Title
Maximum likelihood estimation of the mixture of log-concave densities

Authors
Hu, Hao
Wu, Yichao
Yao, Weixin

Publication Date
2016-09-01

DOI
10.1016/j.csda.2016.03.002

Peer reviewed

Maximum likelihood estimation of the mixture of log-concave densities

Hao Hu[a,*], Yichao Wu[a], Weixin Yao[b]

[a]*Department of Statistics, North Carolina State University, Raleigh, North Carolina 27695, U.S.A.*
[b]*Department of Statistics, University of California, Riverside, California 92521, U.S.A.*

Abstract

Finite mixture models are useful tools and can be estimated via the EM algorithm. A main drawback is the strong parametric assumption about the component densities. In this paper, a much more flexible mixture model is considered, which assumes each component density to be log-concave. Under fairly general conditions, the log-concave maximum likelihood estimator (LCMLE) exists and is consistent. Numeric comparisons are also made to demonstrate that the LCMLE improves the clustering results while comparing with the traditional MLE for parametric mixture models.

Keywords: Consistency, Log-concave maximum likelihood estimator (LCMLE), Mixture model.

1. Introduction

The finite mixture model (McLachlan & Peel, 2000; Mcnicholas & Murphy, 2008) provides a flexible methodology for both theoretical and practical analysis. It has the density of the form

$$f(x) = \sum_{j=1}^{K} \lambda_j g_j(x; \theta_j) \quad x \in \mathbb{R}^p, \tag{1.1}$$

where $\lambda_1, \ldots, \lambda_K$ are the mixing proportions and $g_j(x; \theta_j)$'s are component densities. The unknown parameters in the mixture model (1.1) can be estimated by the EM algorithm, see e.g. Dempster *et al.* (1977) and McLachlan & Krishnan (2007). One major

*Correspondence to: North Carolina State University, 5109 SAS Hall, 2311 Stinson Dr, Raleigh, NC, 27695, United States

Email addresses: `hhu5@ncsu.edu` (Hao Hu), `wu@stat.ncsu.edu` (Yichao Wu), `weixin.yao@ucr.edu` (Weixin Yao)

drawback of the traditional mixture model (1.1) is the strong parametric assumption about the component density g_j. It is often too restrictive and the density estimation may be inaccurate due to the model misspecification. Another drawback is that each model requires a specific EM algorithm based on the parametric assumption.

To relax the parametric assumption, nonparametric shape constraints are becoming increasingly popular. In this paper, we make one fairly general shape constraint for our mixture model. We assume that each component density is log-concave. A density g is log-concave if $\log g$ is concave. Examples of log-concave densities include normal, Laplace, logistic, as well as gamma and beta with certain parameter constraints. Log-concave densities have lots of nice properties as described by Balabdaoui *et al.* (2009). Their nonparametric maximum likelihood estimators were studied by Dümbgen & Rufibach (2009), Cule *et al.* (2010), Cule & Samworth (2010), Chen & Samworth (2013), Pal *et al.* (2007) and Dümbgen *et al.* (2011) (referred as [DSS 2011] thereafter). The convergence rates of these estimators for log-concave densities were studied by Doss & Wellner (2013) and Kim & Samworth (2014). Such estimators provide more generality and flexibility without any tuning parameters.

In our model, we assume that $X_1, \dots, X_n$ are independent d-dimensional random variables with distribution Q_0 and the mixture density f_0. The mixture density f_0 belongs to a given class

$$\mathcal{F} = \{f : f(x) = \sum_{j=1}^{K} f_j(x) = \sum_{j=1}^{K} \lambda_j \exp\{\phi_j(x)\}, \boldsymbol{\lambda} \in \Lambda, \boldsymbol{\phi} \in \Phi\}, \tag{1.2}$$

where $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_K)$, $\Lambda = \{(\lambda_1, \dots, \lambda_K) : 0 < \lambda_j < 1, \sum_{j=1}^{K} \lambda_j = 1\}$, $\boldsymbol{\phi} = (\phi_1, \dots, \phi_K)$, and $\Phi = \{(\phi_1, \dots, \phi_K) : \phi_j \text{ is concave}\}$. We assume that each ϕ_j is continuous and is coercive in the sense that $\phi_j(x) \to -\infty$ as $||x|| \to \infty$ $(j = 1, \dots, K)$.

Note that, similar to traditional normal mixture models with unequal variance, the likelihood functions for mixture of log-concave densities are unbounded as well (e.g. a normal mixture with $x = \mu_1$ and $\sigma_1^2 \to 0$, see Section 3.10 of McLachlan & Peel (2000) for detail discussions). Many methods have been proposed to solve the unboundedness issue of mixture likelihood, see for example, Hathaway (1985), Chen *et al.* (2008), and Yao (2010). Similar to Hathaway (1985), we will define LCMLE on a constrained parameter space. Let $M_j(\boldsymbol{\phi}) = \max_{x\in\mathbb{R}^d}\{\phi_j(x)\}$, $M_{(1)}(\boldsymbol{\phi}) = \min_j\{M_j(\boldsymbol{\phi})\}$, and $M_{(K)}(\boldsymbol{\phi}) = \max_j\{M_j(\boldsymbol{\phi})\}$. We further define the ratio $\mathcal{S}(\boldsymbol{\phi}) = M_{(1)}(\boldsymbol{\phi})/M_{(K)}(\boldsymbol{\phi})$. Here, we borrow the idea of Hathaway (1985) by restricting our interest to a constrained subspace $\mathbf{\Phi}_\eta$ such that $\mathbf{\Phi}_\eta = \{\boldsymbol{\phi}\in\mathbf{\Phi} : |\mathcal{S}(\boldsymbol{\phi})| \geq \eta > 0\}$ for some $\eta\in(0,1]$. This restriction avoids estimating the case that the modes of different components differ a lot. By restricting on $\mathbf{\Phi}_\eta$, we focus our interest on $f\in\mathcal{F}_\eta$, where

$$\mathcal{F}_\eta = \{f : f(x) = \sum_{j=1}^{K} f_j(x) = \sum_{j=1}^{K} \lambda_j \exp\{\phi_j(x)\}, \boldsymbol{\lambda}\in\Lambda, \boldsymbol{\phi}\in\Phi_\eta\}. \tag{1.3}$$

Let Q_n be the empirical distribution of $X_1,\ldots,X_n$. The (restricted) log-concave maximum likelihood estimator (LCMLE) is

$$f_n = f(\cdot|Q_n) = \underset{f\in\mathcal{F}_\eta}{\operatorname{argmax}} \int \log(f)dQ_n. \tag{1.4}$$

In practice, similar to Hathaway (1985), picking η can be tricky for some extreme case. If η is too small, there might be a chance that some boundary point $|\mathcal{S}(\boldsymbol{\phi})| = \eta$ maximizes the log-likelihood and the solution depends on the choice of η. In this paper, we do not focus on the issue of choosing η. The constrained subspace $\mathbf{\Phi}_\eta$ is mainly used for theoretical development. Based on our empirical experience, if we start the algorithm from a reasonable initial value, such as the maximum likelihood estimate assuming all components are normal with equal variance, the unboundedness issue is very rare.

Many methods have been proposed to rexlax the parametric assumption of (1.1). Hunter *et al.* (2007), Bordes *et al.* (2006a), Butucea & Vandekerkhove (2014), and Chee & Wang (2013) considered the extension of (1.1) by assuming all component densities are symmetric but unknown. Bordes *et al.* (2006b), Bordes & Vandekerkhove (2010), Hohmann & Holzmann (2013), Xiang *et al.* (2014), and Ma & Yao (2015) considered the extension of (1.1) when $K = 2$ and one of the component densities is symmetric but unknown. Mixtures of log-concave densities have been studied by Chang & Walther (2007), Cule *et al.* (2010) and Balabdaoui & Doss (2014). Chang & Walther (2007) provided an EM-type algorithm and demonstrated sound numerical results in the simulation study. Cule *et al.* (2010) applied the mixture of log-concave model to Wisconsin breast cancer data set. Balabdaoui & Doss (2014) considered a special case when all components have the same symmetric log-concave densities but with different location parameters, and proved the $\sqrt{n}$-consistency of their proposed M-estimators for mixing proportion as well as location parameters. Note that these models are special cases from the family of $\mathcal{F}$. Therefore, their estimators and asymptotic results cannot be applied here. For example, the mixtures of normal distributions with different component means and variances belongs to $\mathcal{F}$ but do not belong to the model family considered by Balabdaoui & Doss (2014).

To the best of our knowledge, none of existing works has studied the theoretical properties of the estimator for the log-concave mixture model (1.2) under such general conditions. This paper aims to fill in this gap. We show that theoretically, the LCMLE (in the restricted subset $\mathcal{F}_\eta$) exists, and is consistent under fairly general conditions. However, we want to point out that the extension of the properties of the log-concave density to mixtures of log-concave densities is not trivial. The log-density $l_n = l(\cdot|Q_n) = \log f_n$ is no longer guaranteed to be a concave function. Consequently, many nice theoretical properties that stated in DSS 2011 no longer hold for our mixture model.

The rest of the paper is organized as follows. Section 2 introduces the basic setup, model details, and notations. Section 3 states the theoretical properties. We review the EM-type algorithm for log-concave mixture models in Section 4. Simulation studies are conducted in Section 5. We end the article with a short conclusion in Section 6. The proofs and lemmas are presented in the appendix.

2. Log-concave maximum likelihood estimator

Let Q be a distribution on $\mathbb{R}^d$. Our goal is to maximize a log-likelihood-type functional:

$$L(\boldsymbol{\phi}, \boldsymbol{\lambda}, \boldsymbol{\pi}, Q) = \int \log[\sum_{j=1}^{K} \lambda_j \exp\{\phi_j(x)\}] dQ(x) - \sum_{j=1}^{K} \pi_j (\int \exp\{\phi_j(x)\} dx - 1), \tag{2.1}$$

where π_j's are Lagrange multipliers to incorporate the constraint $\int \exp\{\phi_j(x)\}dx = 1$ $(j = 1, \ldots, K)$. We define a profile log-likelihood:

$$L(Q) = \sup_{\boldsymbol{\phi}\in\Phi_\eta,\boldsymbol{\lambda}\in\Lambda,\boldsymbol{\pi}} L(\boldsymbol{\phi}, \boldsymbol{\lambda}, \boldsymbol{\pi}, Q). \tag{2.2}$$

If, for fixed Q, $(\boldsymbol{\psi}, \boldsymbol{\lambda}^*, \boldsymbol{\pi}^*)$ maximizes $L(\boldsymbol{\phi}, \boldsymbol{\lambda}, \boldsymbol{\pi}, Q)$, it will automatically satisfy that:

$$\pi_j^* = E(\pi(j|x)) = \int \frac{\lambda_j^* \exp\{\psi_j(x)\}}{(\sum_{h=1}^K \lambda_h^* \exp\{\psi_h(x)\})} dQ(x); \tag{2.3}$$

$$\int \exp\{\psi_j(x)\}dx = 1 \quad (j = 1, 2, \ldots, K). \tag{2.4}$$

Note that differing from the non-mixture setting in DSS 2011, π_j^* is not equal to 1.

To verify this, note that $\boldsymbol{\phi} + \boldsymbol{c} \in \Phi$ for any fixed vector of functions $\boldsymbol{\phi} \in \Phi$ and arbitrary $\boldsymbol{c} = (c_1, \ldots, c_K) \in \mathbb{R}^K$, and

$$\frac{\partial L(\boldsymbol{\psi} + \boldsymbol{c}, \boldsymbol{\lambda}, \boldsymbol{\pi}, Q)}{\partial c_h}|_{\boldsymbol{c}=0} = (\int \frac{\lambda_h \exp\{\psi_h(x)\}}{\sum_{j=1}^K \lambda_j \exp\{\psi_j(x)\}} dQ(x) - \pi_h \int e^{\psi_h(x)} dx) = 0,$$

$$\frac{\partial L(\boldsymbol{\psi}, \boldsymbol{\lambda}, \boldsymbol{\pi}, Q)}{\partial \pi_h} = 1 - \int \exp\{\psi_h(x)\}dx = 0.$$

The maximizer $(\boldsymbol{\psi}, \boldsymbol{\lambda}^*)$ forms the log-likelihood maximizer $l^*(x) = \log \sum_{j=1}^K \lambda_j^* e^{\psi_j(x)}$.

3. Theoretical Properties

Before we state the main theories, we first define the convex support of a distribution.

Definition For any distribution Q, let $Q(C)$ be the probability measure of the set C. The convex support of Q is the set such that:

$$csupp(Q) = \bigcap \{C : C \subseteq \mathbb{R}^d \text{ closed and convex}, Q(C) = 1\}.$$

The convex support is itself closed and convex with $Q(csupp(Q)) = 1$.

We first show the existence of the maximizer of (2.1) based on the following general assumptions:

(A1) $\int ||x||dQ < \infty$ (We define $||x||$ as Euclidean norm in our paper).

(A2) $interior(csupp(Q)) \neq \emptyset$.

Theorem 1. *For any Q that satisfies (A1) and (A2), the value of $L(Q)$ is real and, with probability 1, there exists a maximizer:*

$$(\boldsymbol{\psi}, \boldsymbol{\lambda}^*, \boldsymbol{\pi}^*) = \underset{\boldsymbol{\phi} \in \Phi_\eta, \boldsymbol{\lambda} \in \Lambda, \boldsymbol{\pi}}{\operatorname{argmax}} L(\boldsymbol{\phi}, \boldsymbol{\lambda}, \boldsymbol{\pi}, Q) \textit{ such that } \int e^{\psi_j(x)} dx = 1 \quad \textit{for} \quad j = 1, \ldots, K.$$

Next, we establish the consistency of the estimated mixture density. In the following text, we refer the concept of convergence of distribution as converging with respect to Mallows distance D_1: $D_1(Q, Q^{'}) = \inf_{(X,X')} E||X - X^{'}||$, where Q and $Q^{'}$ are two distributions and the infimum is taken over all pairs of $(X, X^{'})$ such that $X \sim Q$ and $X^{'} \sim Q^{'}$. The convergence with respect to Mallows distance, i.e. $\lim_{n\to\infty} D_1(Q_n, Q) = 0$, is equivalent with $Q_n \to^w Q$ and $\int \|x\| dQ_n(x) \to \int \|x\| dQ(x)$ as $n \to \infty$.

Theorem 2. *Let a sequence Q_n and the true distribution Q_0 satisfy (A1) and (A2). Moreover, if the following condition holds:*

$$(A3) \lim_{n\to\infty} D_1(Q_n, Q_0) = 0.$$

Then, with probability 1,

$$\lim_{n\to\infty} L(Q_n) = L(Q_0).$$

Let ϕ_{nj}'s and λ_{nj}'s be the maximizer corresponding to profile log-likelihood $L(Q_n)$, i.e, $f_n(x) = \sum \lambda_{nj} \exp\{\phi_{nj}(x)\} = f(\cdot|Q_n) \in \mathcal{F}_\eta$. For $f_0(x) = f(\cdot|Q_0) \in \mathcal{F}_\eta$, we have:

$$\lim_{n\to\infty, x\to y} f_n(x) = f_0(y) \quad \textit{for all} \quad y \notin \partial\{f_0 \geq 0\}, \tag{3.1}$$

$$\lim_{n\to\infty, x\to y} f_n(x) \leq f_0(y) \quad \textit{for all} \quad y \in \mathbb{R}^d, \tag{3.2}$$

$$\lim_{n\to\infty} \int |f_n(x) - f_0(x)| dx = 0. \tag{3.3}$$

The above theorem showed the consistency of the estimated mixture density. If we further assume that the true mixture density $f_0(x)$ is identifiable, then each estimated component densities and mixing proportions are also consistent. We will discuss more about the identifiability issue in Section 6.

4. EM-type algorithm

The EM algorithm of estimating log-concave mixture densities has already been developed by Chang & Walther (2007). Here we just briefly summarize it. First we run the EM algorithm for Gaussian mixture until convergence, which will provide a good initial value. Then we use the outcome as the starting values for our EM-type algorithm. We assume the observed data $\mathbf{x} = (x_1, \ldots, x_n)$ to be incomplete and define the missing value $\mathbf{z} = (z_1^T, ..., z_n^T)$, where z_i is a K-dimension vector:

$$z_{ij} = \begin{cases} 1 & \text{if } x_i \text{ belongs to } j\text{th group} \\ 0 & \text{otherwise} \end{cases}.$$

So the complete log-likelihood is:

$$\log f(\boldsymbol{\phi}, \boldsymbol{\lambda}; \mathbf{x}, \mathbf{z},) = \log \prod_{i=1}^{n} \prod_{j=1}^{K} [\lambda_j e^{\phi_j(x_i)}]^{z_{ij}} = \sum_{i=1}^{n} \sum_{j=1}^{K} z_{ij} [\log \lambda_j + \phi_j(x_i)],$$

where $\mathbf{x} = (x_1, \ldots, x_n)$. In E-step, we replace z_{ij} by

$$z_{ij}^{(t+1)} = \frac{\lambda_j^{(t)} e^{\widehat{\phi}_j^{(t)}(x_i)}}{\sum_{h=1}^{K} \lambda_h^{(t)} e^{\widehat{\phi}_h^{(t)}(x_i)}}.$$

In M-step, first we update λ by $\lambda_j^{(t+1)} = \frac{1}{n} \sum_{i=1}^{n} z_{ij}^{(t+1)}, j = 1, \ldots, K$. Then we update ϕ_j by maximizing $\sum_{i=1}^{n} z_{ij}^{(t+1)} \phi_j(x_i)$ with respect to ϕ_j through the function called `mlelcd` in the R package `LogConcDEAD` (Cule *et al.* (2009)) and get estimator $\widehat{\phi}_j^{(t+1)}$ for $j = 1, \ldots, K$. The estimation of $\widehat{\phi}_j$ has been studied by Walther (2002) and Rufibach (2007). Given i.i.d. data $X_1, \ldots, X_n$ which follow distribution f, the Log-concave Maximum Likelihood Estimator (LCMLE) $\widehat{f}_n$ exists uniquely and has support on the convex hull of the data (by Theorem 2 of Cule *et al.* (2010)). The log-likelihood estimator $\log \widehat{f}_n$ is a piecewise linear function with knots which are a subset of $\{X_1, \ldots, X_n\}$. Walther (2002) and Rufibach (2007) provided algorithms for computing $\widehat{f}_n(X_i), i = 1, \ldots, n$. The entire log-density $\log \widehat{f}_n$ can be computed by linear interpolating between between $\log \widehat{f}_n(X_{(i)})$ and $\log \widehat{f}_n(X_{(i+1)})$. Walther (2002) and Rufibach (2007) also pointed out that it is natural to apply weights for EM-type algorithm. The $z_{1j}^{(t+1)}, \ldots, z_{nj}^{(t+1)}$ can be viewed as weights for $x_1, \ldots, x_n$ when estimating the log-concave density ϕ_j in our algorithm for $j = 1, \ldots, K$.

To avoid the local maximum, we restart the algorithm 20 times and choose the result with the highest log-likelihood. The algorithm stops once the increasing increment is below 10^{-7}.

5. Numeric Results

We first show an example of density estimation for univariate case. 200 observations are generated from a mixture model of $0.3Logistic(0,1) + 0.7Laplace(5,1)$. This setup is at a more general form of Chang & Walther (2007), as Chang & Walther (2007) only considered the case that one component is a location shift of the other. The theoretical values of the component densities and the estimated values of the component densities are shown in Figure 1.

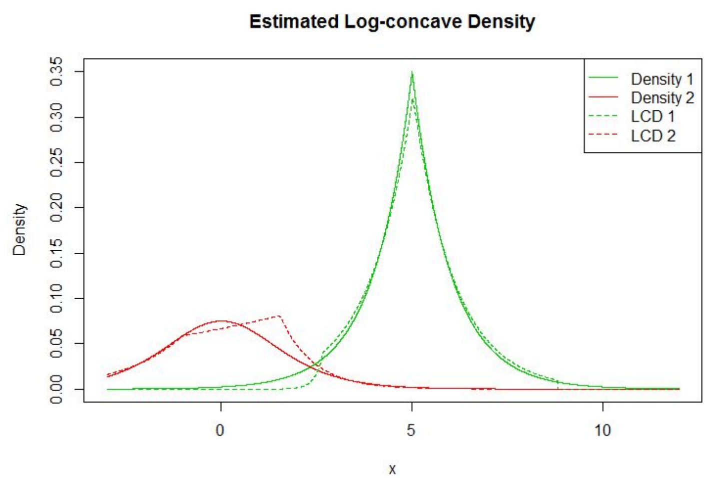


Figure 1: EM-type algorithm estimation for log-concave mixtures. Solid line represents the truth and dashed line represents the estimation results (LCD).

As we don't have tuning issue for LCMLE, the most attractive application of LCMLE is density estimation with dimensionality higher than 1. For a d-dimensional log-concave mixture density, we observe $n = 200$ observations $\mathbf{X}_1, \ldots, \mathbf{X}_n$, where $\mathbf{X}_i = (X_{i1}, \ldots, X_{id}) \in \mathbb{R}^d$. To simplify our simulation, we focus on the model whose univariate marginal distributions are log-concave. We model the dependence structure with a normal copula. Suppose $(N_1, \ldots, N_d)$ be multivariate normal with mean $\mathbf{0}$ and covariance matrix Σ. Let $F_1, \ldots, F_d$ be the CDFs of desired univariate log-concave distributions. Then,

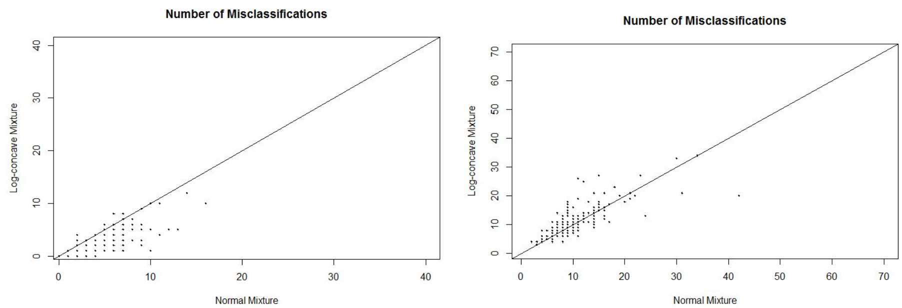
$$(X_{i1}, \ldots, X_{id}) = (F_1^{-1}(\Phi(N_1)), \ldots, F_d^{-1}(\Phi(N_d))).$$

Here, we generate 200 observations for the case $d = 3$, which is a higher dimension case compared with Chang & Walther (2007). The first component (with probability 0.4) is a 3-dimensional normal with mean $\mathbf{0}$ and $\Sigma = [1, 0.5, 0.5; 0.5, 1, 0.5; 0.5, 0.5, 1]$. The second component (with probability 0.6): x-y-z coordinates are independent. The x-coordinate is N(0,1), the y-coordinate is Gamma(2, 1) shifted by 1, and the z-coordinate is Beta(1,4) shifted by 1. The results are replicated 100 times. In Figure (2a), each point represents a single replicate. The x-axis represents the number of misclassification by Normal mixture EM-algorithm. The y-axis represents the number of misclassification by our log-concave mixture EM-algorithm. We observe significant improvement in the sense of misclassification rates.

We are also interested in the price which we have to pay for the flexibility while the data actually are from normal mixtures. We generate 200 observations from a Gaussian mixture, in which the first component (with probability 0.4) is a 3-dimensional normal with mean $\mathbf{0}$ and covariance matrix $[1, 0.5, 0.5; 0.5, 1, 0.5; 0.5, 0.5, 1]$, and the second component (with probability 0.6) is shifted by $(1, 1, 2)$ with same covariance matrix. We also replicate the results 100 times. From Figure (2b), we observe no significant penalty in this case.

6. Conclusion

The log-concave maximum likelihood estimator (LCMLE) provides more flexibility to estimate mixture densities, when compared to the traditional parametric mixture models. The estimation of LCMLE for log-concave mixtures can be achieved by an EM-type algorithm. The LCMLE is not sensitive to the model mis-specification and consequently, only one implementation of EM algorithm is necessary. Through simulation studies, we observed significant improvements in the sense of classification and no significant penalties when the parametric assumption is indeed correct.



(a) Log-concave mixtures

(b) Gaussian Mixtures

Figure 2: Three-dimensional clustering result: normal mixture EM-algorithm vs log-concave mixture EM-algorithm in the sense of number of misclassification. The solid lines represents the identity.

In this paper, we proved the existence of the LCMLE for log-concave mixture models. The consistency is also proved for the estimated mixture density. If the true mixture density is identifiable, then the estimated component densities are also identifiable. However, it is not an easy task to prove the overall identifiability for the most general family of mixtures of log-concave distributions in (1.2) from a nonparametric point of view. Some restrictive conditions, such as symmetry, are needed to ensure identifiability. Hunter *et al.* (2007) and Bordes *et al.* (2006a) proved the identifiability of (1.1) if $K = 2$ and both component densities are symmetric but with different location parameters. Balabdaoui & Doss (2014) has considered a special case of (1.2), when $\phi_j(x; \theta_j) = \phi(x - \theta_j)$ and ϕ is a symmetric concave function about 0, and the identifiability of (1.2) follows from Hunter *et al.* (2007) and Bordes *et al.* (2006a) when $K = 2$.

Acknowledgment

We wish to thank the Associate Editor and two reviewers for their helpful comments and suggestions that led to improvements in this paper. Hu's research is partially supported by National Institutes of Health grant R01-CA149569. Wu's research is partially supported by National Institutes of Health grant R01-CA149569 and National

Science Foundation grant DMS-1055210. Yao's research is supported by NSF grant DMS-1461677.

Appendix A: Lemmas

Lemma 1 is taken from Cule & Samworth (2010). Lemma 2 to Lemma 5 are taken from DSS 2011. Lemma 6 is the extension of Lemma 2.13 of DSS 2011.

Lemma 1. *For any log-concave distribution Q with density f, there exist finite constants $B_1 = B_1(Q) > 0$ and $B_2 = B_2(Q) > 0$ such that $f(x) \leq B_1 \exp(-B_2||x||)$ for all $x \in \mathbb{R}^d$.*

Lemma 2. *The following properties of Q are equivalent:*
(a) $csupp(Q)$ has non-empty interior.
(b) $Q(H) < 1$ for any hyperplane $H \subset \mathbb{R}^d$.
(c) With Leb denoting Lebesgue measure on $\mathbb{R}^d$,

$$\lim_{\delta \downarrow 0} \sup\{Q(A) : A \subset \mathbb{R}^d \text{ closed and convex}, Leb(A) \leq \delta\} < 1.$$

Lemma 3. *Let ϕ be the function such that for any $x, y \in interior(dom(\phi))$ and $t \in (0,1)$, if $tx + (1-t)y \in interior(dom(\phi))$, $\phi(tx + (1-t)y) \geq t\phi(x) + (1-t)\phi(y)$ and for $C \subseteq \mathbb{R}^d$, $\int_C e^{\phi(x)} dx \leq 1$. We define $D_q = \{x \in C : \phi(x) \geq q\}$. For any $r < M \leq \max_{x \in \mathbb{R}^d} \phi(x)$,*

$$Leb(D_r) \leq (M-r)^d e^{-M} / \int_0^{M-r} t^d e^{-t} dt.$$

Lemma 4. *Let $\overline{\phi}, \phi_1, \phi_2, \ldots$ be concave functions and $\phi_n \leq \overline{\phi}$. Further we assume the set $H = \{x : \liminf_{n \to \infty} \phi_n(x) > -\infty\}$ is not empty. Then there exist a subsequence $(\phi_{n(k)})_k$ of $(\phi_n)_n$ and a function ϕ such that $H \subset dom(\phi) \stackrel{d}{=} \{\phi > -\infty\}$:*

$$\lim_{k \to \infty, x \to y} \phi_{n(k)}(x) = \phi(y) \text{ for all } y \in interior(dom(\phi)),$$

$$\lim_{k \to \infty, x \to y} \phi_{n(k)}(x) \leq \phi(y) \text{ for all } y \in \mathbb{R}^d.$$

Lemma 5. *Suppose Q_n is a sequence converged to some distribution Q and h be a nonnegative and continuous function, then*

$$\liminf_{n\to\infty} \int h dQ_n \geq \int h dQ.$$

If the stronger statement $\liminf_{n\to\infty} \int h dQ_n = \int h dQ < \infty$ holds, then for any function f such that $|f|/(1+h)$ is bounded,

$$\lim_{n\to\infty} \int f dQ_n = \int f dQ.$$

Lemma 6. *A point $x \in \mathbb{R}^d$ is an interior point of C if and only if*

$$h(Q,x) = \sup\{Q(E) : E \subset C, E \textit{ closed and convex}, x \notin interior(E)\}/Q(C) < 1.$$

Proof For $x \notin interior(E)$ and closed and convex E, there exits a unit vector $u_j \in R^d$ such that E is contained in the closed set H_C which is a subset of C:

$$C \supseteq H_C(x) = \{y \in C : u^T y \leq u^T x\} \supseteq E.$$

By the definition of $h(Q,x)$ we conclude $h(Q,x) \leq Q(H_C)/Q(C) \leq 1$. There are two cases: $E \subset H_C$ and $E = H_C(x)$. For the case $E \subset H_C$, by definition $h(Q,x) < 1$ strictly. For the case $E = H_C(x)$, as we have $x \notin interior(E)$ but $x \in H_C(x)$, we conclude $x \in \partial H_C(x)$. Now if $x \notin interior(C)$, by definition, $h(Q,x) = 1$. On the other hand, if $h(Q,x) = 1$, then $Q(H_C(x)) = Q(C)$, which leads to $C = H_C(x) = E$. Combined with $x \notin interior(H_C(x))$ we can conclude that $x \notin interior(C)$. Consequently, $x \notin interior(C) \iff h(Q,x) = 1$. Thus, $x \in interior(C) \iff h(Q,x) < 1$.

Appendix B: Proof of Theorem 1

The first thing is to prove the finiteness of the log-likelihood type function.

$L(Q)$ is the supreme of $L(\boldsymbol{\phi}, \boldsymbol{\lambda}, \boldsymbol{\pi}, Q)$ over all $\boldsymbol{\phi} \in \boldsymbol{\Phi}, \boldsymbol{\lambda} \in \boldsymbol{\Lambda}, \boldsymbol{\lambda} \in \mathbb{R}^K$. If we take a special case that $\phi_j^*(x) = -(\log \lambda_j^*) - ||x||$, $L(\boldsymbol{\phi}^*, \boldsymbol{\lambda}^*, \boldsymbol{\pi}, Q) = \log K - \int ||x|| dQ > -\infty$. Consequently, $L(Q) > -\infty$.

Now we show $L(Q) < \infty$. As discussed at the end of Section 2, we do restrict our interest to the $\boldsymbol{\phi}$ such that $\int e^{\phi_j(x)} dx = 1$ for $j = 1, \ldots, K$. Consequently, we define the log-density as $l(x) = \log \sum_{j=1}^{K} \lambda_j e^{\phi_j(x)}$ and rewrite the log-likelihood-type function as $L(l, Q) = L(\boldsymbol{\phi}, \boldsymbol{\lambda}, \boldsymbol{\pi}, Q)$. For the convenience of the proof, we define an envelope function $\overline{\phi}(x) = \max_j\{\phi_j(x)\}$, i.e. $\overline{\phi}(x) \geq l(x)$ for every $x \in \mathbb{R}^d$. This function is continuous but not smooth on $d-1$ dimensional boundaries. These boundaries divide the $csupp(Q)$ into K sets: $C_1, \ldots, C_K$. Each set C_j is defined as $C_j = \{x \in \mathbb{R}^d : \overline{\phi}(x) = \phi_j(x)\}$. The sets $C_1, \ldots, C_K$ are disjoint except on the boundaries and $Leb(C_i \cap C_j) = 0$ for every $i \neq j$. For any $x, y \in C_j$ and $t \in (0, 1)$, $\overline{\phi}(tx + (1-t)y) \geq t\overline{\phi}(x) + (1-t)\overline{\phi}(y)$ and $\int_{C_j} e^{\overline{\phi}(x)} dx \leq 1$. We define $M_j(\boldsymbol{\phi})$ and $\mathcal{S}(\boldsymbol{\phi})$ as stated in Section 1. As $L(l, Q) \leq \sum_{j=1}^{K} Q(C_j) M_j$, $M_j > -\infty$, and the restriction $|\mathcal{S}(\boldsymbol{\phi})| \geq \eta > 0$, we focus our interest on $M_j > 0$ and the only case which we have to worry about is all M_j's increasing to infinity. We define $D_q = \{x \in \mathbb{R}^d : \overline{\phi}(x) \geq q\}$. For any $c > 0$,

$$
\begin{aligned}
L(l, Q) &\leq \int \overline{\phi}(x) dQ = \int_{csupp(Q) \setminus D_{-cM_{(1)}}} \overline{\phi}(x) dQ + \int_{D_{-cM_{(1)}}} \overline{\phi}(x) dQ \\
&\leq -cM_{(1)}(1 - Q(D_{-cM_{(1)}})) + M_{(K)} Q(D_{-cM_{(1)}}) \\
&\leq (1 + c\eta)\Big(Q(D_{-cM_{(1)}}) - \frac{c\eta}{c\eta + 1}\Big) M_{(K)}.
\end{aligned}
$$

We can always find sufficient large c such that the set $D_{-cM_{(1)}}$ is a closed and convex subset of $\mathbb{R}^d$. We define the set $D_{j,q} = \{x \in C_j : \overline{\phi}(x) \geq q\} \subset C_j$. Obviously $Leb(D_{-cM_{(1)}}) = \sum_{j=1}^{K} Leb(D_{j,-cM_{(1)}})$. For any $c > 0$, applying Lemma 3 to set $D_{j,-cM_{(1)}}$ and letting $M = M_{(1)}$ yield $Leb(D_{j,-cM_{(1)}}) \leq (1 + c) M_{(1)}^d e^{-M_{(1)}} / (d! + o(1)) \to 0$ as $M_{(1)} \to \infty$ for every $j = 1, \ldots, K$. Consequently, $Leb(D_{-cM_{(1)}}) \to 0$ as $M_{(1)} \to \infty$. By our definition, $\eta \in (0, 1]$. Thus, by Lemma 2, we can find sufficiently large c and small δ such that

$$
\sup\{Q(D) : D \subset \mathbb{R}^d,\ Leb(D) \leq \delta\} < \frac{c\eta}{c\eta + 1}.
$$

Thus, $L(l, Q) \to -\infty$ as $M_{(1)} \to \infty$, which indicates that when all modes of log-concave densities increase to infinity, the log-likelihood is poorly characterized. On the other hand, $L(l, Q) \leq M_{(K)}$. These considerations show that $L(Q)$ is finite and equals the supremum of $L(l, Q)$ for suitable finite M_j's such that $M_j \in [M_{*j}, M_j^*]$ $(j = 1, \ldots, K)$.

Let $\phi_{m,j}$'s and $\lambda_{m,j}$'s form a sequence $l_m(x) = \log \sum \lambda_{m,j} \exp\{\phi_{m,j}(x)\}$ such that $-\infty < L(l_m, Q) \uparrow L(Q)$ as $m \to \infty$. Next, we will prove that for every $j \in \{1, \ldots, K\}$, there exists a point, say, $x_{0,j} \in interior(csupp(Q))$, such that $\liminf_{m\to\infty} \phi_{m,j}(x_{0,j}) > -\infty$.

We define $\overline{\phi}_m(x) = \max_j\{\phi_{m,j}(x)\}$, $C_{m,j} = \{x \in \mathbb{R}^d : \overline{\phi}_m(x) = \phi_{m,j}(x)\}$, and $M_{m,j} = \max_{x\in\mathbb{R}^d} \phi_{m,j}(x)$. For any $j^* \in \{1, \ldots, K\}$, by picking any $x_{0,j^*} \in C_{m,j^*}$ such that $\phi_{m,j^*}(x_{0,j^*}) \in [M'_{m,j^*}, M_{m,j^*})$, where $M'_{m,j^*} = \max_{x\in\partial\{C_{m,j^*}\}} \phi_{m,j^*}(x)$, there exists a sufficient small $\epsilon \geq 0$ such that the set $E_{m,j^*} = \{x \in C_{m,j^*} : \phi_{m,j^*}(x) \geq \phi_{m,j^*}(x_{0,j^*}) + \epsilon\}$ is a closed and convex subset of C_{m,j^*} and x_{0,j^*} is not an interior point of E_{m,j^*}. Thus,

$$
\begin{aligned}
&L(l_m, Q) = \int l_m dQ \leq \int \overline{\phi}_m(x) dQ = \sum\nolimits_{j\neq j^*} \int_{C_{m,j}} \phi_{m,j}(x) dQ + \int_{C_{m,j^*}} \phi_{m,j^*} dQ \\
&\leq \sum\nolimits_{j\neq j^*} M_{m,j} Q(C_{m,j}) + \phi_{m,j^*}(x_{0,j^*}) Q(C_{m,j^*}) + (M_{m,j^*} - \phi_{m,j^*}(x_{0,j^*})) Q(E_{m,j^*}) \\
&\leq \sum\nolimits_{j=1}^{K} \max(M_{m,j}, 0) + \phi_{m,j^*}(x_{0,j^*}) Q(C_{m,j^*})(1 - h_{j^*}(Q, x_{0,j^*})).
\end{aligned}
$$

These inequalities hold for the case of $\phi_{m,j^*}(x_{0,j^*}) = M_{m,j^*}$ as well ($\epsilon = 0$ accordingly). By Lemma 6, $h_{j^*}(Q, x_{0,j^*}) < 1$. Due to the fact that M_{m,j^*} is finite, $interior(C_{m,j^*})$ is not empty. Consequently, $\liminf_{m\to\infty} Q(C_{m,j^*}) > 0$, which yields

$$
\begin{aligned}
\phi_{m,j^*}(x_{0,j^*}) &\geq -\frac{\sum_{j=1}^{K} \max(M_{m,j}, 0) - L(l_m, Q)}{Q(C_{m,j^*})(1 - h_{j^*}(Q, x_{0,j^*}))} \\
&> -\frac{\sum_{j=1}^{K} \max(M^*_j, 0) - L(l_1, Q)}{Q(C_{m,j^*})(1 - h_{j^*}(Q, x_{0,j^*}))} > -\infty.
\end{aligned}
$$

Hence, the set $H_j = \{x : \liminf_{m\to\infty} \phi_{m,j}(x) > -\infty\}$ is not empty for every $j \in \{1, \ldots, K\}$. From Lemma 1 we conclude that for each ϕ_j, we can find suitable finite positive constants $a_j, b_j > 0$ such that $\phi_j(x) \leq a_j - b_j||x|| \leq a - b||x||$, where $a = \max_j a_j > 0$ and $b = \min_j b_j > 0$. Then by Lemma 4, there exist a subsequence $(\phi_{1,m(k_1)})_{k_1}$ of $(\phi_{1,m})_m$ and a concave function ϕ_1 such that:

$$\lim_{k_1\to\infty, x\to y} \phi_{1,m(k_1)}(x) = \phi_1(y) \text{ for all } y \in interior(dom(\phi_1)),$$

$$\lim_{k_1\to\infty, x\to y} \phi_{1,m(k_1)}(x) \leq \phi_1(y) \text{ for all } y \in \mathbb{R}^d.$$

If we define $\phi_1 = -\infty$ on $\mathbb{R}^d \setminus dom(\phi_1)$, then we can rewrite them as:

$$\limsup_{k_1\to\infty} \phi_{1,m(k_1)}(x) \leq \phi_1(x) \quad \text{for all } x \in \partial\{dom(\phi_1)\},$$
$$\lim_{k_1\to\infty} \phi_{1,m(k_1)}(x) = \phi_1(x) \quad \text{for all } x \in \mathbb{R}^d \setminus \partial\{dom(\phi_1)\}.$$

We can find a sub-subsequence in the original subsequence, which has the similar property for $\phi_{2,m(k_2)}$. Keeping doing this sequentially for all $\phi_{m,j}$'s and $\lambda_{m,j}$'s will yield the common subsequence $l_{m(k)}$ and a function $l^*(x) = \log \sum \lambda_j \exp\{\phi_j(x)\}$ such that:

$$\limsup_{k\to\infty} l_{m(k)}(x) \leq l^*(x) \quad \text{for all } x \in \mathcal{P},$$
$$\lim_{k\to\infty} l_{m(k)}(x) = l^*(x) \quad \text{for all } x \in \mathbb{R}^d \setminus \mathcal{P},$$

where $\mathcal{P} = \cup_{j=1}^K \big(\partial\{dom(\phi_j)\}\big)$ and $Leb(\mathcal{P}) = 0$. The next step is to prove that $l^*(x)$ is the maximizer. Applying Fatou's lemma to the subsequence function $l_{m(k)}(x) \leq a - b||x||$ yields

$$\limsup_{k\to\infty} \int l_{m(k)} dQ \leq \int l^* dQ.$$

Hence,

$$L(Q) \geq l(l^*, Q) \geq \limsup_{k\to\infty} L(l_{m(k)}, Q) = L(Q),$$

from which we conclude $L(l^*, Q) = L(Q)$. The first inequality follows by the definition of $L(Q)$. The last equality follows by the definition that $l_{m(k)}$ is a sequence that maximizes $L(l_{m(k)}, Q)$ to $L(Q)$ as $k \to \infty$. Thus, it concludes the existence of the maximizer l^*, which indicates the existence of λ_j^*'s and ϕ_j^*'s.

Appendix C: Proof of Theorem 2

We proof the theorem for a subsequence of Q_n. Let $L(Q_n) \to \Gamma$. As in the proof of Theorem 1, $l_n(x) \leq a - b||x||$ and $\inf \phi_{n,j}(x_0) > -\infty$ for some $x_0 \in interior(csupp(Q))$. Therefore, for a subsequence of $(Q_n)_n$, there exists a function

l^* such that $l_n(y), l^*(y) \leq a - b||y||$, and

$$\limsup_{k\to\infty} l_{n(k)}(x) \leq l^*(x) \quad \text{for all } x \in \mathcal{P},$$
$$\lim_{k\to\infty} l_{n(k)}(x) = l^*(x) \quad \text{for all } x \in \mathbb{R}^d \setminus \mathcal{P}.$$

By Skorohod's theorem, there exists a probability space with random variables $X_n \sim Q_n$, $X \sim Q$ such that $X_n \to X$ almost surely. We define a random variable $H_n = a - b||X_n|| - l_n(X_n) \geq 0$. Applying Fatou's lemma to H_n yields,

$$\begin{aligned}
\Gamma &= \lim_{n\to\infty} \int l_n dQ_n = \lim_{n\to\infty} \int (a - b||x||) dQ_n - E(H_n) = a - b\gamma - \liminf_{n\to\infty} E(H_n) \\
&\leq a - b\gamma - E\left(\liminf_{n\to\infty}(H_n)\right) \leq a - b\gamma - E\left(a - b||X|| - l^*(X)\right) \\
&= b\left(\int ||x|| dQ_0 - \gamma\right) + \int l^*(X) dQ_0 = L(l^*, Q_0) \leq L(Q_0).
\end{aligned}$$

Let $l_0(x) = \log \sum \lambda_j \phi_j(x)$, i.e. λ_j's and ϕ_j's are the results corresponding with l_0. In the following proof we utilize a special approximation scheme. Let $l^{(\epsilon)}(x) = \log \sum \lambda_j^{(\epsilon)} \phi_j^{(\epsilon)}(x)$, $\lambda_j^{(\epsilon)} = \lambda_j$ and $\phi_j^{(\epsilon)} = \inf_{v,c}(v_j^T x + c_j)$ such that $||v_j|| \leq \epsilon^{-1}$ and $\phi_j(y) \leq v_j^T y + c_j$. DSS 2011 shows that the approximation $\phi_j^{(\epsilon)}$ is real valued and Lipschitz continuous with constant ϵ^{-1}. Consequently, $l^{(\epsilon)}(x)$ is also Lipschitz-continuous with constant ϵ^{-1}. Moreover, $\phi_j^{(\epsilon)} \geq \phi_j$ and $\phi_j^{(\epsilon)} \downarrow \phi_j$ pointwise as $\epsilon \downarrow 0$. Thus, $l^{(\epsilon)} \downarrow l_0$ pointwise as $\epsilon \downarrow 0$ and $l^{(1)} \geq l^{(\epsilon)} \geq l_0$ for $\epsilon \in (0, 1)$. With this approximation, it follows from Lipschitz-continuity, $\int ||x|| dQ_0 = \gamma < \infty$, and the stronger version of Lemma 5 that

$$\begin{aligned}
\Gamma &= \lim_{n\to\infty} \int l_n dQ_n \geq \lim_{n\to\infty} L(l^{(\epsilon)}, Q_n) = \lim_{n\to\infty} \int l^{(\epsilon)} dQ_n - \sum \pi_j \int e^{\phi_j^{(\epsilon)}(x)} dx + 1 \\
&= \int l^{(\epsilon)} dQ_0 - \sum \pi_j \int \exp(\phi_j^{(\epsilon)}(x)) dx + 1.
\end{aligned}$$

Applying monotone convergence theorem to function $l^{(1)} - l^{(\epsilon)}$ and dominated convergence theorem to $\exp\{\phi_j^{(\epsilon)}\}$'s yields, $\lim_{\epsilon\to 0^+} L(l^{(\epsilon)}, Q_0) = L(l_0, Q_0)$. Hence, $\Gamma \geq L(Q_0)$. Combining with $\Gamma \leq L(l^*, Q_0) \leq L(Q_0)$ yields $\Gamma = L(Q_0) = L(l^*, Q_0)$, which indicates that l^* equals the maximizer $l_0 = l(\cdot|Q_0)$ that corresponds to $L(Q_0)$.

Applying to density $f_n = \exp\{l_n\}$ and $f_0 = \exp\{l_0\}$ yields,

$$\lim_{n\to\infty, x\to y} f_n(x) = f_0(y) \text{ for all } x \in \mathbb{R}^d \setminus \mathcal{P},$$
$$\lim_{n\to\infty, x\to y} f_n(x) \leq f_0(y) \text{ for all } y \in \mathcal{P},$$

where $\mathcal{P} = \cup_{j=1}^{K}\big(\partial\{f_{0j} > 0\}\big)$ and $Leb(\mathcal{P}) = 0$. Consequently, $(f_n)_n \to f_0$ almost everywhere with respect to Lebesgue measure. In addition, $|f_n(x)| \leq e^{a-b||x||}$, and $\int e^{a-b||x||}dx$ is finite. Applying Lebesgue's dominated convergence theorem yields,

$$\lim_{n\to\infty} \int |f_n(x) - f_0(x)|dx = 0.$$

Consequently, we claim Theorem 2 to be true for a subsequence of the original sequence $(Q_n)_n$. It remains to show it is true for the entire sequence.

Suppose any assertion about f_n is false, then one could replace the initial sequence $(Q_n)_n$ from the start with a subsequence such that one of the following three conditions is satisfied:

(i)$\lim_{n\to\infty} f_n(x_n) > f_0(y)$ for some sequence $(x_n)_n$ converge to point y;

(ii)$\lim_{n\to\infty} f_n(x_n) < f_0(y)$ for some sequence $(x_n)_n$ converge to point y;

(iii)$\lim_{n\to\infty} \int |f_n(x) - f_0(x)|dx > 0$.

Any of these three properties are transmitted to subsequence of $(Q_n)_n$, which would lead to a contradiction.

References

Balabdaoui, Fadoua, & Doss, Charles R. 2014. Inference for a mixture of symmetric distributions under log-concavity. *arXiv preprint arXiv:1411.4708*.

Balabdaoui, Fadoua, Rufibach, Kaspar, & Wellner, Jon A. 2009. Limit distribution theory for maximum likelihood estimation of a log-concave density. *Annals of statistics*, **37**(3), 1299.

Bordes, Laurent, & Vandekerkhove, Pierre. 2010. Semiparametric two-component mixture model with a known component: an asymptotically normal estimator. *Mathematical Methods of Statistics*, **19**(1), 22–41.

Bordes, Laurent, Mottelet, Stéphane, Vandekerkhove, Pierre, *et al.* 2006a. Semiparametric estimation of a two-component mixture model. *The Annals of Statistics*, **34**(3), 1204–1232.

Bordes, Laurent, Delmas, Céline, & Vandekerkhove, Pierre. 2006b. Semiparametric Estimation of a Two-component Mixture Model where One Component is known. *Scandinavian journal of statistics*, **33**(4), 733–752.

Butucea, Cristina, & Vandekerkhove, Pierre. 2014. Semiparametric mixtures of symmetric distributions. *Scandinavian Journal of Statistics*, **41**(1), 227–239.

Chang, George T, & Walther, Guenther. 2007. Clustering with mixtures of log-concave distributions. *Computational Statistics & Data Analysis*, **51**(12), 6242–6251.

Chee, Chew-Seng, & Wang, Yong. 2013. Estimation of finite mixtures with symmetric components. *Statistics and Computing*, **23**(2), 233–249.

Chen, Jiahua, Tan, Xianming, & Zhang, Runchu. 2008. Inference for normal mixtures in mean and variance. *Statistica Sinica*, **18**(2), 443.

Chen, Yining, & Samworth, Richard J. 2013. Smoothed log-concave maximum likelihood estimation with applications. *Statist. Sinica*, **23**, 1373–1398.

Cule, Madeleine, & Samworth, Richard. 2010. Theoretical properties of the log-concave maximum likelihood estimator of a multidimensional density. *Electronic Journal of Statistics*, **4**, 254–270.

Cule, Madeleine, Gramacy, Robert, & Samworth, Richard. 2009. LogConcDEAD: An R package for maximum likelihood estimation of a multivariate log-concave density. *Journal of Statistical Software*, **29**(2), 1–20.

Cule, Madeleine, Samworth, Richard, & Stewart, Michael. 2010. Maximum likelihood estimation of a multi-dimensional log-concave density. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **72**(5), 545–607.

Dempster, Arthur P, Laird, Nan M, & Rubin, Donald B. 1977. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 1–38.

Doss, Charles, & Wellner, Jon A. 2013. Global rates of convergence of the MLEs of log-concave and s-concave densities. *arXiv preprint arXiv:1306.1438*.

Dümbgen, Lutz, & Rufibach, Kaspar. 2009. Maximum likelihood estimation of a log-concave density and its distribution function: Basic properties and uniform consistency. *Bernoulli*, **15**(1), 40–68.

Dümbgen, Lutz, Samworth, Richard, & Schuhmacher, Dominic. 2011. Approximation by log-concave distributions, with applications to regression. *The Annals of Statistics*, **39**(2), 702–730.

Hathaway, Richard J. 1985. A constrained formulation of maximum-likelihood estimation for normal mixture distributions. *The Annals of Statistics*, 795–800.

Hohmann, Daniel, & Holzmann, Hajo. 2013. Semiparametric location mixtures with distinct components. *Statistics*, **47**(2), 348–362.

Hunter, David R, Wang, Shaoli, & Hettmansperger, Thomas P. 2007. Inference for mixtures of symmetric distributions. *The Annals of Statistics*, 224–251.

Kim, Arlene KH, & Samworth, Richard J. 2014. Global rates of convergence in log-concave density estimation. *arXiv preprint arXiv:1404.2298*.

Ma, Yanyuan, & Yao, Weixin. 2015. Flexible estimation of a semiparametric two-component mixture model with one parametric component. *Electronic Journal of Statistics*, **9**, 444–474.

McLachlan, Geoffrey, & Krishnan, Thriyambakam. 2007. *The EM algorithm and extensions*. Vol. 382. John Wiley & Sons.

McLachlan, Geoffrey, & Peel, David. 2000. *Finite mixture models*. John Wiley & Sons.

Mcnicholas, Paul David, & Murphy, Thomas Brendan. 2008. Parsimonious Gaussian mixture models. *Statistics and Computing*, **18**(3), 285–296.

Pal, Jayanta Kumar, Woodroofe, Michael, & Meyer, Mary. 2007. Estimating a Polya Frequency Function. *Lecture Notes-Monograph Series*, 239–249.

Rufibach, Kaspar. 2007. Computing maximum likelihood estimators of a log-concave density function. *Journal of Statistical Computation and Simulation*, **77**(7), 561–574.

Walther, Guenther. 2002. Detecting the presence of mixing with multiscale maximum likelihood. *Journal of the American Statistical Association*, **97**(458), 508–513.

Xiang, Sijia, Yao, Weixin, & Wu, Jingjing. 2014. Minimum profile Hellinger distance estimation for a semiparametric mixture model. *Canadian Journal of Statistics*, **42**(2), 246–267.

Yao, Weixin. 2010. A profile likelihood method for normal mixture with unequal variance. *Journal of Statistical Planning and Inference*, **140**(7), 2089–2098.