n} \lambda_j \exp(i\mu_j^{\mathrm{T}} t) \phi(t^{\mathrm{T}} \Sigma_j t; \theta_j) = 0, \quad t \in \mathbb{R}^p.$$

Yakowitz & Spragins (1968, p. 211) show that there exists $z \in \mathbb{R}^p$ such that the tupels $\alpha_j^z = (\theta_j, \mu_j^{\mathrm{T}} z, z^{\mathrm{T}} \Sigma_j z)$ are pairwise distinct. If we write $t = uz$ for $u \in \mathbb{R}$ then

$$\sum_{j=1}^{n} \lambda_j \exp(iu\mu_j^{\mathrm{T}} z) \phi(u^2 z^{\mathrm{T}} \Sigma_j z; \theta_j) = 0, \quad u \in \mathbb{R}.$$

Taking the one-dimensional inverse Fourier transform gives

$$\sum_{j=1}^{n} \lambda_j f_{\alpha_j^z, 1}(t) = 0, \quad t \in \mathbb{R},$$

which implies $\lambda_1 = \cdots = \lambda_n = 0$ by our identifiability assumption on the univariate family.

Proof of theorem 1

By lemma 1, it suffices to consider the case $p = 1$. Starting with the relation (14), we take the Fourier transform and obtain

$$\sum_{j=1}^{n} \lambda_j \exp(i\mu_j u) \phi_{\beta_j}(u^2) = 0, \quad u \in \mathbb{R},$$

where $\beta_j = (\sigma_j^2, \theta_j)$ and σ^2 stands for the univariate variance parameter. Without loss of generality we may assume that $\beta_1 \preceq \cdots \preceq \beta_n$. Let $m \geq 1$ be such that $\beta_1 = \cdots = \beta_m \prec \beta_{m+1} \preceq \cdots \preceq \beta_n$; the case $\beta_1 = \cdots = \beta_n$ is treated similarly. Dividing by $e^{i\mu_1 u} \phi_{\beta_1}(u^2)$ and rearranging terms, we get

$$\lambda_1 + \sum_{j=2}^{m} \lambda_j\, e^{iu(\mu_j - \mu_1)} + \sum_{j=m+1}^{n} \lambda_j\, e^{iu(\mu_j - \mu_1)} \frac{\phi_{\beta_j}(u^2)}{\phi_{\beta_1}(u^2)} = 0, \quad u \in \mathbb{R}. \tag{15}$$

From (4) and (15) we find that $\sum_{j=2}^{m} \lambda_j \mathrm{e}^{iu(\mu_j - \mu_1)} \to -\lambda_1$ as $u \to \infty$. Since $\mu_j - \mu_1 \neq 0$ for $j = 2, \ldots, m$, there exists $u_0 \in \mathbb{R}$ such that $u_0(\mu_j - \mu_1) \in (0, 2\pi) \bmod 2\pi$ for $j = 2, \ldots, m$. Now $\sum_{j=2}^{m} \lambda_j \mathrm{e}^{ilu_0(\mu_j - \mu_1)} \to -\lambda_1$ as $l \to \infty$; hence, the arithmetic mean of these terms converges to $-\lambda_1$, too. However, lemma 2.1 in Holzmann *et al.* (2004) implies that

$$\frac{1}{N} \sum_{l=0}^{N-1} \sum_{j=2}^{m} \lambda_j \mathrm{e}^{ilu_0(\mu_j - \mu_1)} \to 0 \quad \text{as} \quad N \to \infty,$$

thereby showing that $\lambda_1 = 0$. An inductive argument completes the proof.

Proof of theorem 2

Suppose that

$$\sum_{j=1}^{m} \lambda_j f_p \left[(x - \mu_j)^{\mathrm{T}} \Sigma_j^{-1} (x - \mu_j); \theta_j \right] = 0, \quad x \in \mathbb{R}^p,$$

where the tuples $\alpha_j = (\mu_j, \Sigma_j, \theta_j)$ are pairwise distinct, and where the $|\Sigma_j|^{-1/2}$ are subsumed into the λ_j. If we put $x = uz$, where u varies over $\mathbb{R}$ and $z \in \mathbb{R}^p$ is considered fixed, then

$$(x - \mu_j)^{\mathrm{T}} \Sigma_j^{-1} (x - \mu_j) = a_j u^2 + b_j u + c_j,$$

where $a_j = z^{\mathrm{T}} \Sigma_j^{-1} z$ and $b_j = -2 \mu_j^{\mathrm{T}} \Sigma_j^{-1} z$, and we conclude that

$$\lambda_1 = -\sum_{j=2}^{m} \lambda_j \frac{f_p \left(a_j u^2 + b_j u + c_j; \theta_j \right)}{f_p \left(a_1 u^2 + b_1 u + c_1; \theta_1 \right)}, \quad u \in \mathbb{R}. \tag{16}$$

The argument in Yakowitz & Spragins (1968, p. 211) shows that there exists $z \in \mathbb{R}^p$ such that the tuples (θ_j, a_j, b_j) are pairwise distinct. Hence, we can assume without loss of generality that $(\theta_1, a_1, b_1) \prec \cdots \prec (\theta_m, a_m, b_m)$. In view of (5), letting $u \to \infty$ in (16) implies $\lambda_1 = 0$, and the proof is completed by an inductive argument.

Proof of theorem 3

Suppose that

$$\sum_{j=1}^{m} \lambda_j f_p \left[(x - \mu_j)^{\mathrm{T}} \Sigma_j^{-1} (x - \mu_j) \right] = 0, \quad x \in \mathbb{R}^p,$$

where the tuples (μ_j, Σ_j) are pairwise distinct. Put $x = uz + \mu_1$, where u varies over $\mathbb{R}$ and $z \in \mathbb{R}^p$ is chosen such that the following hold: $uz + \mu_1 - \mu_j \neq 0 \in \mathbb{R}^p$ for $u \in \mathbb{R}$ unless $\mu_j = \mu_1$ and, furthermore, the $\sigma_j^2 := z^{\mathrm{T}} \Sigma_j^{-1} z$ are pairwise distinct for the indices j with $\mu_j = \mu_1$. Then it follows that

$$\sum_{j|\mu_j = \mu_1} \lambda_j \sigma_j^{2k-1} h^{(2k-1)} (u\sigma_j) = -\sum_{j|\mu_j \neq \mu_1} \lambda_j h_j^{(2k-1)}(u), \quad u \in \mathbb{R} \setminus \{0\}, \tag{17}$$

where $h_j(u) = f_p \left(\| \Sigma_j^{-1/2} (uz + \mu_1 - \mu_j) \|^2 \right)$. The right-hand side of (17) is continuous at $u = 0$, which requires that

$$\sum_{j|\mu_j = \mu_1} \sigma_j^{2k-1} \lambda_j = 0 \quad \text{for} \quad k = k_0, \ldots, k_0 + m - 1,$$

where m denotes the number of terms in the sum. Since the Vandermonde matrix has full rank, this linear system implies that $\lambda_j = 0$ for all indices j such that $\mu_j = \mu_1$. Proceeding inductively, we see that $\lambda_j = 0$ for all j.

Proof of the claim in example 5

By lemma 1, we may suppose that $p = 1$. Starting with (14), we take the Fourier transform and obtain

$$\sum_{j=1}^{n} \lambda_j \, e^{i\mu_j u} \phi(\sigma_j^2 u^2) = 0, \quad u \in \mathbb{R},$$

where we may assume that $\sigma_1 \leq \cdots \leq \sigma_n$ and $\sigma_1 = \cdots = \sigma_m < \sigma_{m+1}$. Division by $\phi(\sigma_1^2 u^2)$ yields $\sum_{j=1}^{m} \lambda_j \, e^{i\mu_j u} = 0$ for $d/\sigma_{m+1}^2 < u^2 < d/\sigma_1^2$. The functions $e^{i\mu_j u}$ are analytic; hence, $\sum_{j=1}^{m} \lambda_j \, e^{i\mu_j u} = 0$ for $u \in \mathbb{R}$, and the proof is completed in analogy to the proof of theorem 1.

Proof of theorem 4

Suppose that $a > 1$ and consider the decomposition

$$L_G(au) = \int_0^{u^{-1/2}} e^{-aur} g(r) \, dr + \int_{u^{-1/2}}^{\infty} e^{-aur} g(r) \, dr = I_1 + I_2.$$

In order to estimate $I_2/L_G(u)$, note that $e^{-aur} g(r) \leq e^{-ur} g(r) \, \exp(-(a-1)u^{1/2})$ for $r \geq u^{-1/2}$, which implies

$$\frac{I_2}{L_G(u)} \leq \frac{\int_{u^{-1/2}}^{\infty} e^{-aur} g(r) \, dr}{\int_{u^{-1/2}}^{\infty} e^{-ur} g(r) \, dr} \leq \exp(-(a-1)u^{1/2}) \to 0 \quad \text{as} \quad u \to \infty.$$

Concerning I_1, we find that

$$\frac{I_1}{L_G(u)} \leq \frac{\int_0^{u^{-1/2}} e^{-aur} g(r) \, dr}{\int_0^{au^{-1/2}} e^{-ur} g(r) \, dr} = \frac{\int_0^{au^{1/2}} e^{-s} g[s/(au)] \, ds}{a \int_0^{au^{1/2}} e^{-s} g(s/u) \, ds}. \tag{18}$$

From (11),

$$\sup_{s \in [0, \, au^{1/2}]} \frac{g[s/(au)]}{g(s/u)} = \sup_{r \in [0, \, u^{-1/2}]} \frac{g(r)}{g(ar)} \to 0 \quad as \quad u \to \infty,$$

and this implies that the quotient in (18) tends to zero, thereby proving (10).