

Computation of an efficient and robust estimator in a semiparametric mixture model

Jingjing Wu, Weixin Yao & Sijia Xiang

To cite this article: Jingjing Wu, Weixin Yao & Sijia Xiang (2017) Computation of an efficient and robust estimator in a semiparametric mixture model, Journal of Statistical Computation and Simulation, 87:11, 2128-2137, DOI: [10.1080/00949655.2017.1318136](https://doi.org/10.1080/00949655.2017.1318136)

To link to this article: <https://doi.org/10.1080/00949655.2017.1318136>



Published online: 25 Apr 2017.



Submit your article to this journal [↗](#)



Article views: 72



View related articles [↗](#)



View Crossmark data [↗](#)



Computation of an efficient and robust estimator in a semiparametric mixture model

Jingjing Wu^a, Weixin Yao^b and Sijia Xiang^c

^aDepartment of Mathematics and Statistics, University of Calgary, Calgary, Alberta, Canada; ^bDepartment of Statistics, University of California, Riverside, CA, USA; ^cSchool of Data Sciences, Zhejiang University of Finance and Economics, Hangzhou, People's Republic of China

ABSTRACT

In this article, we propose an efficient and robust estimation for the semiparametric mixture model that is a mixture of unknown location-shifted symmetric distributions. Our estimation is derived by minimizing the profile Hellinger distance (MPHD) between the model and a nonparametric density estimate. We propose a simple and efficient algorithm to find the proposed MPHD estimation. Monte Carlo simulation study is conducted to examine the finite sample performance of the proposed procedure and to compare it with other existing methods. Based on our empirical studies, the newly proposed procedure works very competitively compared to the existing methods for normal component cases and much better for non-normal component cases. More importantly, the proposed procedure is robust when the data are contaminated with outlying observations. A real data application is also provided to illustrate the proposed estimation procedure.

ARTICLE HISTORY

Received 10 January 2017
Accepted 7 April 2017

KEYWORDS

Mixture models; semiparametric EM algorithm; semiparametric mixture models; minimum Hellinger distance

1. Introduction

The m -component mixture model has the density

$$h(x) = \sum_{j=1}^m \pi_j f_j(x), \quad x \in \mathbb{R}^k, \quad (1)$$

where f_j is the j th component density and (π_1, \dots, π_m) are the mixing proportions with $\sum_{j=1}^m \pi_j = 1$. When m is unknown, there are various articles that discuss the selection of m ; see, for example, Roeder [1], McLachlan and Peel [2], Chen et al. [3,4] and Chen and Li [5]. We assume throughout the article that m is fixed and known. When f_j belongs to a parametric family $\mathcal{F} = \{f(\cdot; \mu) : \mu \in \mathbb{R}^d\}$, i.e. $f_j(\cdot) = f(\cdot; \mu_j)$, model (1) has been well studied in the literature; see, for example, Lindsay [6], Böhning [7], McLachlan and Peel [2], and Frühwirth-Schnatter [8].

In practice, however, the choice of parametric family \mathcal{F} is difficult and it may affect the model estimation. Many authors have been trying to relax parametric assumptions on f_j . Note that model (1) is generally unidentifiable if no restrictions are placed on f_j , simply due to the fact that f_j alone could be another mixture of several densities. Bordes et al. [9] and Hunter et al. [10] considered a

location-shifted mixture model by assuming $f_j(\cdot) = f(\cdot - \mu_j)$, i.e.

$$h(x; \theta, f) = \sum_{j=1}^m \pi_j f(x - \mu_j), \tag{2}$$

where $\theta = (\pi_j, \mu_j, j = 1, \dots, m)$ and f is an unknown density function *symmetric* about zero. Model (2) is called semiparametric since it involves both unknown functional part f and unknown parametric part θ . Bordes et al. [9] proved the identifiability of model (2) for $m = 2$. Hunter et al. [10] established the identifiability of model (2) for both $m = 2$ and $m = 3$. Bordes et al. [11] provided an EM type algorithm to estimate model (2) and claimed that the new EM-type algorithm has better performance, according to their numerical studies, than the methods provided by Hunter et al. [10] and Bordes et al. [9]. Benaglia et al. [12] further improved the estimation procedure over Bordes et al. [11]. However, their estimates based on semiparametric EM algorithm (SPEM) do not fully use the model information in the M step and thus are not efficient.

In this article, we propose to estimate the semiparametric model (2) by minimizing the profile Hellinger distance between a nonparametric density estimation and assumed model (2). The resulted estimation is called minimum profile Hellinger distance (MPHD) estimation and was first introduced in [13]. It is well known that minimum Hellinger distance (MHD) estimator has been proved to be efficient and robust [6,14] for parametric models. MHD estimation for semiparametric models has been studied extensively in the past few years by many authors, including Wu and Karunamuni [13,15,16], Wu et al. [17], Tang and Karunamuni [18], Zhu et al. [19], and Xiang et al. [20], among others. The MPHD estimation introduced recently by Wu and Karunamuni [13] has the advantage, over above mentioned works, that it does not require a separate estimation of the infinite-dimensional nuisance parameter. It has been shown in [13] that the MPHD estimation is efficient, adaptive and robust against model misspecification. The robust and efficient properties of the MPHD estimation proposed in this paper, particularly for the mixture model (2), are also demonstrated by our Monte Carlo simulation studies. Based on our empirical study, the proposed MPHD estimation works very competitively with the existing methods for normal cases and much better for non-normal cases. More importantly, the proposed MPHD estimation is robust when the data are contaminated with outlying observations. A real data application is also provided to illustrate the proposed estimation procedure.

The remainder of this paper is organized as follows. In Section 2, we introduce the new MPHD estimation method for the semiparametric mixture model (2) and propose an effective algorithm to find the proposed estimation. In Section 3, we present both a Monte Carlo study and a real data example to compare the proposed MPHD estimation with those based on some other different methods. Finally, some discussions are given in Section 4.

2. New estimation method

In this section, we will develop the MPHD estimator for model (2). We first introduce the idea of MPHD estimators [13]. Then we apply this method to model (2).

2.1. Introduction of MPHD estimator

Let $(\mathcal{X}, \mathcal{S}, \nu)$ be a measure space and \mathcal{H} be a semiparametric model of ν -densities of the form

$$\mathcal{H} = \{h_{\theta, f} : \theta \in \Theta, f \in \mathcal{F}\}, \tag{3}$$

where Θ is a subset of \mathbb{R}^p and \mathcal{F} is an arbitrary set of infinite dimension. Let \mathcal{G} be a class of ν -densities that contains \mathcal{H} . For members a and b of $L_2(\nu)$ we write $\langle a, b \rangle$ for their inner products in $L_2(\nu)$ and

$\|a\|$ for the $L_2(\nu)$ -norm. The Hellinger distance between two members g_1 and g_2 of \mathcal{G} is given by

$$d_H(g_1, g_2) = \|g_1^{1/2} - g_2^{1/2}\| = \sqrt{2 - 2\langle g_1^{1/2}, g_2^{1/2} \rangle} = \sqrt{2\langle g_2^{1/2} - g_1^{1/2}, g_2^{1/2} \rangle}.$$

Assume that X_1, \dots, X_n are independent and identically distributed (i.i.d.) \mathcal{X} -valued random variables (r.v.s) with density $h_0 = h_{\theta_0, f_0}$ with θ_0 an interior point of Θ and $f_0 \in \mathcal{F}$. Suppose \hat{h}_n is a nonparametric \mathcal{G} -valued estimator of h_0 based on observed data X_1, \dots, X_n . Then, the MHD estimators of f_0 and θ_0 are

$$(\hat{f}, \hat{\theta}) = \arg \min_{f \in \mathcal{F}, \theta \in \Theta} \|h_{\theta, f}^{1/2} - \hat{h}_n^{1/2}\|. \tag{4}$$

Note that the arguments in Equation (4) contain both parametric part θ and nonparametric part f . In order to simplify the computation, Wu and Karunamuni [13] proposed the MPHD estimation method based on the profile idea.

Suppose model \mathcal{H} is identifiable in the sense that $\|h_{\theta_1, f_1}^{1/2} - h_{\theta_2, f_2}^{1/2}\| = 0$ implies $\theta_1 = \theta_2$ and $f_1 = f_2$. For each $\theta \in \Theta$, let

$$f_\theta(\hat{h}_n) = \arg \min_{f \in \mathcal{F}} \|h_{\theta, f}^{1/2} - \hat{h}_n^{1/2}\|. \tag{5}$$

Then, the profile Hellinger distance function $D(\theta)$ is defined as

$$D(\theta) = \inf_{f \in \mathcal{F}} \|h_{\theta, f}^{1/2} - \hat{h}_n^{1/2}\| = \|h_{\theta, f_\theta(\hat{h}_n)}^{1/2} - \hat{h}_n^{1/2}\|, \quad \theta \in \Theta. \tag{6}$$

The MPHD estimator of θ_0 proposed by Wu and Karunamuni [13] is

$$\hat{\theta} = \arg \min_{\theta \in \Theta} D(\theta).$$

Therefore, the profile method reduce the problem of minimizing the complicated objective function in Equation (4), which contains both parametric part θ and nonparametric part f , to the problem of minimizing the simple Hellinger distance function $D(\theta)$, which only contains the parametric part θ . Wu and Karunamuni [13] have proved the efficiency and robustness of MPHD estimators for semiparametric models (3).

2.2. Application of MPHD estimator to the semiparametric mixture model (2)

Next we apply the MPHD method to the location-shifted mixture model (2). For simplicity we consider the case that $m = 2$. The following result could be easily extended to the cases with $m > 2$. First, model (2) could be rewritten as

$$\mathcal{H} = \{h_{\theta, f}(x) = \pi f(x - \mu_1) + (1 - \pi)f(x - \mu_2) : \theta \in \Theta, f \in \mathcal{F}\}, \tag{7}$$

where

$$\Theta = \left\{ \theta = (\pi, \mu_1, \mu_2) : \pi \in \left(0, \frac{1}{2}\right) \cup \left(\frac{1}{2}, 1\right), \mu_1 < \mu_2, \mu_i \in \mathbb{R}, i = 1, 2 \right\},$$

$$\mathcal{F} = \left\{ f : f \geq 0, f(-x) = f(x), \int f(x) dx = 1 \right\}. \tag{8}$$

Note that in Equation (8), we assume $\mu_1 < \mu_2$, following Hunter et al. [10], to avoid the label switching issue [21,22]. By Theorem 2 of Hunter et al. [10], the mixture model \mathcal{H} in Equation (7) is

identifiable over Θ and \mathcal{F} given in Equation (8). Therefore, the idea of MPHD estimation method can be applied to semiparametric mixture model (7) to produce an efficient and robust estimate.

Note that in order to use MPHD estimator, we need to have a nonparametric estimate of $h_{\theta,f}$. Suppose we observe X_1, \dots, X_n from a population with density $h_{\theta,f} \in \mathcal{H}$. Let \hat{h}_n denote the nonparametric kernel density estimator of $h_{\theta,f}$, i.e.

$$\hat{h}_n(x) = \frac{1}{nb} \sum_{j=1}^n K\left(\frac{X_j - x}{b}\right),$$

where $K(\cdot)$ is a density function (called kernel) and b is the bandwidth. In this paper, we use Gaussian density for $K(\cdot)$, and $b = b(n) \rightarrow 0$ as $n \rightarrow \infty$. Then, following Wu and Karunamuni [13], we propose an MPHD estimator of as

$$\hat{\theta} = \arg \min_{\theta \in \Theta} \|h_{\theta,f}^{1/2} - \hat{h}_n^{1/2}\|, \tag{9}$$

where f_{θ} is defined in Equation (5).

2.3. Computation algorithm

Since the MPHD estimate $\hat{\theta}$ defined in Equation (9) does not have an explicit expression, we propose the following algorithm to calculate $\hat{\theta}$ iteratively. Suppose the initial estimates of $\theta = (\pi, \mu_1, \mu_2)$ and f are, respectively, $\theta^{(0)} = (\pi^{(0)}, \mu_1^{(0)}, \mu_2^{(0)})$ and $f^{(0)}$.

Step 1 For fixed $\pi^{(k)}, \mu_1^{(k)}, \mu_2^{(k)}$ and $f^{(k)}$, find $f^{(k+1)}$ which minimizes

$$\|[\pi^{(k)} f^{(k)}(\cdot - \mu_1^{(k)}) + (1 - \pi^{(k)}) f^{(k+1)}(\cdot - \mu_2^{(k)})]^{1/2} - \hat{h}_n^{1/2}(\cdot)\|.$$

It turns out [13] that the solution is

$$f^{(k+1)}(x - \mu_2^{(k)}) = \begin{cases} \frac{\alpha}{1 - \pi^{(k)}} \hat{h}_n(x) - \frac{\pi^{(k)}}{1 - \pi^{(k)}} f^{(k)}(x - \mu_1^{(k)}), & \text{if } x \in M, \\ 0, & \text{if } x \in M^C, \end{cases}$$

where

$$M = \{x : \alpha \hat{h}_n(x) \geq \pi^{(k)} f^{(k)}(x - \mu_1^{(k)})\},$$

$$\alpha = \sup_{0 < a \leq 1} \left\{ \pi^{(k)} \int_M f^{(k)}(x - \mu_1^{(k)}) dx + (1 - \pi^{(k)}) \geq a \int_M \hat{h}_n(x) dx \right\}.$$

If both $f^{(k)}$ and \hat{h}_n are continuous, then

$$\alpha = \frac{(1 - \pi^{(k)}) + \pi^{(k)} \int_M f^{(k)}(x - \mu_1^{(k)}) dx}{\int_M \hat{h}_n(x) dx}.$$

Since $f(\cdot)$ is symmetric about 0, i.e. $f(x) = f(-x)$, similar to Bordes et al. [11], we can symmetrize $f^{(k+1)}(x)$ by

$$\frac{f^{(k+1)}(x) + f^{(k+1)}(-x)}{2}.$$

Step 2. For fixed $f^{(k+1)}$, find $\pi^{(k+1)}$, $\mu_1^{(k+1)}$ and $\mu_2^{(k+1)}$ which minimize

$$\|[\pi^{(k+1)}f^{(k+1)}(\cdot - \mu_1^{(k+1)}) + (1 - \pi^{(k+1)})f^{(k+1)}(\cdot - \mu_2^{(k+1)})]^{1/2} - \hat{h}_n^{1/2}(\cdot)\|. \tag{10}$$

Then go to Step 1.

The minimization of Equation (10) requires numerical integration and optimization. To give the initial estimates $\theta^{(0)}$ and $f^{(0)}$, one may use the following way, similar to the idea in [10]. For fixed $\theta^{(0)}$, define

$$f^{(0)}(x) = \frac{u(x)I_{\{u(x) \geq 0\}}}{\int u(x)I_{\{u(x) \geq 0\}} dx},$$

where

$$u(x) = \pi^{(0)}[\hat{h}_n(\mu_1^{(0)} + x) + \hat{h}_n(\mu_1^{(0)} - x)] - (1 - \pi^{(0)})[\hat{h}_n(\mu_2^{(0)} + x) + \hat{h}_n(\mu_2^{(0)} - x)].$$

The estimator $f^{(0)}$ of f has appealing properties such as it is symmetric about zero, and $f^{(0)} = f$ when $\hat{h}_n = \pi^{(0)}f(x - \mu_1^{(0)}) + (1 - \pi^{(0)})f(x - \mu_2^{(0)})$. Now $\theta^{(0)}$ is the value which minimizes

$$\|[\pi^{(0)}f^{(0)}(\cdot - \mu_1^{(0)}) + (1 - \pi^{(0)})f^{(0)}(\cdot - \mu_2^{(0)})]^{1/2} - \hat{h}_n^{1/2}(\cdot)\|.$$

3. Simulation studies and real data application

In this section, we use a Monte Carlo simulation study to illustrate the finite sample performance of our proposed MPHED estimator and compare it with some existing methods. In addition, we illustrate the newly proposed procedure with an empirical analysis of a real data example.

Example 3.1: This example is designed to assess the finite sample performance and robustness of the proposed estimator and compare it with some existing methods. We generate independent and identically distributed data $\{x_1, \dots, x_n\}$ from a population with density function

$$h(x) = \pi_1 f(x - \mu_1) + (1 - \pi_1) f(x - \mu_2), \tag{11}$$

where (π_1, μ_1, μ_2) are unknown parameters and f is an unknown density that is symmetric about zero. We consider the following four cases:

Case I: $X \sim 0.25N(-1, 1) + 0.75N(2, 1) \Rightarrow f(x) \sim N(0, 1), \mu_1 = -1, \mu_2 = 2, \pi_1 = 0.25$.

Case II: $X \sim 0.3U(-1, 1) + 0.7U(0, 2) \Rightarrow f(x) \sim U(-1, 1), \mu_1 = 0, \mu_2 = 1, \pi_1 = 0.3$.

Case III: $f(x) \sim t_5, \mu_1 = 0, \mu_2 = 3, \pi_1 = 0.3$.

Case IV: $X \sim 0.25N(-1, 1) + 0.75N(2, 1)$ contaminated with 2% of outliers from $U(10, 20) \Rightarrow f(x) \sim N(0, 1), \mu_1 = -1, \mu_2 = 2, \pi_1 = 0.25$.

We use Case I to test the efficiency of our robust estimation compared to the traditional MLE when the error is normally distributed. Case II is also the model used by Bordes et al. [9] and Benaglia et al. [12] to show the performance of their SPEM. Case II and Case III are used to demonstrate that the proposed new method can be adaptive to the non-normal component densities. Case IV is the same as Case I except for adding 2% outliers. Therefore, Case IV can be used to compare the robustness of different estimation methods.

To estimate the unknown parameters in model (11), we consider the following three methods: (a) MLE assuming normal assumption with equal variance (MLE); (b) SPEM proposed by Benaglia et al. [12]; (c) the proposed MPHED estimator. For MLE, we use 20 random initial values and select the converged value which has the largest likelihood. For SPEM, we use the MLE as the initial value (note

that for SPEM, there is no objective function and thus it is difficult to choose the right root if multiple initial values are used). For MPHD, we also use 20 random initial values and select the converged value which minimizes the corresponding objective function.

To assess the performance, we look at both the mean and the root-mean-squared error (RMSE) of each estimate:

$$\text{mean}(\hat{\theta}) = \bar{\theta} =: \frac{1}{N} \sum_{t=1}^N \hat{\theta}_t,$$

$$\text{RMSE}(\hat{\theta}) = \sqrt{\frac{1}{N} \sum_{t=1}^N (\hat{\theta}_t - \bar{\theta})^2},$$

where N is the number of repetitions and $\hat{\theta}_t$ is the estimate based on the t th replicate. Here we take $N = 200$, and $\hat{\theta}$ is either MLE, SPEM, or MPHD estimate of $\theta = (\pi_1, \mu_1, \mu_2)$.

For the four different cases, tables 1–3 report the mean and the RMSE of the parameter estimates, based on the three methods under consideration, when $n = 30, 100$ and 300 , respectively. Figure 1 shows the boxplots of RASE of Cases I–IV with $n = 100$. From the tables and figure, we have the following findings.

1. The MPHD estimator has overall better performance than the MLE and SPEM. The proposed MPHD estimate provides similar results to the MLE and SPEM estimate for normal component density. However, MPHD shows much better performance than the MLE and SPEM estimate for uniform and t_5 component densities. In addition, based on Case IV, we can see that the MPHD

Table 1. Mean (RMSE) of point estimates for $\theta = (\pi_1, \mu_1, \mu_2)$ when $n = 30$ for Example 3.1.

Component density	True	MLE	SPEM	MPHD
I: Norm	$\mu_1 : -1$	−0.998(0.582)	−0.630(1.115)	−0.841(0.641)
	$\mu_2 : 2$	2.042(0.379)	1.806(0.539)	1.910(0.335)
	$\pi_1 : 0.25$	0.272(0.121)	0.232(0.134)	0.255(0.134)
II: Uniform	$\mu_1 : 0$	−0.005(0.277)	0.313(0.518)	0.055(0.279)
	$\mu_1 : 1$	1.210(0.299)	0.896(0.317)	1.016(0.187)
	$\pi_1 : 0.3$	0.433(0.228)	0.311(0.164)	0.351(0.192)
III: t_5	$\mu_1 : 0$	−0.352(1.494)	0.061(1.865)	−0.008(1.353)
	$\mu_2 : 3$	3.219(1.046)	2.832(1.130)	2.932(0.579)
	$\pi_1 : 0.3$	0.326(0.199)	0.267(0.207)	0.311(0.196)
IV: Norm+Outliers	$\mu_1 : -1$	1.244(2.265)	1.135(2.341)	−0.686(0.872)
	$\mu_2 : 2$	15.02(13.37)	1.758(1.679)	1.758(0.553)
	$\pi_1 : 0.25$	0.968(0.718)	0.340(0.235)	0.247(0.195)

Table 2. Mean (RMSE) of point estimates for $\theta = (\pi_1, \mu_1, \mu_2)$ when $n = 100$ for Example 1.

Component density	True	MLE	SPEM	MPHD
I: Norm	$\mu_1 : -1$	−1.030(0.318)	−0.957(0.601)	−0.968(0.315)
	$\mu_2 : 2$	2.010(0.146)	1.864(0.270)	2.013(0.161)
	$\pi_1 : 0.25$	0.250(0.057)	0.211(0.088)	0.250(0.060)
II: Uniform	$\mu_1 : 0$	0.037(0.181)	0.382(0.528)	0.019(0.170)
	$\mu_1 : 1$	1.126(0.205)	0.790(0.270)	0.978(0.153)
	$\pi_1 : 0.3$	0.401(0.170)	0.276(0.160)	0.301(0.130)
III: t_5	$\mu_1 : 0$	−0.331(2.330)	0.490(1.320)	0.032(0.605)
	$\mu_2 : 3$	3.574(2.080)	2.624(0.627)	2.964(0.361)
	$\pi_1 : 0.3$	0.345(0.230)	0.246(0.170)	0.299(0.130)
IV: Norm+Outliers	$\mu_1 : -1$	1.262(2.268)	1.112(2.310)	−0.982(0.468)
	$\mu_2 : 2$	15.04(13.20)	1.662(0.462)	1.947(0.232)
	$\pi_1 : 0.25$	0.980(0.730)	0.318(0.158)	0.237(0.105)

Table 3. Mean(RMSE) of Point Estimates for $\theta = (\pi_1, \mu_1, \mu_2)$ when $n = 300$ for Example 1.

Component density	True	MLE	SPEM	MPHD
I: Norm	$\mu_1 : -1$	-0.985(0.165)	-0.993(0.169)	-0.938(0.169)
	$\mu_2 : 2$	2.022(0.085)	1.962(0.104)	2.034(0.093)
	$\pi_1 : 0.25$	0.252(0.033)	0.236(0.039)	0.254(0.034)
II: Uniform	$\mu_1 : 0$	0.083(0.139)	0.196(0.295)	0.003(0.072)
	$\mu_1 : 1$	1.126(0.163)	0.882(0.154)	1.002(0.048)
	$\pi_1 : 0.3$	0.414(0.150)	0.280(0.077)	0.299(0.051)
III: t_5	$\mu_1 : 0$	-0.446(2.030)	0.207(0.571)	0.044(0.241)
	$\mu_2 : 3$	3.396(2.150)	2.780(0.359)	2.995(0.130)
	$\pi_1 : 0.3$	0.309(0.190)	0.259(0.096)	0.301(0.061)
IV: Norm+Outliers	$\mu_1 : -1$	1.257(2.259)	1.466(2.487)	-0.970(0.165)
	$\mu_2 : 2$	15.05(13.10)	1.558(0.470)	2.004(0.102)
	$\pi_1 : 0.25$	0.980(0.730)	0.509(0.276)	0.247(0.032)

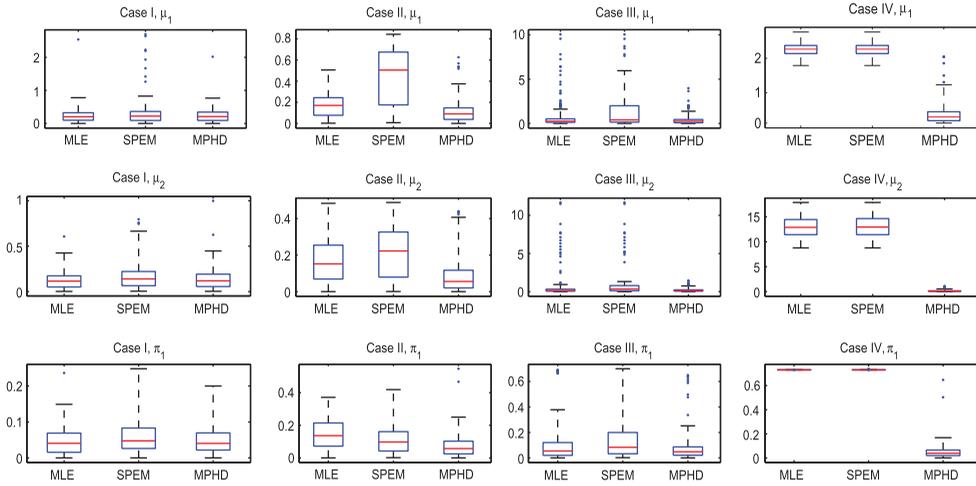


Figure 1. Boxplot of RASE of Cases I-IV with $n = 100$.

estimator is resistant to outliers while the MLE and SPEM are very sensitive to the outliers and fail to provide reasonable estimates when outliers are present.

2. The MLE, assuming normal components, works the best when the component density is exactly normal but does not work well for uniform and t_5 component densities.
3. The SPEM estimate provides worse results than the MLE for normal component density, especially for small sample size, but better results for uniform and t_5 component densities, especially for larger sample size.

Example 3.2: In this example, we illustrate the proposed methodology with an empirical analysis of the Old Faithful geyser data. The Old Faithful geyser data, which is available in the standard R distribution, records the waiting time between eruptions and the duration of the eruption for the Old Faithful geyser in the Yellowstone National Park, Wyoming, USA. The time units are minutes. The data set contains 272 observations on 2 variables. The waiting time variable has been used by Bordes et al. [9], Hunter et al. [10] and Bordes et al. [11] as a benchmark for the location shifted mixture model (2). The histogram of the data is shown in Figure 2(a). From the plot, we can see that the waiting time variable can be approximated well by a two-component normal mixture model. Table 4 reports the parameter estimates based on different methods. It can be seen that our proposed MPHD, along with SPEM, provide very close results to the parametric MLE, but without any parametric assumption

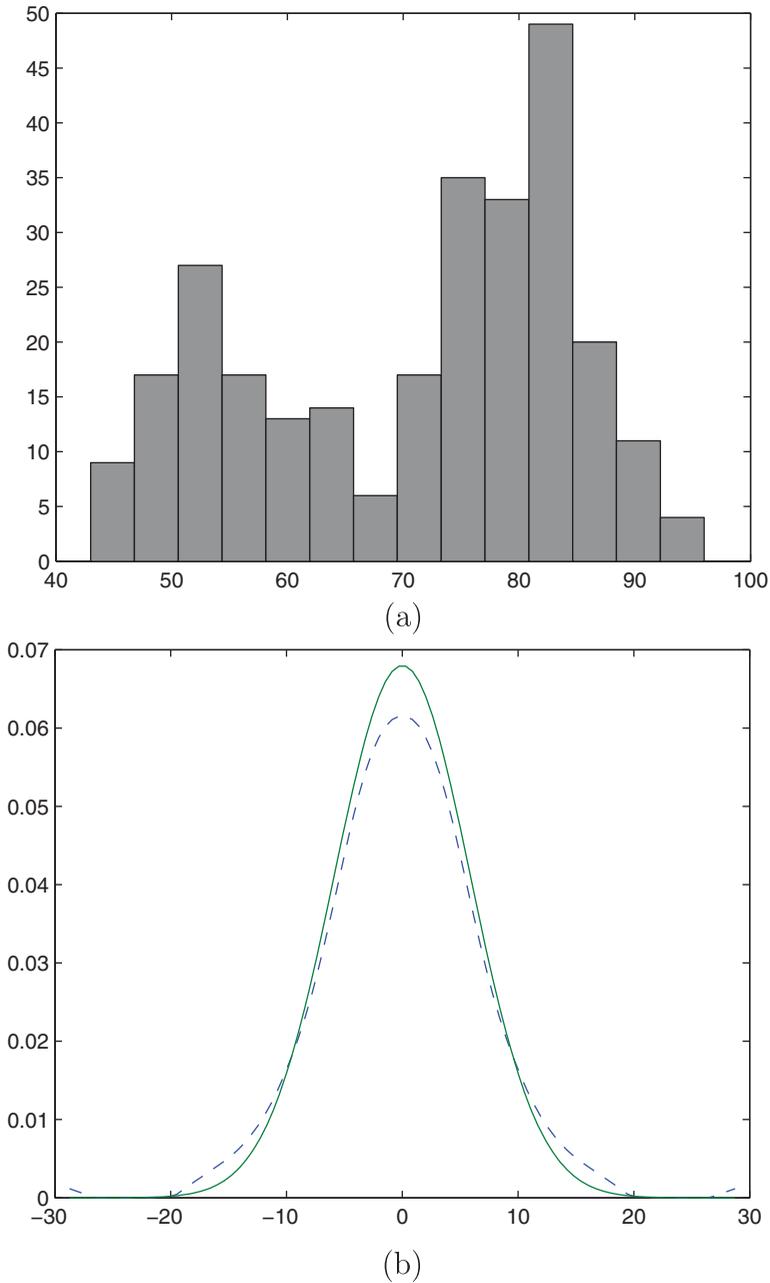


Figure 2. (a) Histogram of the Old Faithful geyser waiting time data; (b) The estimated density plots of f for MLE (solid) and MPHD (dashed).

about the component density. Figure 2(b) shows the estimated densities for f based on MLE and MPHD. From the plot, we can see that the component density is close to normal.

To check the robustness of different methods, we add five identical outliers ‘0’ to the original data set (the range of original data is from 43 to 96). The results are reported in Table 5, from which we can see that MPHD provides almost the same estimates as the ones without outliers in Table 4 and thus is robust and resistant to outliers. Comparatively, MLE considers the five outliers as one component

Table 4. Parameter estimates for the Old Faithful geyser waiting data based on different estimation methods.

Method	π_1	μ_1	μ_2
MLE	0.361	54.61	80.09
SPEM	0.360	54.61	80.06
MPHD	0.352	54.81	79.97

Table 5. Parameter estimates for the Old Faithful geyser waiting data **with added five outliers '0'** based on different estimation methods.

Method	π_1	μ_1	μ_2
MLE	0.018	0.061	70.90
SPEM	0.236	69.61	69.62
SPEM(MPHD)	0.338	51.04	79.11
MPHD	0.355	54.75	80.29

and the rest of the data as the other component. Therefore, the outliers have a big impact on MLE. The estimate provided by SPEM, using MLE as initial, is essentially a one component mixture model. For comparison, we also add the SPEM using MPHD as the initial, denoted by SPEM(MPHD). From Table 5, we can see that the SPEM(MPHD) provides much more reasonable estimate than SPEM, although the outliers still create some bias for the first component mean of SPEM(MPHD). Therefore, *a good initial value for SPEM is crucial.*

4. Concluding remark

We have developed an estimation procedure for a class of semiparametric mixture models, whose components are unknown location-shifted symmetric densities. The proposed estimation procedure is based on MHD. Using some Monte Carlo studies and data analysis, we have demonstrated that the new MHD-based estimation for this semiparametric model inherits the desired robustness and efficiency properties of traditional parametric MHD estimator. Our simulation studies further confirm this conclusion.

Though in this paper we focus on the mixture of location shifted models, the proposed methodology can be easily adapted to other mixture models, such as mixture of linear regression models and mixture of Poisson regression models. Those circumstances are of our great interest to further the investigation of MHD inference.

Acknowledgments

The authors thank the editor, the associate editor, and the reviewers for their constructive comments that have led to dramatic improvement of the earlier version of this article.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

Wu's research is supported by a grant from the Natural Sciences and Engineering Research Council of Canada [grant no. RGPIN 355970-2013]. Yao's research was supported by NSF [grant no. DMS-1461677] and Department of Energy Award No: 10006272. Xiang's research was supported by Zhejiang Provincial NSF of China [grant no. LQ16A010002] and NSF of China [grant no. 11601477].

References

- [1] Roeder K. A graphical technique for determining the number of components in a mixture of normals. *Journal of the Amer Statist Assoc.* 1994;89:487–495.
- [2] McLachlan GJ, Peel D. *Finite mixture models*. New York: Wiley; 2000.
- [3] Chen H, Chen J, Kalbfleisch JD. A modified likelihood ratio test for homogeneity in finite mixture models. *J R Statist Soc, Ser B.* 2001;63:19–29.
- [4] Chen H, Chen J, Kalbfleisch JD. Testing for a finite mixture model with two components. *J R Statist Soc, Ser B.* 2004;66:95–115.
- [5] Chen J, Li P. Hypothesis test for normal mixture models: The EM approach. *Ann Stat.* 2009;37:2523–2542.
- [6] Lindsay BG. *Mixture models: Theory, geometry and applications*. Hayward, CA: Institute of Mathematical Statistics; 1995. (NSF-CBMS Regional Conference Series in Probability and Statistics; Vol. 5).
- [7] Böhning D. *Computer-assisted analysis of mixtures and applications*. Boca Raton, FL: Chapman and Hall/CRC; 1999.
- [8] Frühwirth-Schnatter S. *Finite mixture and Markov switching models*. New York: Springer; 2006.
- [9] Bordes L, Mottelet S, Vandekerkhove P. Semiparametric estimation of a two-component mixture model. *Ann Stat.* 2006;34:1204–1232.
- [10] Hunter D, Wang S, Hettmansperger TP. Inference for mixtures of symmetric distributions. *Ann Stat.* 2007;35:224–251.
- [11] Bordes L, Chauveau D, Vandekerkhove P. A stochastic EM algorithm for a semiparametric mixture model. *Comput Stat Data Anal.* 2007;51:5429–5443.
- [12] Benaglia T, Chauveau D, Hunter DR. An EM-like algorithm for semi- and non-parametric estimation in multivariate mixtures. *J Comput Graph Stat.* 2009;18:505–526.
- [13] Wu J, Karunamuni RJ. Profile Hellinger distance estimation. *Statistics.* 2015;49:711–740.
- [14] Beran R. Minimum Hellinger distance estimates for parametric models. *Ann Stat.* 1977;5:445–463.
- [15] Wu J, Karunamuni RJ. On minimum Hellinger distance estimation. *Canad J Stat.* 2009;37:514–533.
- [16] Wu J, Karunamuni RJ. Efficient Hellinger distance estimates for semiparametric models. *J Multivariate Anal.* 2012;107:1–23.
- [17] Wu J, Karunamuni RJ, Zhang B. Minimum Hellinger distance estimation in a two-sample semiparametric model. *J Multivariate Anal.* 2010;101:1102–1122.
- [18] Tang Q, Karunamuni RJ. Minimum distance estimation in a finite mixture regression model. *J Multivariate Anal.* 2013;120:185–204.
- [19] Zhu Y, Wu J, Lu X. Minimum Hellinger distance estimation for a two-sample semiparametric cure rate model with censored survival data. *Comput Stat.* 2013;28:2495–2518.
- [20] Xiang S, Yao W, Wu J. Minimum profile Hellinger distance estimation for a semiparametric mixture model. *Canad J Stat.* 2014;42:246–267.
- [21] Stephens M. Dealing with label switching in mixture models. *J R Statist Soc, Ser B.* 2000;62:795–809.
- [22] Yao W, Lindsay BG. Bayesian mixture labeling by highest posterior density. *J Amer Statist Assoc.* 2009;104:758–767.