Bernoulli **24**(2), 2018, 1053–1071
DOI: 10.3150/16-BEJ864

Inference for a two-component mixture of symmetric distributions under log-concavity

FADOUA BALABDAOUI[1] and CHARLES R. DOSS[2]

[1] *CEREMADE, Université Paris-Dauphine, 75775 Paris, France and Seminar für Statistik, ETH Zürich, 8092 Zürich, Schweiz. E-mail: fadoua.balabdaoui@stat.math.ethz.ch*
[2] *University of Minnesota, School of Statistics, 313 Ford Hall, 224 Church Street SE, Minneapolis, MN 55455, USA. E-mail: cdoss@stat.umn.edu*

In this article, we revisit the problem of estimating the unknown zero-symmetric distribution in a two-component location mixture model, considered in previous works, now under the assumption that the zero-symmetric distribution has a log-concave density. When consistent estimators for the shift locations and mixing probability are used, we show that the nonparametric log-concave Maximum Likelihood estimator (MLE) of both the mixed density and that of the unknown zero-symmetric component are consistent in the Hellinger distance. In case the estimators for the shift locations and mixing probability are $\sqrt{n}$-consistent, we establish that these MLE's converge to the truth at the rate $n^{-2/5}$ in the L_1 distance. To estimate the shift locations and mixing probability, we use the estimators proposed by (*Ann. Statist.* **35** (2007) 224–251). The unknown zero-symmetric density is efficiently computed using the R package `logcondens.mode`.

Keywords: bracketing entropy; consistency; empirical processes; global rate; Hellinger metric; log-concave; mixture; symmetric

1. Introduction

Let us assume that $X_1, \dots, X_n$ are independent and identically distributed (i.i.d.) draws from a mixture distribution, with cumulative distribution function (c.d.f.) G^0 given by

$$G^0(x) = \sum_{i=1}^{k} \pi_i^0 F_i^0(x), \qquad x \in \mathbb{R}, \tag{1}$$

for some integer $k \geq 2$, where F_i^0 are c.d.f.s, $\pi_i^0 \geq 0$, and $\sum_{i=1}^{k} \pi_i^0 = 1$. Such mixture distributions are very common in statistical modeling, in part because a variety of data generating frameworks lead to mixture models; for instance, one common approach to clustering problems leads to estimation of a mixture density [24]. Another reason for this popularity is that they are very flexible and many distributions can be well approximated by some mixture model (see, e.g., [23,38], or [29]).

In this paper, we revisit the semi-parametric mixture model already studied by [6] and [26]. In this model, it is assumed that the mixing distributions F_i, $1 \leq i \leq k$ in (1) are such that

$$F_i(x) = F^0\bigl(x - u_i^0\bigr)$$

1350-7265

for $u_i^0 \in \mathbb{R}$, $i = 1, \dots, k$, and F^0 is a distribution function restricted to be symmetric about 0, that is, $F^0(-x) = 1 - F^0(x-)$. This model was also studied more recently by [7]. All of these authors have actually focused on the case $k = 2$:

$$G^0(x) = \pi^0 F^0\bigl(x - u_1^0\bigr) + \bigl(1 - \pi^0\bigr) F^0\bigl(x - u_2^0\bigr), \qquad x \in \mathbb{R}. \tag{2}$$

This is still, in fact, a flexible model which is useful in many scenarios (see our data applications in Section 6). As the main goal is to estimate the mixing parameters and the mixing component F^0, it is crucial to be assured that there exists a unique solution $(\pi^0, u_1^0, u_2^0, F^0)$ for a given G^0 determined by (2). Bordes, Mottelet and Vandekerkhove [6] and Hunter, Wang and Hettmansperger [26] were able to establish that identifiability holds under some suitable conditions on the mixing parameters. Their result states that if $u_1^0 < u_2^0$ and $\pi^0 \notin \{0, 1/2, 1\}$, then G^0 given (2) is identifiable for any zero-symmetric distributions F^0. Furthermore, the condition is necessary and sufficient since any distribution G^0 that is symmetric about its median clearly cannot be 2-identifiable; see Theorem 2 of [26].

After having shown identifiability, [26] put their focus on estimating $(\pi^0, u_1^0, u_2^0, F^0)$. They have shown that their estimator of the parametric component (π^0, u_1^0, u_2^0) is consistent and asymptotically normal. However, the obtained estimator of F^0 is not even guaranteed to have the properties of a genuine c.d.f. (i.e., it is not necessarily nondecreasing). On the other hand, [6,7], and [9] use a KDE approach to estimation of F^0. The resulting estimators are proper distribution functions, but the procedures involve a model-selection procedure (cross-validation or Akaike or Bayesian information criterion) to choose the tuning parameter. The estimators of [26] and [7] for the mixture parameters are shown to converge weakly to a multivariate Gaussian at the parametric rate $n^{-1/2}$ under some regularity conditions on F^0 which are related to smoothness in the case of [7]. Bordes, Mottelet and Vandekerkhove [6] obtain also a convergence rate under smoothness assumptions, but their rate of convergence is much slower (of order $n^{-1/4+\alpha}$, for any $\alpha > 0$). Bordes, Mottelet and Vandekerkhove [6] show that the same rates of convergence are inherited by their kernel estimator of F^0 in the supremum norm, under the assumption that the location parameters u_1^0 and u_2^0 are unknown. If F^0 is assumed to admit a density f^0, then [6] provide only almost sure consistency in the supremum norm. For their kernel estimator, [7] obtain, for pointwise convergence, a rate of order $n^{-(2\beta-1)/(4\beta)}$ in the quadratic risk assuming smoothness of level $\beta > 1/2$ and assuming that the bandwidth is chosen optimally (the authors suggest using cross validation).

Hence, the proposed estimators of F^0 in the aforementioned works suffer various practical difficulties, including slow rates of convergence (or as-of-yet unknown rates) for it or its density, the estimator not being a proper c.d.f., or the need for model-selection procedures to choose a tuning parameter. Our goal in this paper is to circumvent those issues by constructing an estimator of the density f^0 which

- converges to the truth with a provably good convergence rate,
- can be efficiently computed,
- does not require a tuning parameter, and furthermore,
- is unimodal.

Unimodality is a natural constraint to enforce; when using a mixture model, it is somewhat unnatural to imagine a multimodal mixture component density. However, using unimodality involves some technical difficulties: enforcing unimodality on f^0 is not directly feasible, because the class of unimodal densities is too large and the MLE of a unimodal density fails to exist even in the simple one-dimensional setting (with no mixing). We propose instead to assume that f^0 satisfies the shape constraint of log-concavity (i.e., $\log f^0$ is concave).

Log-concave functions are always unimodal, and have been used to great success in nonparametric modeling. Unlike the class of unimodal densities, the log-concave class admits an MLE [39]. Many papers have studied the log-concave MLE on $\mathbb{R}$ or $\mathbb{R}^d$ and much is already known about its large sample properties, both local and global; see, for example, [5,10,13,14,16,18,20,31,35], and [27]. Balabdaoui *et al.* [4] studied asymptotics and confidence intervals of the discrete log-concave MLE of a probability mass function in the well- and misspecified settings. Dümbgen, Hüsler and Rufibach [17], Rufibach [33] and Dümbgen and Rufibach [19] study algorithms for computation of the MLE, allowing unequal weights to be assigned to the observations, an important feature of which we will take advantage.

In the present context, we need to consider the class of *zero-symmetric* log-concave densities on $\mathbb{R}$, which has not been considered before. To do so, we note that if f is zero-symmetric and log-concave on $\mathbb{R}$, then $f^+(t) := 2f(t)\mathbb{I}_{t\in[0,\infty)}$ is log-concave with mode at 0. Thus, through a simple transformation of the data, it can be shown that the original estimation problem is equivalent to maximizing the log-likelihood over the class of log-concave densities on $[0,\infty)$ with mode at 0. We can then compute the maximum of the log-likelihood easily by alternating between the EM algorithm [15] and the active set algorithm provided in the R package `logcondens.mode` which computes the log-concave MLE with a fixed mode. We use the fact that the active set algorithm allows for unequal weights to be assigned to the data points: here, the weights assigned are proportional to the posterior probabilities from the EM algorithm.

We are able to show that the zero-symmetric log-concave MLE converges in probability to the true zero-symmetric log-concave component density in the Hellinger distance and in the supremum norm on sets of continuity of the true density. Furthermore, it can be shown that our estimator converges to the truth at the rate $n^{-2/5}$ in the L_1-distance. Although the risk measure we use here is different from the one considered by [7], it seems that the rate of convergence of our MLE, when the true mixture component is log-concave, is faster than that given in their Theorem 4 for their KDE when the smoothness parameter β satisfies $\beta \in (1/2, 5/2)$. Note for an estimator $\widehat{g}_n$ of g_0 in the direct density estimation problem based on i.i.d. observations from g_0 (as opposed to the mixture setting) when g_0 has smoothness β the optimal pointwise rate of convergence of $|\widehat{g}_n(x_0) - g_0(x_0)|$ at a fixed point x_0 is $n^{-\beta/(2\beta+1)}$ [37]. Note also that [18] find a rate of convergence of $(\log n)^{\frac{\beta}{2\beta+1}} n^{-\frac{\beta}{2\beta+1}}$ in the uniform norm on compact sets for the log-concave MLE, in the direct density estimation problem, when the true density g_0 is log-concave and also lies in a Hölder class with smoothness β, that is,

$$\begin{aligned} \left|g_0(x) - g_0(y)\right| &\le L|x-y|, \qquad \text{if } \beta = 1, \\ \left|g_0'(x) - g_0'(y)\right| &\le L|x-y|^{\beta-1}, \qquad \text{if } \beta > 1 \end{aligned}$$

for some $L > 0$. This rate is optimal for nonparametric estimation with smoothness β (the log factor being due to the supremum norm [25]), and no bandwidth needs to be chosen.

We note that, although we refer to our estimator as the log-concave MLE, we do not use a “pure” maximum likelihood approach since we feed in other estimators of (π^0, u_1^0, u_2^0) to our likelihood, which we maximize to estimate f^0 and thus g^0, the density of the mixed distribution G^0. An alternative approach is to estimate both the parametric and nonparametric components simultaneously by maximum likelihood. However, there are many additional difficulties in that approach, due to the complicated non-concave nature of the log-likelihood function; see Section 2.

We also note that we are not the first to use log-concavity in mixture modeling; [8] and [22] consider univariate mixtures of log-concave densities, and [14] consider multivariate mixtures of log-concave densities. However, in none of those settings was symmetry imposed, perhaps because the authors were not worried about the (often fundamental) question of identifiability. Thus, their work does not directly apply in our setting.

The paper will be structured as follows. In Section 2, we establish existence of the MLE and provide a necessary condition for a candidate to be equal to the estimator. In Section 3, we establish consistency in the Hellinger distance. This implies other forms of consistency by the results of [13]. The techniques we used are re-adapted from [13,31] and [35] to deal with the additional difficulties of a mixture model. In Section 4, we find that the MLEs of f^0 and g^0 converge to the truth at a rate of order $n^{-2/5}$. In Section 5, we develop a likelihood ratio procedure based on our estimator in the problem of testing absence of mixing. We also consider the problem of clustering where we use the estimators of the posterior probabilities obtained via our log-concave MLE. In both problems, we compare our method to alternative or existing approaches. In Section 6, we present two data applications. Section 7 gathers some conclusions. Proofs and technical details can be found in the supplementary material [3].

2. The model and estimation via maximum likelihood

Let $X_1, \ldots, X_n$ to be n independent observations assumed to come from the location mixture with c.d.f. G^0 which we now assume has a density, given by

$$g^0(x) = \pi^0 f^0\bigl(x - u_1^0\bigr) + \bigl(1 - \pi^0\bigr) f^0\bigl(x - u_2^0\bigr), \tag{3}$$

for some $\pi^0 \in (0, 1) \setminus \{1/2\}$, $u_1^0, u_2^0 \in \mathbb{R}$ such that $u_1^0 \neq u_2^0$. We assume that f^0 is a zero-symmetric log-concave density, that is, $f^0 \in \mathcal{SLC}_1$ where

$$\mathcal{SLC}_1 := \mathcal{SLC} \cap \left\{ f : \int_{\mathbb{R}} f(u)\,du = 1 \right\}, \quad \text{and} \quad \mathcal{SLC} := \bigl\{ e^{\psi} : \psi \in \mathcal{SC} \bigr\},$$

and $\mathcal{SC}$ is the class of concave functions on $\mathbb{R}$ that are upper semi-continuous ("closed") and proper [32], and satisfy $\psi(x) = \psi(-x)$. The upper semi-continuity condition is made only for the purpose of uniqueness. Then

$$L(\pi, u_1, u_2, f) := \sum_{j=1}^{n} \log\bigl(\pi f(X_j - u_1) + (1 - \pi) f(X_j - u_2)\bigr) \tag{4}$$

is the log-likelihood in this problem. In the case of estimation of a log-concave density on $\mathbb{R}$, the log-likelihood is a concave function [18,31,33]. However, [21] study a semiparametric model incorporating log-concavity and find a non-concave likelihood; see their Section 3.3 including a plot on page 18. Unfortunately, our objective function L is also far from concave. Consider order statistics $X_{(1)}, \dots, X_{(n)}$, a fixed $\pi \in (0,1) \setminus \{1/2\}$, and (zero-symmetric log-concave) f with support given by $\operatorname{supp}(f) = [-s, s]$ and $\inf_{x \in \operatorname{supp}(f)} f(x) > 0$. Assume $u_1 < u_2$ are such that $[u_1 - s, u_1 + s] \cup [u_2 - s, u_2 + s] \supset [X_{(1)}, X_{(n)}]$ so that $L(\pi, u_1, u_2, f) > -\infty$. Let $j := \min\{i : X_{(i)} > u_2 - s\}$ be the index of the smallest order statistic contained in the support of the second component, and let $\tilde{\delta} := X_{(j)} - (u_2 - s)$. Then not only does $L(\pi, u_1, \cdot, f)$ fail to be concave, but it is in fact discontinuous at $u_2 + \tilde{\delta}$.

We now describe our estimation approach. Let $\check{\pi}_n, \check{u}_{1,n}, \check{u}_{2,n}$ be estimators of π, u_1, u_2, where we assume $\check{\pi}_n \in (0,1) \setminus \{1/2\}$ and $\check{u}_{1,n} < \check{u}_{2,n}$. We will generally think of these estimators as being $\sqrt{n}$-consistent. We will then consider maximizing the log-likelihood

$$f \mapsto \sum_{j=1}^{n} \log\bigl(\check{\pi}_n f(X_j - \check{u}_{1,n}) + (1 - \check{\pi}_n) f(X_j - \check{u}_{2,n})\bigr) \tag{5}$$

over $\mathcal{SLC}_1$. Using the Lagrange penalty term introduced by [36], this is equivalent to maximizing the criterion Φ_n defined as

$$\Phi_n(\psi) = \frac{1}{n} \sum_{j=1}^{n} \log\bigl[\check{\pi}_n e^{\psi(X_i - \check{u}_{1,n})} + (1 - \check{\pi}_n) e^{\psi(X_i - \check{u}_{2,n})}\bigr] - \int_{\mathbb{R}} e^{\psi(x)}\, dx$$

over $\mathcal{SC}$. We will abusively use the term MLE for our estimators of f^0 and g^0 despite the fact that the mixing parameters (π, u_1, u_2) are not a part of the space over which the likelihood is maximized. In the next proposition, we establish existence of the MLE, and describe its nature.

Proposition 2.1. *The criterion* Φ_n *admits a maximizer* $\widehat{\psi}_n$. *Furthermore, the following holds true almost surely, letting* $\widehat{f}_n = e^{\widehat{\psi}_n}$.

- $\widehat{f}_n$ *is in* $\mathcal{SLC}_1$.
- *For* $i = 1, \dots, n$, *let*

$$Z_{2i-1} = |X_i - \check{u}_{1,n}| \quad \text{and} \quad Z_{2i} = |X_i - \check{u}_{2,n}|. \tag{6}$$

Then, on $[0, \infty)$ *the MLE* $\widehat{\psi}_n$ *changes slope only at points belonging to the set*

$$\{Z_1, Z_2, \dots, Z_{2n-1}, Z_{2n}\}.$$

Furthermore, $\widehat{\psi}_n'(0) = 0$, *and* $\widehat{\psi}_n(x) = -\infty$ *if and only if* $x \notin [-Z_{(2n)}, Z_{(2n)}]$ *where* $Z_{(2n)}$ *is the largest order statistic of* $Z_1, Z_2, \dots, Z_{2n-1}, Z_{2n}$.

All proofs are deferred to the supplementary material [3]. The MLE of f will be denoted by $\widehat{f}_n$ throughout, and that of g by $\widehat{g}_n$. In the following, we give a necessary condition for a log-concave function $f = \exp(\psi)$ to be the MLE.

Proposition 2.2 is interesting to compare with the characterization of [18] for the log-concave MLE. The result is also useful in combination with the EM-algorithm described below as its non-fulfillment indicates that convergence is not yet reached.

Proposition 2.2. *Let* ψ *be a zero-symmetric concave function on* $\mathbb{R}$ *such that* $\psi(x) = -\infty$ *if and only if* $x \notin [-Z_{(2n)}, Z_{(2n)}]$ *where* $Z_{(2n)}$ *is defined in Proposition* 2.1, *and* $\psi'(0) = 0$. *If* $\exp(\psi) = \widehat{f}_n$ *is the MLE, then for any real zero-symmetric function* Δ *such that* $\psi + \varepsilon\Delta \in \mathcal{SC}$ *for some* $\varepsilon > 0$ *we have that*

$$\frac{1}{n}\sum_{i=1}^{n}\bigl\{\widehat{p}_n(X_i)\Delta(X_i - \check{u}_{1,n}) + \bigl(1 - \widehat{p}_n(X_i)\bigr)\Delta(X_i - \check{u}_{2,n})\bigr\} \le \int_{\mathbb{R}} \widehat{f}_n(x)\Delta(x)\,dx, \tag{7}$$

where

$$\widehat{p}_n(X_i) = \frac{\check{\pi}_n \widehat{f}_n(X_i - \check{u}_{1,n})}{\check{\pi}_n \widehat{f}_n(X_i - \check{u}_{1,n}) + (1 - \check{\pi}_n)\widehat{f}_n(X_i - \check{u}_{2,n})} = \frac{\check{\pi}_n \widehat{f}_n(X_i - \check{u}_{1,n})}{\widehat{g}_n(X_i)}, \tag{8}$$

for $i = 1, \dots, n$.

Next, we give the condition in (7) under an alternative form. Dümbgen and Rufibach [18] shows that the log-concave MLE is uniquely characterized by the fact that the first integral of the c.d.f. of the MLE stays below the first integral of the empirical distribution, while touching it exactly at the points where the logarithm of the MLE changes slope. To derive a related result, let $\widehat{\mathbb{F}}_n$ denote the c.d.f. of the discrete distribution putting mass $\widehat{p}_n(X_i)/n$ at Z_{2i-1} and $(1 - \widehat{p}_n(X_i))/n$ at Z_{2i} for $i = 1, \dots, n$, where $\widehat{p}_n(X_i)$ was defined in (8) and Z_i was defined in (6). That is,

$$\widehat{\mathbb{F}}_n = \frac{1}{n}\sum_{i=1}^{n}\bigl(\widehat{p}_n(X_i)\delta_{Z_{2i-1}} + \bigl(1 - \widehat{p}_n(X_i)\bigr)\delta_{Z_{2i}}\bigr),$$

where $\delta_x(t) = 1_{\{[x,\infty)\}}(t)$. Let $\widehat{f}_n^+(x) = 2\widehat{f}_n(x)1_{x\in[0,\infty)}$, $\widehat{\psi}_n^+ = \log(\widehat{f}_n^+)$ and let $\widehat{F}_n^+$ be the c.d.f. of $\widehat{f}_n^+$.

Proposition 2.3. *If* $\widehat{f}_n$ *is the MLE of the component* $f^0 \in \mathcal{SLC}_1$ *then*

$$\int_0^z \widehat{F}_n^+(x)\,dx \begin{cases} \le \int_0^z \widehat{\mathbb{F}}_n(x)\,dx, & \text{for } z \in [0, Z_{(2n)}], \\ = \int_0^z \widehat{\mathbb{F}}_n(x)\,dx, & \text{if } \widehat{\psi}_n^+(z-) > \widehat{\psi}_n^+(z+). \end{cases} \tag{9}$$

3. Consistency

The main result of this section is to establish consistency in the Hellinger distance of the MLEs $\widehat{g}_n$ and $\widehat{f}_n$ as $n \to \infty$, where the Hellinger distance is defined by

$$H(p,q) := \sqrt{(1/2)\int \bigl(\sqrt{p(x)} - \sqrt{q(x)}\bigr)^2\,dx}.$$

We will also find consistency for $\widehat{f}_n$ in certain exponentially weighted metrics. Our approach to the problem follows the idea of [31] and [18] but will require handling carefully the extra complexity induced by the mixture. As in [13,31] and [35], we will first need to show that the MLE of the mixed density and hence the MLE of the log-concave component are bounded. Here, the claimed boundedness will be only in probability, which is weaker than the almost sure boundedness proved in the aforementioned articles. Those articles, however, were able to take advantage of the fact that the level sets of a bounded unimodal function are convex and compact; such a statement does not hold if we consider a mixture of two unimodal functions instead of a single unimodal function, even if the two components are log-concave. So, instead of studying how the empirical distribution behaves over the class of compact intervals, we will instead need to study its behavior over more complicated classes of functions. This is what is done in Propositions A.2 and A.3.

Theorem 3.1. *Let g^0 be as in (3) and $\widehat{g}_n$ be the MLE of g^0. Then we have that*

$$H\left(\widehat{g}_n, g^0\right) = o_p(1).$$

Consistency of the log-concave component, $\widehat{f}_n$, follows now from Theorem 3.1.

Corollary 3.1. *Let f^0 denote again the true log-concave zero-symmetric density. Then,*

$$H\left(\widehat{f}_n, f^0\right) = o_p(1),$$

and for any $a \in (0, a_0)$ such that $f^0(x) \le \exp(-a_0 x + b)$ for some $b \in \mathbb{R}$, then

$$\int_{\mathbb{R}} e^{at} \left| \widehat{f}_n(t) - f^0(t) \right| = o_p(1),$$

and

$$\sup_{t \in [-A, A]} e^{at} \left| \widehat{f}_n(t) - f^0(t) \right| = o_p(1)$$

on any continuity set $[-A, A]$ of f^0, where A may be ∞ if f^0 is continuous on all of $\mathbb{R}$.

4. Rates of convergence

In this section, we aim at refining the convergence result obtained in the previous section to attain a rate of convergence for both $\widehat{f}_n$ and $\widehat{g}_n$ in the L_1 distance. To this goal, we need first to recall some definitions from empirical processes theory. Given a class of functions $\mathcal{F}$, the bracketing number of $\mathcal{F}$ under some distance $\|\cdot\|$ is defined as

$$N_{[\,]}\left(\varepsilon, \mathcal{F}, \|\cdot\|\right) = \min\left\{k : \exists \underline{f}_1, \bar{f}_1, \dots, \underline{f}_k, \bar{f}_k \text{ s.t. } \|\underline{f}_i - \bar{f}_j\| \le \varepsilon, \mathcal{F} \subset \bigcup_{i=1}^{k} [\underline{f}_i, \bar{f}_i]\right\},$$

where $[l, u] = \{f :\in \mathcal{F} : l \le f \le u\}$. In this section, we refine the consistency result above by deriving the rate of convergence of the MLE's $\widehat{g}_n$ and $\widehat{f}_n$ of the mixed density and the zero-symmetric log-concave component, respectively.

For fixed $M > 0$, $a_0 < b_0$ and $\delta \in (0, (b_0 - a_0)/2)$, consider the class of functions

$$\begin{aligned}\mathcal{G} = \big\{&\lambda f(\cdot - a) + (1-\lambda) f(\cdot - b),\ f \in \mathcal{SLC},\ f(0) \in [1/M, M], \lambda \in [0, 1],\\ &(a, b) \in [a_0 - \delta, a_0 + \delta] \times [b_0 - \delta, b_0 + \delta]\big\}.\end{aligned}$$

Here, the parameters a_0 and b_0 play the role of the true location shifts u_1^0 and u_2^0. Consistency of the estimates $\check{u}_{1,n}$ and $\check{u}_{2,n}$ ensures that they are stay within distance 2δ from the truth with increasing probability. Also, uniform consistency of the log-concave MLE, $\widehat{f}_n$, on continuity sets of f^0 implies consistency at the point 0 (the common mode of f^0 and $\widehat{f}_n$). Thus, we can find $M > 0$ such that $\widehat{f}_n(0) \in [1/M, M]$ with increasing probability. The following proposition gives a bound on the bracketing entropy for the class $\mathcal{G}$.

Proposition 4.1. *For $\varepsilon \in (0, \varepsilon_0]$, we have that*

$$\log N_{[\,]}(\varepsilon, \mathcal{G}, H) \lesssim \frac{1}{\sqrt{\varepsilon}},$$

where ε_0 and $\lesssim$ depend only on a_0, b_0, δ and M.

Now, we are ready to state our main theorem. We find a rate of convergence of at least $n^{-2/5}$ in the L_1 norm, both for $\widehat{f}_n$ and $\widehat{g}_n$. Although we consider L_1 and [7] consider L_2 distance, the rate of Theorem 4.1 is an improvement over the corresponding L_2 rate $n^{-(2\beta-1)/(4\beta)}$ of [7] whenever $\beta < 5/2$. (Note that (log-)concave functions are Lebesgue-almost-everywhere twice differentiable by Alexandrov's theorem [30], so roughly correspond to β being 2 or larger.)

Theorem 4.1. *Let $\widehat{f}_n$ and $\widehat{g}_n$ be again the MLE's of the zero-symmetric log-concave component and mixed density, respectively. If $\sqrt{n}(\check{u}_{1,n} - u_1^0) = O_p(1)$, $\sqrt{n}(\check{u}_{2,n} - u_1^0) = O_p(1)$, and $\sqrt{n}(\check{\pi}_n - \pi^0) = O_p(1)$, then*

$$L_1\big(\widehat{f}_n, f^0\big) = O_p\big(n^{-2/5}\big), \quad \textit{and} \quad L_1\big(\widehat{g}_n, g^0\big) = O_p\big(n^{-2/5}\big),$$

where $L_1(d_1, d_2) = \int_{\mathbb{R}} |d_1(x) - d_2(x)|\,dx$.

To illustrate the theory, a simulated example is given in Figure 1. The true zero-symmetric component f^0 is taken to be the density of a standard Gaussian with mixing probability $\pi^0 = 1/3$ and shift locations $u_1^0 = 0$ and $u_2^0 = 4$. The plot on the top (bottom) shows our MLE of g^0 (f^0) based on a sample of size $n = 500$. The bullets in the right plot depict the knot points of the zero-symmetric log-concave MLE, that is, the points where the logarithm of the log-concave MLE changes its slope.

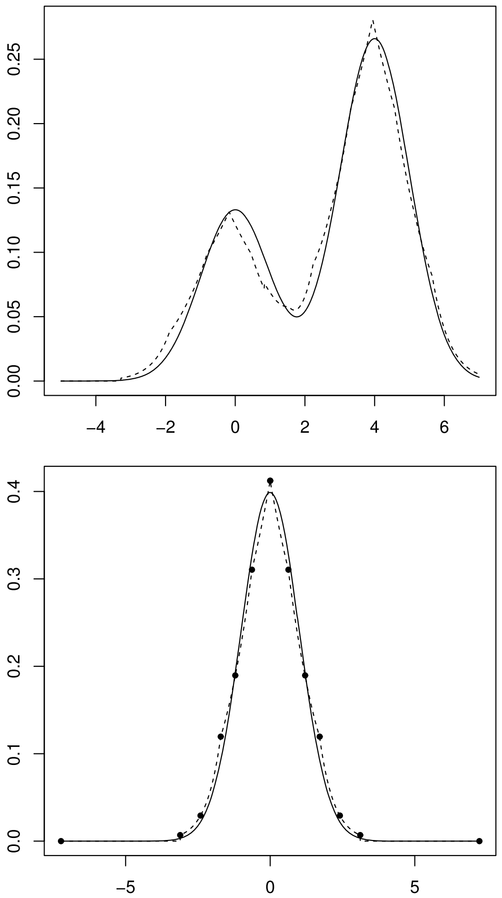

Figure 1. Top: plot of the true density of $(1/3)\mathcal{N}(0,1) + (2/3)\mathcal{N}(4,1)$ (solid line) and its log-concave estimator $\widehat{g}_n$ (dotted line). Bottom: plot of the density of $\mathcal{N}(0,1)$ and its zero-symmetric log–concave MLE $\widehat{f}_n$. The MLE was based on $n = 500$ independent data drawn from the mixture density $(1/3)\mathcal{N}(0,1) + (2/3)\mathcal{N}(4,1)$.

5. Testing and clustering

Bordes, Mottelet and Vandekerkhove [6], Hunter, Wang and Hettmansperger [26] and Butucea and Vandekerkhove [7] propose three different ways of estimating the mixture parameters π^0, u_1^0 and u_2^0. As we are interested here in $\sqrt{n}$-consistent estimators of these parameters, we prefer the work by [26] and [7]. Also, due to some numerical instabilities encountered when computing the estimators proposed by [7], we adopt the approach of [26] which has been already implemented in R; one could either used the code posted at http://www.stat.psu.edu/~dhunter/code or the function in *the mixtools* package. The latter option was kindly brought to the attention of the first author by David Hunter in a private communication.

Once the estimates $\check{\pi}_n$, $\check{u}_{1,n}$ and $\check{u}_{2,n}$ of π^0, u_1^0 and u_2^0 are computed, we maximize

$$f^+ \mapsto \frac{1}{n}\sum_{i=1}^{n} \log\bigl(\check{\pi}_n f^+(Z_{2i-1}) + (1-\check{\pi}_n) f^+(Z_{2i})\bigr),$$

where $f^+ = 2f1_{[0,\infty)}$, $Z_{2i-1} = |X_i - \check{u}_{1,n}|$ and $Z_{2i} = |X_i - \check{u}_{2,n}|$ for $i = 1, \dots, n$. This is equivalent to maximizing (5), which we have shown to admit a maximizer. Since the log-likelihood is not concave and it is not clear how to maximize it directly, we will appeal to the EM algorithm [15]. Although we are fixing the parameters π, u_1, and u_2 we may still introduce the standard-in-mixture-models complete data of (X_i, Δ_i), where $\Delta_i \sim \text{Bernoulli}(\pi)$ and $X_i|\{\Delta_i = 1\} \sim f^0(\cdot - \check{u}_{1,n})$ and $X_i|\{\Delta_i = 0\} \sim f^0(\cdot - \check{u}_{2,n})$. An iteration of the EM algorithm in this setup is then, given an estimate $\widehat{f}_n^{+,(r)}$, to compute

$$\widehat{f}_n^{+,(r+1)} := \operatorname*{argmax}_{f^+} \sum_{i=1}^{n} \bigl\{\widehat{p}_{n,i}^{(r)} \log f^+(Z_{2i-1}) + \bigl(1 - \widehat{p}_{n,i}^{(r)}\bigr) \log f^+(Z_{2i})\bigr\}, \tag{10}$$

where the argmax is over log-concave densities on $[0, \infty)$ with mode at 0, and

$$\widehat{p}_{n,i}^{(r)} = \frac{\check{\pi}_n \widehat{f}^{+,(r)}(Z_{2i-1})}{\check{\pi}_n \widehat{f}^{+,(r)}(Z_{2i-1}) + (1-\check{\pi}_n)\widehat{f}^{+,(r)}(Z_{2i})}$$

is the conditional expectation of Δ_i given X_i. To initialize the EM algorithm, we start with the density of a centered Gaussian distribution with variance equal to the estimate given in formula (11) of [26] for the true variance of the zero-symmetric component, that is $\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2 - \check{\pi}_n(1-\check{\pi}_n)(\check{u}_{2,n} - \check{u}_{1,n})^2$, or 1 if this estimate is negative (this may occur for moderate sample sizes). The argmax in (10) can be computed by the R package `logcondens.mode`.

5.1. Testing the absence of mixing

Recall that the mixing model we consider in this paper is given by

$$g^0 = \pi^0 f^0\bigl(\cdot - u_1^0\bigr) + \bigl(1 - \pi^0\bigr) f^0\bigl(\cdot - u_2^0\bigr)$$

with f^0 a log-concave zero-symmetric density on $\mathbb{R}$, $\pi^0 \notin \{0, 1/2, 1\}$ and $u_1^0 < u_2^0$. We now use our log-concave MLE to test for the absence of mixing, i.e. to test for the null hypothesis that $u_1^0 = u_2^0$, against the alternative that $u_1^0 \neq u_2^0$ and $\pi^0 \neq 1/2$ under the assumption that f^0 is zero-symmetric and log-concave.

To test for mixing, we consider the likelihood ratio statistic. Under the null hypothesis, we take the estimator of the true density to be equal to the log-concave MLE which is symmetric around the median of the data. If $\widehat{g}_n^0$ denotes this estimator, then our test statistic is given by

$$\Lambda_n = \frac{\prod_{i=1}^{n} \widehat{g}_n(X_i)}{\prod_{i=1}^{n} \widehat{g}_n^0(X_i)}. \tag{11}$$

The null hypothesis is then rejected when Λ_n is too large. We use the null hypothesis estimator to find critical values; that is, we bootstrap from the symmetric log-concave estimator $\widehat{g}_n^0$. The critical values of Λ_n are then computed in the usual way: based on the bootstrapped samples from $\widehat{g}_n^0$, we compute the estimators of the mixing probability and mixture locations and the corresponding MLE $\widehat{g}_n$. The order statistics of the bootstrapped values of the likelihood ratio are then obtained to compute upper empirical quantiles of a given order. We also compare our test for mixing (hereafter referred to as the LR test) to the following procedures:

- the naive symmetric bootstrap (NSBS): we re-sample with replacement n random variables $Z_1^*, \dots, Z_n^*$ from $\{\pm|X_1 - \widehat{m}_n|, \dots, \pm|X_n - \widehat{m}_n|\}$ with $\widehat{m}_n$ the median of $X_1, \dots, X_n$ and set $X_i^* = \widehat{m}_n + Z_i^*, i = 1, \dots, n$. Then, the bootstrapped estimators of the location mixture of [26], u_1^* and u_2^*, are computed based on $Y_i^*, i = 1, \dots, n$. We repeat this procedure B times and compute the empirical $(1-\alpha)$-quantile of the distribution of $u_2^* - u_1^*$. The null hypothesis is rejected if the observed $\check{u}_{2,n} - \check{u}_{1,n}$ is larger than this quantile.
- the naive symmetric bootstrap based on symmetric kernel density estimation (NSBSKDE): the method is similar to the one described above except that a standard kernel density estimator is fitted to $\widehat{m}_n \pm |X_i - \widehat{m}_n|$ and $X_1^*, \dots, X_n^*$ are now drawn from the fitted estimator at each bootstrap iteration.
- the likelihood ratio based on symmetric kernel density estimation (LRSKDE): two kernel density estimators are computed, one under the full model, that is, based on $X_1, \dots, X_n$, and one under the null model, that is, based on $\widehat{m}_n \pm |X_i - \widehat{m}_n|$. The likelihood ratio of these estimators is then computed. Bootstrap samples are obtained by simulation from the kernel estimator under the null hypothesis and then the empirical $(1-\alpha)$-quantile of the likelihood ratio is thereby computed. The null hypothesis is rejected if the observed likelihood ratio is larger that this quantile.

Note that the NSBS provides a comparison procedure not based on density estimation of the components. In assessing the power, we take the true zero-symmetric component f^0 to be one of the following distributions: (1) a standard Gaussian, (2) a double exponential, and (3) a uniform on $[-1, 1]$, Also, we take the true parameters to be $\pi^0 \in \{0.20, 0.40\}$ and $(u_1^0, u_2^0) \in \{(0, 0), (0, 1), (0, 3)\}$. We give the estimated probability of rejecting the null hypothesis based on $R = 500$ replications with $B = 49$ bootstrap samples in Table 1, for $n = 250$. The simulation results show the LR and LRSKDE tests are both outperforming the NSBS and NSBSKDE with power nearly equal or equal to 1 for the well-separated mixtures. However, all the considered tests seem to have a level larger than the specified level $\alpha = 0.1$ for the uniform distribution. Further simulations, which we do not report here, show that this improves when the sample size is increased to $n = 500$. Note that the mixtures with mixture probability $\pi^0 = 0.4$ are more difficult to distinguish than those with $\pi^0 = 0.2$. This is to be expected as the former mixtures are close to being symmetric around the mid-point $(u_1^0 + u_2^0)/2$.

It would be interesting to know whether the level of our testing procedure based on the bootstrapped likelihood ratio test equals the theoretical level. The problem is however far from being trivial. Deriving the asymptotic level for example would require establishing the limit distribution of our statistic under the null hypothesis and also showing that it admits a continuous cumulative distribution function. Establishing such results requires a thorough study of the global asymptotics of the log-concave MLE. As this is outside the scope of this paper, the question remains open.

Table 1. Values of the bootstrapped power for LR, NSBS, NSBSKDE and LRSKDE tests when the true density is $\pi^0 f^0(\cdot - u_1^0) + (1-\pi^0) f^0(\cdot - u_2^0)$, where $u_1^0 = 0$, $u_2^0 - u_1^0 \in \{0, 1, 3\}$, $\pi^0 \in \{0.2, 0.4\}$, and f^0 is one of the zero-symmetric log-concave densities shown in the first column. The nominal level is $\alpha = 0.1$. The sample size is $n = 250$, the number of bootstraps and the number of replications were taken to be $B = 49$ and $R = 500$, respectively. The common value of the power under H_0 is replaced by "$*$"

			$u_2^0 - u_1^0$		
Distribution	π^0	Test	0	1	3
$\mathcal{N}(0, 1)$	0.2	LR	0.11	0.15	1.00
		NSBS	0.06	0.07	0.34
		NSBSKDE	0.07	0.08	0.40
		LRSKDE	0.11	0.11	1.00
	0.4	LR	$*$	0.14	0.86
		NSBS	$*$	0.05	0.28
		NSBSKDE	$*$	0.08	0.26
		LRSKDE	$*$	0.11	0.87
$\mathcal{L}(1)$	0.2	LR	0.11	0.22	0.99
		NSBS	0.11	0.01	0.01
		NSBSKDE	0.13	0.11	0.02
		LRSKDE	0.12	0.19	0.98
	0.4	LR	$*$	0.14	0.89
		NSBS	$*$	0.08	0.12
		NSBSKDE	$*$	0.08	0.03
		LRSKDE	$*$	0.18	0.86
$\mathcal{U}[-1, 1]$	0.2	LR	0.19	0.92	1.00
		NSBS	0.11	0.07	0.16
		NSBSKDE	0.11	0.08	0.14
		LRSKDE	0.25	0.99	1.00
	0.4	LR	$*$	0.60	1.00
		NSBS	$*$	0.09	0.78
		NSBSKDE	$*$	0.07	0.75
		LRSKDE	$*$	0.54	1.00

5.2. Gaussian versus symmetric log-concave clustering

We now consider the problem of clustering, that is, of assigning to each observation in a dataset a label without being given any "training" labels. We will assume that the data can be clustered into two groups, which we will do by fitting the two-component mixture (2) and assigning a label to an observation X based on whether our estimate of the posterior probability

$$\frac{\pi^0 f^0(X - u_1^0)}{\pi^0 f^0(X - u_1^0) + (1-\pi^0) f(X - u_2^0)} \tag{12}$$

is greater than $1/2$ or not.

Table 2. Comparison of the four different clustering methods given by the column labels, see text for more details. The reported numbers are the average number of misclassifications out of $n = 500$ samples over $R = 5000$ replications under each of the three log-concave densities in the left column. Here $u_2^0 - u_1^0 = 1$ and $\pi^0 = 0.2$. The numbers in parentheses are the corresponding standard errors

	G	HG	SLC	KDE
$\mathcal{N}(0,1)$	163 (0.44)	170 (0.30)	170 (0.30)	182 (0.88)
$\mathcal{L}(1)$	223 (0.64)	173 (0.32)	174 (0.28)	154 (0.82)
$\mathcal{U}(-1,1)$	144 (0.15)	66 (0.11)	55 (0.10)	105 (0.28)

We fit the mixture three different ways. In the first basic approach, labeled "G," we maximize the likelihood (5) under the assumption that the component f is a normal density. We use the EM algorithm to maximize the likelihood. Our next two approaches both use the method of [26] to estimate the mixture components u_1^0, u_2^0, and π^0. Then we either fit the components using a *G*aussian density (denoted "HG"), with variance estimate also given by [26], or we use the symmetric log-concave density estimator (denoted "SLC") for the components. The fourth approach is based on the estimators of [26] and the kernel density estimator based on the inversion formula given in (9) by [6] where we truncate the infinite sum at some large integer $K > 0$. Precisely, let $\bar{g}_n$ be a standard density estimator of the mixed density g^0. Then, the KDE of f^0 we use is given by $\max(1/2(\bar{f}_n(x) + \bar{f}(-x)), 0)$ where

$$\bar{f}_n(x) = \sum_{k=0}^{K} \left(\frac{-\check{\pi}_n}{1 - \check{\pi}_n} \right)^k \bar{g}_n\big(x + \check{u}_{2,n} + k(\check{u}_{2,n} - \check{u}_{1,n})\big).$$

We should note that this formula is only valid, when $\check{\pi}_n < 1/2$. Hence, $1 - \check{\pi}_n$ and $(\check{u}_{2,n}, \check{u}_{1,n})$ should replace $\check{\pi}_n$ and $(\check{u}_{1,n}, \check{u}_{2,n})$ when $\check{\pi}_n \geq 1/2$.

We record the average missclassification count when the true density is one of the densities in the left column of Table 2. In all cases, the number of replications is $R = 5000$, the sample size is $n = 500$, $u_2^0 - u_1^0 = 1$ and $\pi^0 = 0.2$.

The performances of the four approaches were then compared, and the results are reported in Table 2. The KDE approach does clearly worse than the three other methods. The SLC outperforms HG by 16% when the true density is $\mathcal{U}[-1, 1]$. In the other cases, they perform similarly. All four methods define the two cluster regions by dividing the real line into two half-lines. The HG and SLC methods have the same mixture components so the shape of the component density estimates have to be dramatically different (e.g., uniform instead of normal) in order to noticeably change the results; note this somewhat deceiving outcome is not totally in contradiction with the finding of [14] about the performance of their two-dimensional log-concave classifier applied to the Breast cancer data of Wisconsin; see [14] for details. The authors found that the log-concave MLE reduces the percentage of misclassification from 10.36% obtained for the Gaussian estimate to only 8.43% for that particular data set. The posterior probabilities of cluster membership, which can be used as a measurement of uncertainty, can also differ noticeably between the HG method and our SLC method.

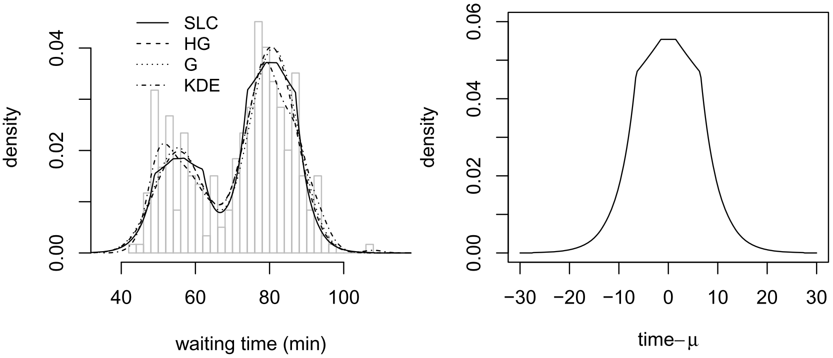

Figure 2. Time between eruptions of Old Faithful Geyser (min). The "SLC" and "HG" estimates both use the method of [26] to estimate the mixture parameters u_1^0, u_2^0, and π^0. "SLC" then fits with symmetric log-concave components and "HG" fits with Gaussian components. The "G" estimate is the maximum likelihood estimate of a mixture model with two Gaussian components with equal variances. The "KDE" is a standard kernel density estimator with an optimal bandwidth.

6. Data application

In this section, we apply our new estimation approach to two different datasets.

6.1. Old faithful data

The data to which we first apply our estimation procedure are the times, in minutes, between eruptions of the Old Faithful geyser in Yellowstone National park. There are many forms of the Old Faithful data. As far as we know, the oldest version of the data was collected by S. Weisberg from R. Hutchinson in August 1978. The data we analyze were collected between August 1 and August 15, 1985 continuously, and are from [1]. The following explanation from [40] motivates interest in the data:

> Old Faithful Geyser is an important tourist attraction, with up to several thousand people watching it erupt on pleasant summer days. The park service uses data like these to obtain a prediction equation for the time to the next eruption.

In Figure 2, we have two plots related to the Old Faithful data. The plot on the left depicts a descriptive histogram of the data with around 30 bins (which is too many for optimal estimation) along with the plots of four mixture density estimates. The "SLC" (*s*ymmetric *l*og-*c*oncave) estimate is the mixture model where u_1^0, u_2^0 and π^0 are estimated using the method of [26], and then the components are estimated using our symmetric log-concave estimator. The "HG" (*H*unter *et al.* and *G*aussian components) estimate is given by again using the method of [26] to find estimates of the mixture parameters whereas the nonparametric components are taken to be Gaussian components (the same Gaussian density for both mixture components). The estimates for u_1^0, u_2^0, and π^0 given by [26] are 55.5, 80.5, and 0.33, respectively. The "G" (*G*aussian) estimate in the plot is based on simply using a Gaussian mixture model with two components with equal variances. Assuming equal variances forces the two components to be identical, which makes the model analogous to the others. In this case, we estimated u_1^0, u_2^0, and π^0 by the EM algorithm [15], with estimated values of 55.3, 81.0, and 0.339. The normal components are slightly more peaked than the log-concave ones, but the overall fit is fairly similar; in large part this is because the locations and weights are very similar. Finally, the "KDE" is a standard kernel density estimator with an optimal bandwidth.

The plot on the right is that of the zero-symmetric log-concave component, centered at 0, used in the mixture density. As expected from the known theoretical properties of this estimator, it has a flat interval about the origin, and is the exponential of a concave piecewise linear and zero-symmetric function.

6.2. Height data

We next examine 1766 human height observations. We look at the heights of the population of Campora, a village in the south of Italy. This population is studied by the "Genetic Park of Cilento and Vallo di Dano Project" [11], which is interested in identifying geographically and genetically isolated populations. Such populations are of particular interest because in addition to "genetic homogeneity," they have a "uniformity of diet, life style and environment." These homogeneities are valuable in the study of genetic risk factors for complex pathologies such as "hypertension, diabetes, obesity, cancer, and neurodegenerative diseases," by allowing for a "simplification of the complexity of genetic models" involved, because of the population's homogeneity [11].

Colonna *et al.* [12] provide evidence that this population is indeed genetically isolated. Because of this feature, the distribution of heights of this population is not necessarily the same as that of the global population at large, so estimating its distribution is of interest. Height data are often modeled as mixtures of two components, corresponding to the two sexes.

We present plots related to the height data in Figure 3. The height data do not exhibit multimodality, but two-component mixtures still fit the data well. The three approaches that we consider fit similarly, but the log-concave components are able to capture a bit more asymmetry near to the mode.

The plot on the right includes the mixture component density (labeled "All"), in black. The data include the sex of each individual, so, using this extra information we can also estimate the true component densities separately: the zero-symmetric log-concave density estimate can be compared to the estimates of the density of the heights for either sex considered alone. Figure 4 shows the plots of the (descriptive) histograms of the heights for men and women and fitted standard kernel density estimators. The assumption of symmetry of the distribution of the heights for each of the genders seems to be reasonable to make. The observed proportion of women is 0.57, whereas the observed medians of the heights for women and men were found to be 156.0 and 168.7, respectively. Using the estimation method of [26], we found $\check{\pi}_n = 0.72$, $\check{u}_{1,n} = 157.5$, and $\check{u}_{2,n} = 170.5$. Here $\check{\pi}_n$ and $\check{u}_{1,n}$ correspond to the component for women. The components estimated by using the labels for men and women differ from that using the mixture model without the labels especially towards the center. We believe that this is essentially due to the difference between the estimates of the mixture parameters π^0, μ_1^0 and μ_2^0 obtained by ignoring or using the information available about the gender. In the latter case, the locations are estimated

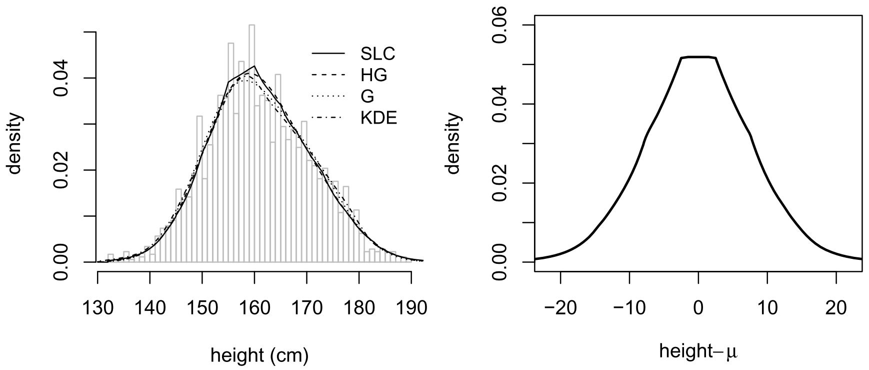

Figure 3. Height data of the population of Campora. The "SLC" and "HG" estimates both use the method of [26] to estimate the mixture parameters u_1^0, u_2^0, and π^0. "SLC" then fits with symmetric log-concave components and "HG" fits with Gaussian components. The "G" estimate is the maximum likelihood estimate of a mixture model with two Gaussian components with equal variances. The "KDE" is a standard kernel density estimator with an optimal bandwidth.

by the respective medians. It does appear that the distributions of heights of men and women are somewhat different near those centers, with women having a more peaked density and men having a flatter one. Thus, in the mixture model, without using the labels, the component density estimate is somewhere in between the two shapes.

7. Conclusions

The goal in this paper is to make use of the log-concavity constraint to estimate the unknown density component in a semi-parametric location mixture model assuming that this unknown density

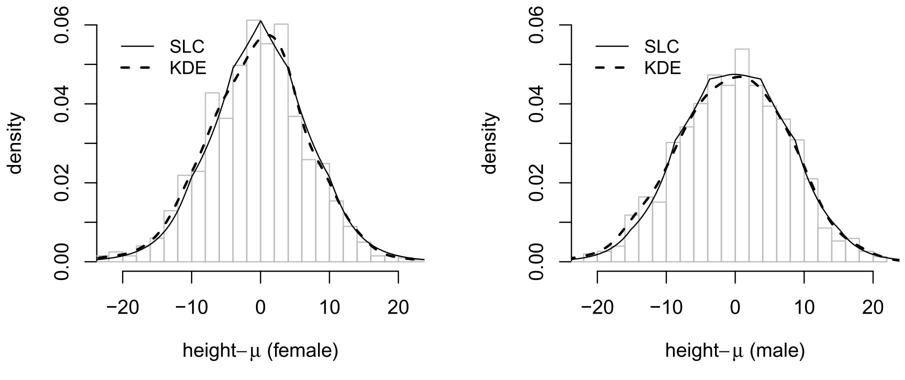

Figure 4. Descriptive histograms of the height data for women (left) and men (right) after centering around the median. The "SLC" is the log-concave MLE of the true density constrained to have mode at 0. The "KDE" is a standard kernel density estimator with an optimal bandwidth.

is symmetric around the origin. The first motivation for choosing this approach is that many densities are log-concave. The second one is to build an estimation procedure that does not depend on a tuning parameter. Our log-concave MLE is computed by maximizing the log-likelihood function after estimating the mixture parameters using the approach of [26]. The computation is easily implementable using the EM algorithm in combination with an active set algorithm already implemented in the R package `logcondens.mode`.

As already mentioned, our method is not advocated for heavy-tailed densities. In such cases, other shape constraints may be more appropriate, specifically, s-concavity, as studied in [28] and [16]. Unfortunately, the theory of estimators of s-concave densities is less developed than that of log-concave MLEs, which remains a barrier to using s-concavity in our current context.

Finally, [26] give sufficient conditions on the mixing probabilities and mixture locations for the model to be 3-identifiable. In this case, the mixture parameters can still be computed using the method of [26], and the log-concave MLE can be computed as described in this paper. However, it is not immediate in that case whether the same proof approaches would still yield the same rate of convergence. Recently, [2] proved that the number of components k, the mixture parameters and the unknown density are identifiable provided that the density is Pólya frequency (of infinite order) such that its expectation is equal to zero. For a precise definition of Pólya frequence functions, we refer to [34]. The obtained identifiability result can be used of course in the case of symmetry but it is certainly not a requirement. Note that imposing the log-concave constraint in this setting is natural since the class of Pólya frequency functions is a subset of the log-concave class as shown by [34]. One may argue that non-parametric classes such as symmetric densities or Pólya frequency functions with expectation equal to zero are not large enough. However, it seems that identifiability is hard to obtain if one allows for large classes.

Supplementary Material

Supplement to "Inference for a two-component mixture of symmetric distributions under log-concavity" (DOI: 10.3150/16-BEJ864SUPP; .pdf). In the supplement, we provide the proofs and other technical details that were omitted from the main paper.

References

[1] Azzalini, A. and Bowman, A.W. (1990). A look at some data on the Old Faithful geyser. *J. R. Stat. Soc., Ser. C* **39** 357–365.

[2] Balabdaoui, F. and Butucea, C. (2014). On location mixtures with Pólya frequency components. *Statist. Probab. Lett.* **95** 144–149. MR3262962

[3] Balabdaoui, F. and Doss, C.R. (2016). Supplement to "Inference for a two-component mixture of symmetric distributions under log-concavity." DOI:10.3150/16-BEJ864SUPP.

[4] Balabdaoui, F., Jankowski, H., Rufibach, K. and Pavlides, M. (2013). Asymptotics of the discrete log-concave maximum likelihood estimator and related applications. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **75** 769–790. MR3091658

[5] Balabdaoui, F., Rufibach, K. and Wellner, J.A. (2009). Limit distribution theory for maximum likelihood estimation of a log-concave density. *Ann. Statist.* **37** 1299–1331. MR2509075

[6] Bordes, L., Mottelet, S. and Vandekerkhove, P. (2006). Semiparametric estimation of a two-component mixture model. *Ann. Statist.* **34** 1204–1232. MR2278356

[7] Butucea, C. and Vandekerkhove, P. (2014). Semiparametric mixtures of symmetric distributions. *Scand. J. Stat.* **41** 227–239. MR3181141

[8] Chang, G.T. and Walther, G. (2007). Clustering with mixtures of log-concave distributions. *Comput. Statist. Data Anal.* **51** 6242–6251. MR2408591

[9] Chee, C.-S. and Wang, Y. (2013). Estimation of finite mixtures with symmetric components. *Stat. Comput.* **23** 233–249. MR3016941

[10] Chen, Y. and Samworth, R.J. (2013). Smoothed log-concave maximum likelihood estimation with applications. *Statist. Sinica* **23** 1373–1398. MR3114718

[11] Ciullo, M. (2009). Genetic park of Cilento and Vallo di Dano project. Available at http://www.igb.cnr.it/cilentoisolates.

[12] Colonna, V., Nutile, T., Astore, M., Guardiola, O., Antoniol, G., Ciullo, M. and Persico, M.G. (2007). Campora: A young genetic isolate in South Italy. *Hum. Hered.* **64** 123–135.

[13] Cule, M. and Samworth, R. (2010). Theoretical properties of the log-concave maximum likelihood estimator of a multidimensional density. *Electron. J. Stat.* **4** 254–270. MR2645484

[14] Cule, M., Samworth, R. and Stewart, M. (2010). Maximum likelihood estimation of a multi-dimensional log-concave density. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **72** 545–607. MR2758237

[15] Dempster, A.P., Laird, N.M. and Rubin, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Statist. Soc. Ser. B* **39** 1–38. MR0501537

[16] Doss, C.R. and Wellner, J.A. (2016). Global rates of convergence of the MLEs of log-concave and s-concave densities. *Ann. Statist.* **44** 954–981. MR3485950

[17] Dümbgen, L., Hüsler, A. and Rufibach, K. (2010). Active set and EM algorithms for log-concave densities based on complete and censored data. Technical report, Univ. Bern. Available at arXiv:0707.4643.

[18] Dümbgen, L. and Rufibach, K. (2009). Maximum likelihood estimation of a log-concave density and its distribution function: Basic properties and uniform consistency. *Bernoulli* **15** 40–68. MR2546798

[19] Dümbgen, L. and Rufibach, K. (2011). logcondens: Computations related to univariate log-concave density estimation. *J. Stat. Softw.* **39** 1–28.

[20] Dümbgen, L., Samworth, R. and Schuhmacher, D. (2011). Approximation by log-concave distributions, with applications to regression. *Ann. Statist.* **39** 702–730. MR2816336

[21] Dümbgen, L., Samworth, R. and Schuhmacher, D. (2011). Approximation by log-concave distributions, with applications to regression. Available at arXiv:1002.3448v3.

[22] Eilers, P.H.C. and Borgdorff, M.W. (2007). Non-parametric log-concave mixtures. *Comput. Statist. Data Anal.* **51** 5444–5451. MR2370883

[23] Everitt, B.S. and Hand, D.J. (1981). *Finite Mixture Distributions*. London: Chapman & Hall. MR0624267

[24] Fraley, C. and Raftery, A.E. (2002). Model-based clustering, discriminant analysis, and density estimation. *J. Amer. Statist. Assoc.* **97** 611–631. MR1951635

[25] Has'minskiĭ, R.Z. (1978). A lower bound for risks of nonparametric density estimates in the uniform metric. *Teor. Veroyatnost. i Primenen.* **23** 824–828. MR0516279

[26] Hunter, D.R., Wang, S. and Hettmansperger, T.P. (2007). Inference for mixtures of symmetric distributions. *Ann. Statist.* **35** 224–251. MR2332275

[27] Kim, A.K.H. and Samworth, R.J. (2014). Global rates of convergence in log-concave density estimation. Available at arXiv.org:1404.2298v1.

[28] Koenker, R. and Mizera, I. (2010). Quasi-concave density estimation. *Ann. Statist.* **38** 2998–3027. MR2722462

[29] McLachlan, G. and Peel, D. (2000). *Finite Mixture Models*. *Wiley Series in Probability and Statistics*: *Applied Probability and Statistics*. New York: Wiley. MR1789474

[30] Niculescu, C.P. and Persson, L.-E. (2006). *Convex Functions and Their Applications*: *A Contemporary Approach*. *CMS Books in Mathematics/Ouvrages de Mathématiques de la SMC*, 23. New York: Springer. MR2178902

[31] Pal, J.K., Woodroofe, M. and Meyer, M. (2007). Estimating a Polya frequency function$_2$. In *Complex Datasets and Inverse Problems*. *Institute of Mathematical Statistics Lecture Notes—Monograph Series* **54** 239–249. Beachwood, OH: IMS. MR2459192

[32] Rockafellar, R.T. (1970). *Convex Analysis*. *Princeton Mathematical Series*, *No*. 28. Princeton, NJ: Princeton Univ. Press. MR0274683

[33] Rufibach, K. (2007). Computing maximum likelihood estimators of a log-concave density function. *J*. *Stat*. *Comput*. *Simul*. **77** 561–574. MR2407642

[34] Schoenberg, I.J. (1951). On Pólya frequency functions. I. The totally positive functions and their Laplace transforms. *J*. *Anal*. *Math*. **1** 331–374. MR0047732

[35] Schuhmacher, D. and Dümbgen, L. (2010). Consistency of multivariate log-concave density estimators. *Statist*. *Probab*. *Lett*. **80** 376–380. MR2593576

[36] Silverman, B.W. (1982). On the estimation of a probability density function by the maximum penalized likelihood method. *Ann*. *Statist*. **10** 795–810. MR0663433

[37] Stone, C.J. (1980). Optimal rates of convergence for nonparametric estimators. *Ann*. *Statist*. **8** 1348–1360. MR0594650

[38] Titterington, D.M., Smith, A.F.M. and Makov, U.E. (1985). *Statistical Analysis of Finite Mixture Distributions*. *Wiley Series in Probability and Mathematical Statistics*: *Applied Probability and Statistics*. Chichester: Wiley. MR0838090

[39] Walther, G. (2002). Detecting the presence of mixing with multiscale maximum likelihood. *J*. *Amer*. *Statist*. *Assoc*. **97** 508–513. MR1941467

[40] Weisberg, S. (2005). *Applied Linear Regression*, 3rd ed. *Wiley Series in Probability and Statistics*. Wiley: Hoboken, NJ. MR2112740

Received November 2014 and revised May 2016