University of Calgary

PRISM Repository **https://prism.ucalgary.ca**

The Vault Open Theses and Dissertations

2017

Minimum Hellinger Distance Estimation for a Two-component Mixture Model

Zhou, Xiaofan

Zhou, X. (2017). Minimum Hellinger Distance Estimation for a Two-component Mixture Model (Master's thesis, University of Calgary, Calgary, Canada). Retrieved from https://prism.ucalgary.ca. doi:10.11575/PRISM/26922

http://hdl.handle.net/11023/3992

Downloaded from PRISM Repository, University of Calgary

UNIVERSITY OF CALGARY

Minimum Hellinger Distance Estimation for a Two-Component Mixture Model

by

Xiaofan Zhou

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE

DEGREE OF MASTER OF SCIENCE

DEPARTMENT OF MATHEMATICS AND STATISTICS

CALGARY, ALBERTA

July, 2017

© Xiaofan Zhou 2017

Abstract

Over the last two decades, semiparametric mixture model receives increasing attention, simply due to the fact that mixture models arise frequently in real life. In this thesis we consider a semiparametric two-component location-shifted mixture model. We propose to use the minimum Hellinger distance estimator (MHDE) to estimate the two location parameters and the mixing proportion. A MHDE is obtained by minimizing the Hellinger distance between an assumed parametric model and a nonparametric estimation of the model. MHDE was proved to have asymptotic efficiency and excellent robustness against small deviations from assumed model. To construct the MHDE, we propose to use a bounded linear operator introduced by Bordes et al. (2006) to estimate the unknown nuisance parameter (an unknown function). To facilitate the calculation of the MHDE, we develop an iterative algorithm and propose a novel initial estimation of the parameters of our interest. To assess the performance of the proposed estimations, we carry out simulation studies and a real data analysis and compare the results with those of the minimum profile Hellinger distance estimator (MPHDE) proposed by Wu et al. (2017) for the same model. The results show that the MHDE is very competitive with the MPHDE in terms of bias and mean squared error, while the MHDE is on average about 2.7 times computationally faster than the MPHDE. The simulation studies also demonstrate that the proposed initial is much more robust than the one used in Wu et al. (2017).

Acknowledgements

First I would like to express my sincere appreciation to my supervisor Dr. Jingjing Wu. When I got stuck in my research, she always helped me out with her extensive knowledge and extreme patience. Besides guidance for my academic development, Dr. Wu also gave me many advices for my life in Canada, which is very helpful and heartwarming for me as a student who studies abroad for the first time.

I would like to thank my committee members Dr. Haocheng Li and Dr. Shawn Liu for their careful reading of my thesis and providing me with constructive and thoughtful advices and comments.

I also want to thank Dr. Gemai Chen and Dr. Xuewen Lu. Whenever I went to their office with my questions, they spared no effort to help me.

I am also grateful to the financial support from the Department of Mathematics and Statistics. Without it, this work can not happen.

Finally, many thanks go to my friends Heng Zhu, Ji Ruan, Wenyan Zhong, Qicheng Zhang and Longlong Huang for their support to my research.

Table of Contents

List of Tables

List of Figures and Illustrations

Chapter 1

INTRODUCTION

In this chapter, we introduce the model under our consideration and give a brief description of the proposed research in this thesis. In Sections 1.1 we review the mixture model and particularly the semiparametric location-shifted mixture model. In Section 1.2 we present the two-component location-shifted mixture model that will be discussed throughout the thesis. In Section 1.3, we describe the proposed research in this thesis and give the structure of the thesis.

1.1 Mixture model

The probability density function (p.d.f.) of m-component mixture model is defined as

$$h(x) = \sum_{j=1}^{m} \pi_j f_j(x) \tag{1.1}$$

where π_j's are the unknown mixture proportions with $\sum_{j=1}^{m} \pi_j = 1$ and f_j's are the unknown p.d.f.s of the components. It's common to assume that f_j's belong to a parametric family $\mathcal{F} = \{f(\cdot|\xi), \xi \in \mathbb{R}^d\}$ so that $h(x)$ becomes

$$h_{\theta,f}(x) = \sum_{j=1}^{m} \pi_j f(x|\xi_j), \tag{1.2}$$

where $\theta = (\pi_1 \dots \pi_m, \xi_1 \dots \xi_m)$ is the unknown parameter vector subject to constraint $\sum_{j=1}^{m} \pi_j = 1$ and $f(\cdot|\xi_j)$ is specified up to the unknown parameter ξ_j.

When m is known, model (1.2) has been studied extensively in literatures, such as Fr$\ddot{u}$hwirth-Schnatter (2006), Lindsay (1995), McLachlan and Basford (1988) and McLachlan and Peel (2000). Model (1.2) can be made more flexible with m unspecified. Many articles discussed the estimation of m, such as Dacunha-Castelle and Gassiat (1999), Keribin (2000), Roeder (1994) and Saraiva, Louzada and Milan(2014). Model (1.2) can be applied to many fields such as unsupervised pattern recognition, speech recognition and medical imaging. A well known example of its application is given in Hosmer (1973) where the author used a two-component mixture model to estimate the proportion of male halibuts with use of observations on length of halibuts.

Since the parametric assumption of model (1.2) can lead to model misspecification while model (1.1) is nonparametrically unidentifiable, many researchers started to pay more attention on semiparametric models. Vaart (1996) considered MLE in several examples of semiparametric mixture models including the exponential frailty model and the errors-in-variables model. Zou, Fine and Yandell (2002) studied the partial likelihood for a two-component semiparametric mixture model in which the component densities are related by an exponential tilt but are unspecified otherwise. Bordes and Vandekerkhove (2010) studied the semiparametric estimation of a two-component mixture model where one component is known. Xiang (2014) proposed a class of semiparametric mixture of regression models where the mixing proportions and variances are constants but the component regression functions are nonparametric functions of a covariate. In this thesis, we will focus on a semiparametric two-component location-shifted mixture model.

Semiparametric location-shifted mixture model is defined as

$$h(x) = \sum_{j=1}^{m} \pi_j f(x - \mu_j), \tag{1.3}$$

where m is known and fixed and the density f is symmetric about 0. We call model (1.3) a semiparametric model because we know nothing about f except its symmetry. Bordes et al. (2006) established the identifiability of (1.3) for $m = 2$. Hunter, Wang and Hettmansperger (2007) established the identifiability of (1.3) for $m = 2$ and $m = 3$ and proposed a generalized Hodges-Lehmann estimator for the unknown parameters. Bordes et al. (2007) used a stochastic EM algorithm for (1.3). Benaglia, Chauveau and Hunter (2009) eliminated the stochasticity of the algorithm of Bordes et al. (2007) and proposed a more flexible and easily applicable algorithm which can be extended to any number of mixture components. That's to say, we can use this algorithm when identifiability is difficult to establish conclusively. Seo(2017) proposed an estimation method using the doubly smoothed maximum likelihood and this method can effectively eliminate potential biases while maintaining a high efficiency.

1.2 Two-component location-shifted mixture model

In this thesis we consider model (1.3) with $m = 2$, i.e. the following semiparametric two-component location-shifted mixture model:

$$h_{\theta,f}(x) = h(x) = \pi f(x-\mu_1) + (1-\pi) f(x-\mu_2), \tag{1.4}$$

where $\theta = (\pi, \mu_1, \mu_2)$ is the unknown parameter vector and f is the unknown p.d.f. symmetric about 0. Here θ is the parameter of our interest and we treat f as a nuisance parameter which we are not interested in but will bring difficulty to estimating θ. According to Bordes et al. (2007), for identifiability we assume $\theta \in \Theta$ with parameter space $\Theta = \{(\pi, \mu_1, \mu_2) : \pi \in (0, 0.5) \cup (0.5, 1), \mu_1 < \mu_2, \mu_1, \mu_2 \in \mathbb{R}\}$.

Many attempts have been done to estimate the unknown parameter θ in (1.4). Bordes et al. (2006) presented a distance-based estimator based on a linear bounded operator and a contrast function proposed in the paper. Butucea and Vandekerkhove (2013) proposed a class of M-estimators based on a Fourier approach. Wu et al. (2017) proposed to use the minimum profile Hellinger distance (MPHD) to estimate θ in (1.4). The resulted MPHD estimator is very competitive when the components are normal and much better when the components are non-normal, compared to existing methods such as the SPEM proposed by Benaglia, Chauveau and Hunter (2009). Chee and Wang (2013)presented a semiparametic MLE while Kottas and Fillingham (2012) proposed a Baysian approach to estimate model (1.4).

1.3 Proposed research

In this thesis we propose a new procedure to estimate θ in model (1.4). Wu and Karunamuni (2012) demonstrated that the minimum Hellinger distance estimation (MHDE) for semiparametric models is efficient at the assumed model and simultaneously is robust when the true model deviates from the assumed model (for example when outlying observations are present). In the construction of MHDE, an appropriate estimate of the unknown nuisance parameter f is required which may based on the current sample from the mixture or other resource. Bordes et al. (2006) provided a way to estimate f based on the current sample from the mixture. They derived the inversion formula of f from the mixture $h_{\theta,f}$ from which f could be recovered. Thus, we can estimate f by applying this formula, or essentially a bounded linear operator, to a nonparametric estimate of $h_{\theta,f}$. The proposed estimation of θ in (1.4) will combine the two methods in Wu and Karunamuni (2012) and Bordes et al. (2006).

This thesis is organized in the following way. In Chapter 2, we propose a MHDE of θ in model (1.4) and present an algorithm for numerical calculation of the estimator. In Chapter 3, we evaluate the performance of the proposed estimator through simulation studies and make comparison with the MPHD estimator in Wu et al. (2017). We also apply the proposed estimation to analyze the Old Faithful Geyser data. Finally the conclusions and discussions are presented in Chapter 4.

Chapter 2

MHDE OF LOCATION-SHIFTED MIXTURE MODEL

In this chapter, we propose a MHDE for the two-component location-shifted mixture model (1.4). In Section 2.1, we review the MHDE for semiparametric models. Since the considered MHDE requires an appropriate estimate of the nuisance parameter, i.e. f for model (1.4), in Section 2.2 we use an inversion operator to obtain an estimation of f. Based on this estimation, we propose in Section 2.3 the MHDE of the parameter θ in (1.4) and develop an iterative algorithm to calculate it. Finally in Section 2.4, we present in detail a novel way to choose initial estimates for the algorithm.

2.1 Review of MHDE

For parametric model of general form, Beran (1977) introduced the very first time the MHDE based on Hellinger distance. Suppose we observe independent and identically distributed (i.i.d.) random variables (r.v.s) $X_1, \ldots, X_n$ with density function a member of the parametric model family $\{f_\theta : \theta \in \Theta\}$, where Θ is a compact subset of $\mathbb{R}^p$. The Hellinger distance between two density functions is defined as the L_2-distance between the respective root densities. The MHDE $\hat{\theta}$ of the unknown parameter θ is defined as

$$\hat{\theta} = \min_{t \in \Theta} \|f_t^{1/2} - \hat{f}^{1/2}\|, \tag{2.1}$$

where $\|\cdot\|$ denotes the L_2-norm, and $\hat{f}$ is an appropriate nonparametric (e.g. kernel) density estimator based on the sample X_i's. Beran (1977) proved that the MHDE defined in (2.1) is asymptotically efficient under assumed parametric model and simultaneously is minimax robust against small deviations from assumed parametric model.

MHDE has been studied for various parametric models in the literature. For example, Simpson (1987) examined MHDE for discrete data models. Yang (1991) and Ying (1992) studied MHDE for censored data. Sriram and Vidyashankar (2000) and Woo and Sriram (2006, 2007) investigated MHDE for branching processes and the mixture complexity in a finite mixture model, respectively. MHDE for finite mixture models and their variants were studied in Woodward et al. (1995), Cutler and Cordero-Brana (1996), Karlis and Xekalaki (1998), Lu et al. (2003) and Xiang et al. (2008). Takada (2009), N'drin and Hili (2013) and Prause et al. (2016) studied MHDE in respective stochastic volatility model, one-dimensional diffusion process and bivariate time series. Karunamuni and Wu (2011) proposed a one-step MHDE to overcome computational drawbacks.

MHDE for semiparametric models hasn't been fully investigated until recently. Wu and Karunamuni (2009, 2012) extended the MHDE from parametric model to semiparametric model of general form and the resulted MHDE has been proved to retain the efficiency and robustness properties under certain conditions. Wu et al. (2010) studied the MHDE in a two-sample semiparametric model, Zhu et al. (2013) investigated MHDE under the same model but for survival data with care rate, while Chen and Wu (2013) applied the model to the classification of leukemia patients. Xiang et al. (2014) and Wu et al. (2017) examined the MHDE under two different semiparametric mixture models. Karunamuni and Wu (2017) examined the hypothesis testing based on MHDE in seminparametric models.

Suppose we observe i.i.d. r.v.s $X_1, \ldots, X_n$ with density $g_0 = f_{\theta,\eta}$ being a member of the following general semiparametric model

$$\mathcal{F} = \{f_{t,h} : \theta \in \Theta, \eta \in \mathcal{H}\}, \tag{2.2}$$

where Θ is a compact subset of $\mathbb{R}^p$ and $\mathcal{H}$ is an arbitrary set of infinite dimension. In general, θ is the parameter of interest with η the nuisance parameter. To make estimating θ meaningful, assume $\mathcal{F}$ is identifiable in the sense that, if the Hellinger distance between f_{t_1,h_1} and f_{t_2,h_2} is 0, i.e. $\|f_{t_1,h_1}^{1/2} - f_{t_2,h_2}^{1/2}\| = 0$, then $t_1 = t_2$ and $h_1 = h_2$.

For model (2.2), Wu and Karunamuni (2012) focused on the case when an estimator $\hat{\eta}$ of η, based on either the same data or other resource, is available. They proposed a plug-in MHDE given by

$$\hat{\theta} = \arg\min_{t\in\Theta} \|f_{t,\hat{\eta}}^{1/2} - \hat{f}^{1/2}\|$$

and proved under certain conditions that this MHDE has good properties such as consistency, asymptotic normality, efficiency and robustness.

For model (2.2), if an estimator of η is not available, Wu and Karunamuni (2015) first introduced the MPHDE. The MPHDE is obtained by first profiling out the infinite-dimensional nuisance parameter and then minimizing the profiled Hellinger distance. Wu et al. (2017) applied the MPHDE to the two-component location-shifted mixture model in (1.4) and developed an algorithm to calculate the MPHDE. Their results showed that the MPHDE is very competitive under model assumption than some existing methods such as MLE while it has much better performance when outliers are present. However their algorithm involves updating the nuisance parameter estimate through an optimization which is comparatively computing extensive. In this thesis we explore an alternative of their estimation which is computationally easier. For model (1.4), we follow the outline of Wu and Karunamuni (2012) and propose to use a particular estimator of f, based on the same data sample, to construct a plug-in MHDE of the parameter $\theta = (\pi, \mu_1, \mu_2)^T$.

2.2 Estimation of the nuisance parameter f

For model (1.4), Bordes et al. (2006) expressed f as a function of h and θ by inverting. The key idea is to rewrite (1.4) as

$$f(x) = \frac{1}{1-\pi} h(x+\mu_2) + \frac{-\pi}{1-\pi} f(x+\delta), \quad \forall x \in \mathbb{R}, \tag{2.3}$$

where $\delta = \mu_2 - \mu_1 \neq 0$, and then use (2.3) as a recurrence formula. Let m be a positive integer. By using (2.3) m times, we have

$$f(x) = \frac{1}{1-\pi} \sum_{i=0}^{m-1} \left(\frac{-\pi}{1-\pi}\right)^i h(x+\mu_2+i\delta) + \left(\frac{-\pi}{1-\pi}\right)^m f(x+m\delta), \quad \forall x \in \mathbb{R}. \tag{2.4}$$

Let us show that when $\pi \in (0, 0.5)$,

$$f(x) = \frac{1}{1-\pi} \sum_{i \geq 0} \left(\frac{-\pi}{1-\pi} \right)^i h(x + \mu_2 + i\delta) \quad \text{for almost all } x \in \mathbb{R}. \tag{2.5}$$

Let $\|\cdot\|_1$ denote the L_1-norm. If we denote by $g(x)$ the right-hand side in (2.5), then by (2.4) and $\pi \in (0, 0.5)$ we have, for all $m \geq 1$,

$$\begin{aligned}
\|f - g\|_1 &= \Bigg\| \frac{1}{1-\pi} \sum_{i=0}^{m-1} \left(\frac{-\pi}{1-\pi} \right)^i h(x + \mu_2 + i\delta) + \left(\frac{-\pi}{1-\pi} \right)^m f(x + m\delta) \\
&\qquad - \frac{1}{1-\pi} \sum_{i \geq 0} \left(\frac{-\pi}{1-\pi} \right)^i h(x + \mu_2 + i\delta) \Bigg\|_1 \\
&= \left\| \left(\frac{-\pi}{1-\pi} \right)^m f(x + m\delta) - \frac{1}{1-\pi} \sum_{i \geq m} \left(\frac{-\pi}{1-\pi} \right)^i h(x + \mu_2 + i\delta) \right\|_1 \\
&\leq \left| \left(\frac{-\pi}{1-\pi} \right)^m \right| \cdot \|f(x + m\delta)\|_1 + \frac{1}{1-\pi} \sum_{i \geq m} \left| \left(\frac{-\pi}{1-\pi} \right)^i \right| \cdot \|h(x + \mu_2 + i\delta)\|_1 \\
&= \left(\frac{\pi}{1-\pi} \right)^m + \frac{1}{1-\pi} \sum_{i \geq m} \left(\frac{\pi}{1-\pi} \right)^i \\
&= \left(\frac{\pi}{1-\pi} \right)^m + \left(\frac{1}{1-2\pi} \right) \left(\frac{\pi}{1-\pi} \right)^m \\
&= \frac{2(1-\pi)}{1-2\pi} \left(\frac{\pi}{1-\pi} \right)^m.
\end{aligned}$$

By $\pi \in (0, 0.5)$ again we have

$$0 \;\leq\; \|f - g\|_1 \;\leq\; \lim_{m \to \infty} \frac{2(1-\pi)}{1-2\pi} \left(\frac{\pi}{1-\pi} \right)^m \;=\; 0,$$

i.e. $\|f(x) - g(x)\|_1 = 0$. Thus $f(x) = g(x)$ for almost every $x \in \mathbb{R}$, i.e. (2.5) holds.

Note that (2.5) is valid only for $\pi \in (0, 0.5)$. Since the parameter space for π in model (1.4) is $(0, 0.5) \cup (0.5, 1)$, we now consider the case of $\pi \in (0.5, 1)$. Because model (1.4) is invariant by permutation of (π, μ_1, μ_2) and $(1 - \pi, \mu_2, \mu_1)$, by replacing (π, μ_1, μ_2) with $(1 - \pi, \mu_2, \mu_1)$ in (2.5) we can easily obtain the inversion formula for $\pi \in (0.5, 1)$ as

$$f(x) = \frac{1}{\pi} \sum_{i \geq 0} \left(\frac{\pi - 1}{\pi} \right)^i h(x + \mu_1 + i(\mu_1 - \mu_2)) \quad \text{for almost all } x \in \mathbb{R}.$$

Suppose we have an appropriate nonparametric estimator $\hat{h}$ of h and a consistent estimator $\hat{\theta}^{(0)} = (\hat{\pi}^{(0)}, \hat{\mu}_1^{(0)}, \hat{\mu}_2^{(0)})^T$ of $\theta = (\pi, \mu_1, \mu_2)^T$. If $\hat{\pi}^{(0)} < 0.5$, then by applying (2.5) to $\hat{h}$ we can estimate

f by

$$\tilde{f}(x) = \frac{1}{1-\hat{\pi}^{(0)}} \sum_{i\geq 0} \left(\frac{-\hat{\pi}^{(0)}}{1-\hat{\pi}^{(0)}} \right)^i \hat{h}\left(x + \hat{\mu}_2^{(0)} + i(\hat{\mu}_2^{(0)} - \hat{\mu}_1^{(0)}) \right). \tag{2.6}$$

Similarly, if $\hat{\pi}^{(0)} > 0.5$ then we estimate f by

$$\tilde{f}(x) = \frac{1}{\hat{\pi}^{(0)}} \sum_{i\geq 0} \left(\frac{\hat{\pi}^{(0)} - 1}{\hat{\pi}^{(0)}} \right)^i \hat{h}\left(x + \hat{\mu}_1^{(0)} + i(\hat{\mu}_1^{(0)} - \hat{\mu}_2^{(0)}) \right). \tag{2.7}$$

Since f in model (1.4) is symmetric about 0, we can symmetrize $\tilde{f}$ and thus estimate f by

$$\tilde{f}_1(x) = \frac{\tilde{f}(-x) + \tilde{f}(x)}{2}. \tag{2.8}$$

Though h belongs to the mixture model (1.4), this is no longer true for its approximate $\hat{h}$ and, as a result, $\tilde{f}_1$ given in (2.8) through (2.6) or (2.7) can not be a p.d.f.. Even though $\tilde{f}_1$ still has integral 1, but it is not non-negative. To resolve this problem, Bordes et al. (2006) transformed $\tilde{f}_1$ into a density given by

$$\hat{f}(x) = \frac{\tilde{f}_1(x) I_{\{\tilde{f}_1(x)\geq 0\}}}{\int \tilde{f}_1(x) I_{\{\tilde{f}_1(x)\geq 0\}} dx} \tag{2.9}$$

with I_A the indicator function over set A. Bordes et al. (2006) showed that $\|\hat{f} - f\|_1 \to 0$ almost surely.

2.3 MHDE of the parameter θ

Suppose a random sample $X_1, \ldots, X_n$ is from the population $h_{\theta,f}$ given in (1.4). Based on this random sample, we can estimate $h_{\theta,f}$ nonparametrically by

$$\hat{h}(x) = \frac{1}{nb_n} \sum_{j=1}^{n} K\left(\frac{x - X_j}{b_n} \right), \tag{2.10}$$

where $K(\cdot)$ is a kernel density function and b_n is a sequence of bandwidths satisfying $b_n \to 0$ and $nb_n \to \infty$ as $n \to \infty$.In this thesis,we use Gaussian density for $K(\cdot)$ With this $\hat{h}$ and some initial estimate $\hat{\theta}^{(0)} = (\hat{\pi}^{(0)}, \hat{\theta}_1^{(0)}, \hat{\theta}_2^{(0)})^T$, we can calculate the estimate $\hat{f}$ given in (2.9). Finally the MHDE of θ is given by

$$\hat{\theta} = \arg\min_{\theta\in\Theta} \|h_{\theta,\hat{f}}^{1/2} - \hat{h}^{1/2}\|. \tag{2.11}$$

Since the MHDE $\hat{\theta}$ defined in (2.11) doesn't have an explicit expression, we will do the optimization numerically. In addition, to relax the dependence on the initial estimate $\hat{\theta}^{(0)}$, we will do iterations to update $\hat{f}$ and $\hat{\theta}$ until the latter converges. Therefore, we propose the following iterative algorithm to calculate the MHDE $\hat{\theta}$. Suppose the k^{th} estimates of $\theta = (\pi, \mu_1, \mu_2)^T$ and f are respectively $\hat{\theta}^{(k)} = (\hat{\pi}^{(k)}, \hat{\mu}_1^{(k)}, \hat{\mu}_2^{(k)})^T$ and $\hat{f}^{(k)}$, and $\hat{h}$ is the kernel estimate of h given in (2.10).

Step 1. For fixed $\hat{\pi}^{(k)}$, $\hat{\mu}_1^{(k)}$ and $\hat{\mu}_2^{(k)}$, find $\hat{f}^{(k+1)}$ in the following way:

If $\hat{\pi}^{(k)} < 0.5$, then

$$\tilde{f}^{(k+1)}(x) = \frac{1}{1-\hat{\pi}^{(k)}} \sum_{i \geq 0} \left(\frac{-\hat{\pi}^{(k)}}{1-\hat{\pi}^{(k)}} \right)^i \hat{h}\left(x + \hat{\mu}_2^{(k)} + i(\hat{\mu}_2^{(k)} - \hat{\mu}_1^{(k)})\right).$$

If $0.5 < \hat{\pi}^{(k)} < 1$, then

$$\tilde{f}^{(k+1)}(x) = \frac{1}{\hat{\pi}^{(k)}} \sum_{i \geq 0} \left(\frac{\hat{\pi}^{(k)} - 1}{\hat{\pi}^{(k)}} \right)^i \hat{h}\left(x + \hat{\mu}_1^{(k)} + i(\hat{\mu}_1^{(k)} - \hat{\mu}_2^{(k)})\right).$$

Then calculate

$$\tilde{f}_1^{(k+1)}(x) = \frac{\tilde{f}^{(k+1)}(-x) + \tilde{f}^{(k+1)}(x)}{2},$$

$$\hat{f}^{(k+1)}(x) = \frac{\tilde{f}_1^{(k+1)}(x) I_{\{\tilde{f}_1^{(k+1)}(x) \geq 0\}}}{\int \tilde{f}_1^{(k+1)}(x) I_{\{\tilde{f}_1^{(k+1)}(x) \geq 0\}} dx}.$$

Step 2. For fixed $\hat{f}^{(k+1)}$, find $\hat{\pi}^{(k+1)}$, $\hat{\mu}_1^{(k+1)}$ and $\hat{\mu}_2^{(k+1)}$ by minimizing

$$\left\| \left[\hat{\pi}^{(k+1)} \hat{f}^{(k+1)}(\cdot - \hat{\mu}_1^{(k+1)}) + (1 - \hat{\pi}^{(k+1)}) \hat{f}^{(k+1)}(\cdot - \hat{\mu}_2^{(k+1)}) \right]^{1/2} - \hat{h}^{1/2}(\cdot) \right\|.$$

Then go to Step 1.

In Step 2, we use the package 'fminsearch' in Matlab to minimize the Hellinger distance.

2.4 Choice of initial parameter estimate

In this thesis, the method used to obtain initial estimates of $\theta = (\pi, \mu_1, \mu_2)$ is based on an intuitive idea. The idea is that if we plot the kernel density estimate $\hat{h}$ given in (2.10) and see a bi-mode shape, then we expect that μ_1 and μ_2 are the two modes of h and thus can use the two modes of $\hat{h}$ to estimate μ_1 and μ_2. To practically implement this idea, we find the point $\hat{\mu}$ at which the global maximum of $\hat{h}$ is obtained. To determine whether $\hat{\mu}$ is a more reasonable estimate of μ_1 or μ_2, we calculate $\int_{\hat{\mu}}^{\infty} \hat{h}(x)dx$. If this integral is greater than 0.5, then we set the initial estimate of μ_1 as $\hat{\mu}_1^{(0)} = \hat{\mu}$, otherwise $\hat{\mu}_2^{(0)} = \hat{\mu}$. In practice this integral could be replaced by its approximation $\int_{\hat{\mu}}^{X_{(n)}+5b_n} \hat{h}(x)dx$, where $X_{(n)}$ is the n^{th} order statistic (the maximum) of the observations and b_n is the bandwidth used in $\hat{h}$. An example is given in Figure 2.1 where $\hat{\mu} = 1.9963$ and the area of the shaded tail is 0.338. Thus we use 1.9963 as the initial estimate of μ_2.

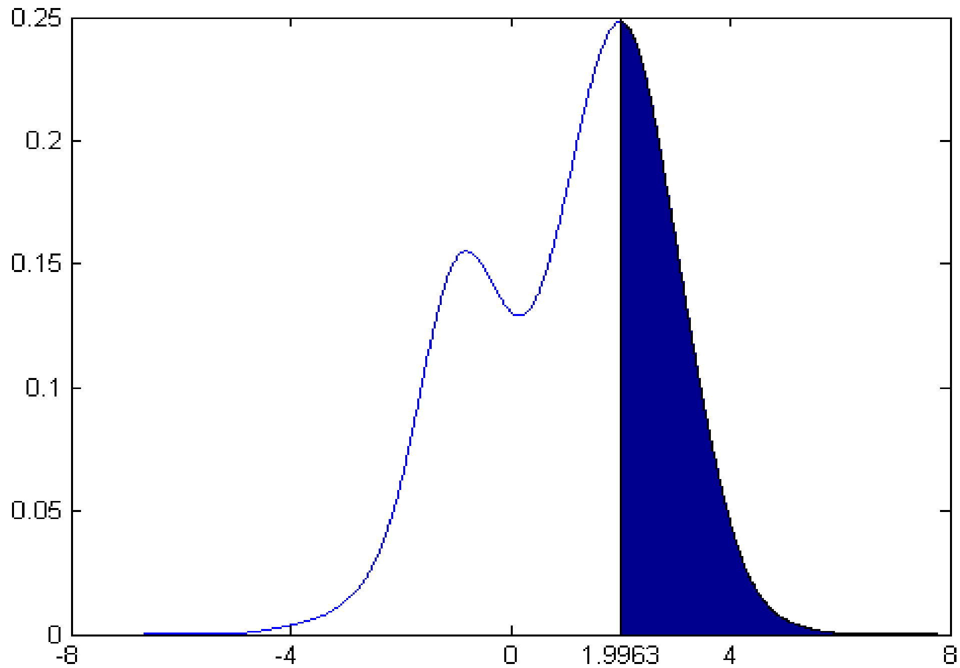

Figure 2.1: An example of bi-model $\hat{h}$.

Now we assume $\hat{\mu}_2^{(0)} = \hat{\mu}$ and discuss how to choose $\hat{\mu}_1^{(0)}$, and the process will be similar for the case that $\hat{\mu}_1^{(0)} = \hat{\mu}$. For ideal cases such as the one shown in Figure 2.1, there exist at least one local maximum less than $\hat{\mu}_2^{(0)}$ and among those local maxima we take the point at which the highest value is obtained as $\hat{h}(x)$ as $\mu_1^{(0)}$. If there is no local maximum less than $\hat{\mu}_2^{(0)}$, like the $\hat{h}$ displayed in Figure 2.2, we will use the following 'reflection' method to find $\hat{\mu}_1^{(0)}$. We first find the reflection of

$\hat{h}(x)$ over $x \in [\hat{\mu}_2^{(0)}, \infty)$ with respect to $x = \hat{\mu}_2^{(0)}$, and we denote it as $t(x)$. In practice we only need to find the reflection of $\hat{h}$ over a large enough interval, say, $[\hat{\mu}_2^{(0)}, X_{(n)} + 5b_n]$ instead of $[\hat{\mu}_2^{(0)}, \infty)$ for simpler calculation but still with good accuracy. An example of $t(x)$ is shown in Figure 2.2 as the dotted line. Now we define a new function $r(x)$ as

$$r(x) = \begin{cases} \hat{h}(x), & \text{if } x \leq 2\hat{\mu}_2^{(0)} - X_{(n)} - 5b_n, \\ \hat{h}(x) - t(x), & \text{if } 2\hat{\mu}_2^{(0)} - X_{(n)} - 5b_n < x \leq \hat{\mu}_2^{(0)}. \end{cases}$$

In another word, $r(x)$ is the function over $(-\infty, \hat{\mu}_2^{(0)}]$ that measures the difference between $\hat{h}$ and t. The function $r(x)$ 'wipe off' the weighted second component $(1 - \pi)f(x - \mu_2)$ from h and thus serves as a crude estimate of the weighted first component $\pi f(x - \mu_1)$. Finally we use the mode of $r(x)$ as $\hat{\mu}_1^{(0)}$, the initial estimate of μ_1. The corresponding function r of the example in Figure 2.2 is given in Figure 2.3.

To give the initial estimate $\hat{\pi}^{(0)}$ of π, note that

$$\begin{aligned} \frac{h(\mu_1)}{h(\mu_1) + h(\mu_2)} - \pi &= \frac{\pi f(0) + (1 - \pi) f(\mu_1 - \mu_2)}{f(0) + f(\mu_1 - \mu_2)} - \pi \\ &= \frac{(1 - 2\pi) f(\mu_1 - \mu_2)}{f(0) + f(\mu_1 - \mu_2)}. \end{aligned} \tag{2.12}$$

If π is close to 0.5 or if f is uni-model and μ_1 is quite different from μ_2, then the quantity calculated in (2.12) will be close to 0. So intuitively we can use

$$\hat{\pi}^{(0)} = \frac{\hat{h}(\hat{\mu}_1^{(0)})}{\hat{h}(\hat{\mu}_1^{(0)}) + \hat{h}(\hat{\mu}_2^{(0)})}$$

as the initial estimate of π. Note that we don't need the initial estimate of f and the first estimate of f is $\hat{f}^{(1)}$ calculated based on $\hat{h}$ given in (2.10) and the initial estimates $\hat{\mu}_1^{(0)}$, $\hat{\mu}_2^{(0)}$ and $\hat{\pi}^{(0)}$.

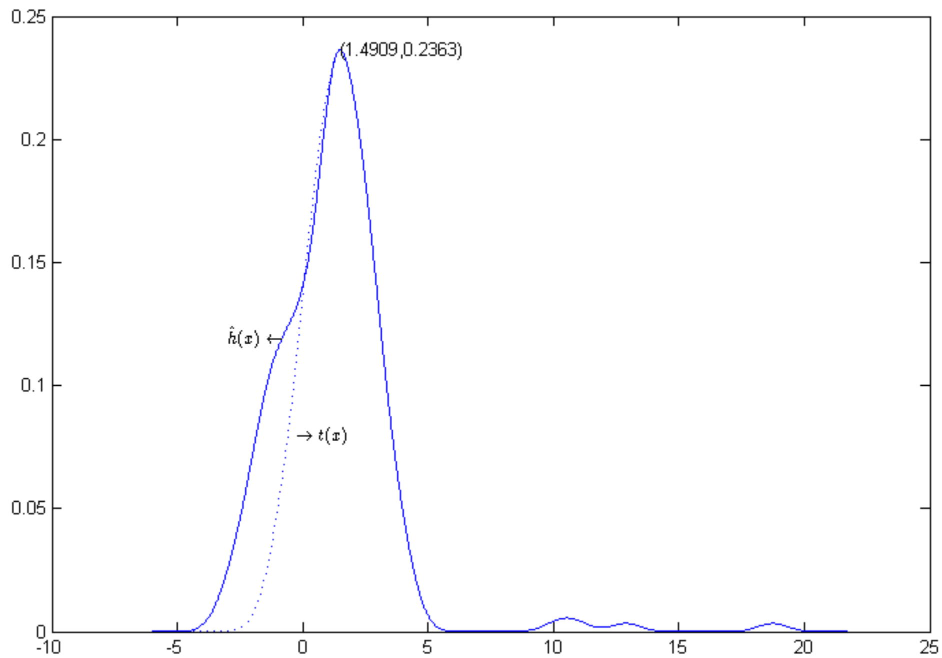

Figure 2.2: An example of $\hat{h}$ that uses the reflection technique.

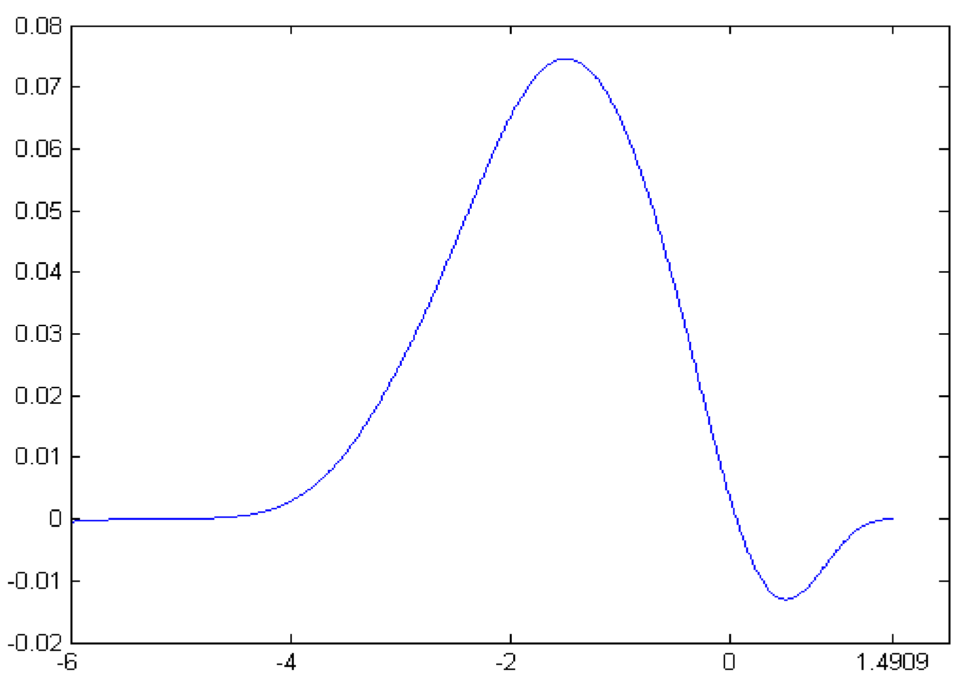

Figure 2.3: The function $r(x)$ of the example in Figure 2.2.

Chapter 3

NUMERICAL STUDIES

In this chapter, we use simulation studies to assess the performance of the proposed MHDE given in (2.11) and the iterative algorithm for calculating it. We analyze the Old Faithful Geyser data to illustrate the application of the proposed methodology. Particularly we compare the proposed MHDE with the MPHDE in Wu et al. (2017) through both simulation studies and real data analysis. In Section 3.1 we study the efficiency of the MHDE when the assumed model (1.4) is strictly correct. In Section 3.2 we explore the robustness properties (particularly resistance to outliers) of the MHDE when data is contaminated by outlying observations. Finally in Section 3.3, we analyze the waiting time between eruptions of the Old Faithful Geyser.

3.1 Efficiency study

To assess efficiency, we simulate the data strictly from the assumed two-component location-shifted model (1.4). In our simulation we consider mixtures of normal distributions, t distributions and uniform distributions. Specifically we assume the i.i.d. random samples are from one of the following populations:

Case 1: $X \sim \pi N(-1, 1) + (1 - \pi)N(2, 1) \Longrightarrow f = N(0, 1), \mu_1 = -1, \mu_2 = 2.$

Case 2: $f \sim t_5, \mu_1 = 0, \mu_2 = 3.$

Case 3: $X \sim \pi U(-1, 1) + (1 - \pi)U(0, 2) \Longrightarrow f = U(-1, 1), \mu_1 = 0, \mu_2 = 1.$

For every case, we consider varying π values: 0.15, 0.3 and 0.45. For cases $\pi > 0.5$, the simulation results are similar. Case 1 is used to test the efficiency of our proposed MHDE in normal cases. We use Cases 2 and 3 to demonstrate that the proposed MHDE can be adaptive to the non-normal component densities.

The measurements used to assess the efficiency of the MHDE given in (2.11) are the estimated bias and root mean squared error (RMSE) given respectively by

$$\widehat{Bias}(\hat{\theta}) = \frac{1}{N}\sum_{t=1}^{N}\left(\hat{\theta}_t - \theta\right),$$

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

where N is the number of repetitions, $\hat{\theta}_t$ is the estimate based on t^{th} replication and $\bar{\hat{\theta}}$ is the average of $\hat{\theta}$ over the N repetitions. Here we take $N = 200$.

We compare the simulation results of the proposed MHDE $\hat{\theta}$ given in (2.11) with those of the MPHDE $\tilde{\theta}$ presented in Wu et al. (2017). The reason why we compare it with MPHDE only but not with some other popular estimations, such as MLE, is that in Wu et al. (2017) the authors have already compared the MPHDE with MLE and SPEM (Benaglia et al., 2009). So the results presented below indicate the comparison of the MHDE with MLE. On the other hand, as initial value affects the performance of an estimator, we compare the proposed initial estimator $\hat{\theta}^{(0)}$ given in Section 2.3 with that in Wu et al. (2017), denoted by $\tilde{\theta}^{(0)}$, which is another special MHDE of θ calculated with E-M algorithm. It is obvious that $\hat{\theta}^{(0)}$ is computational easier than $\tilde{\theta}^{(0)}$. Considering different combinations, the simulation studies compare four estimations M1-M4:

M1: $\tilde{\theta}$ with initial estimate $\tilde{\theta}^{(0)}$.

M2: $\hat{\theta}$ with initial estimate $\tilde{\theta}^{(0)}$.

M3: $\hat{\theta}$ with initial estimate $\hat{\theta}^{(0)}$.

M4: modified M3.

M1': modified M1.

Estimations M4 and M1' are modified M3 and M1 respectively with the same modification technique used in Wu et al. (2017). In another word, M1' is the final estimation given by Wu et al. (2017) while M4 is our final proposed estimation. Taking M4 as example, the modification is described as following. If the resulted estimate makes the Hellinger distance (the objective function (2.11)) less than a predetermined small value ϵ, then we take it as the final MHDE $\hat{\theta}$ of θ.

Otherwise, we randomly choose another initial estimate, from the interval formed by the range of data, to carry out the calculation of MHDE. We repeatedly choose random initial estimates until the resulted MHDE makes the Hellinger distance in (2.11) less than ϵ. To limit computational expense, we allow maximum 49 repetitions of this recalculation of MHDE. After 49 repetitions if we still have no Hellinger distance less than ϵ, then we take the estimate that gives the smallest Hellinger distance among the 50 (49 repetitions plus the original one with $\hat{\theta}^{(0)}$ as the initial) as our final MHDE $\hat{\theta}$.

Tables 3.1, 3.2 and 3.3 show the simulation results for $n = 50$, 100 and 200 respectively. We make pairwise comparisons of M1 vs M2, M2 vs M3, M3 vs M4, and M4 vs M1' for the reason that we want to see the separate effect of the proposed MHDE $\hat{\theta}$ compared with $\tilde{\theta}$ and that of the proposed initial $\hat{\theta}^{(0)}$ compared with $\tilde{\theta}^{(0)}$. The comparison of M1 vs M2 tells the difference between the proposed MHDE $\hat{\theta}$ and the MPHDE $\tilde{\theta}$ in Wu et al. (2017). The comparison of M2 vs M3 tells the difference between the proposed initial $\hat{\theta}^{(0)}$ and the $\tilde{\theta}^{(0)}$ in Wu et al. (2017), when applied to the proposed MHDE. The comparison of M3 vs M4 tells the performance of the proposed modification to MHDE $\hat{\theta}$ with initial $\hat{\theta}^{(0)}$. The comparison of M4 vs M1' tells the difference between our final proposed MHDE and the final proposed MPHDE in Wu et al. (2017).

M1 vs M2: When sample size is small ($n = 50$), the proposed MHDE $\hat{\theta}$ tend to give larger bias but smaller RMSE than the MPHDE $\tilde{\theta}$. The difference diminishes as sample size increases ($n = 100, 200$) and when $n = 200$ the two estimators perform equivalently well. From the comparison between M1 and M2 we can see that the proposed MHDE $\hat{\theta}$ is very competitive with the MPHDE $\tilde{\theta}$ in Wu et al. (2017).

In terms of computing time, M2 is on average 2.7 times faster than M1.

M2 vs M3: When sample size is small ($n = 50$), the proposed initial estimate $\hat{\theta}^{(0)}$ performs slightly better than the initial $\tilde{\theta}^{(0)}$ used in Wu et al. (2017), when used to calculate the MHDE, in terms of both bias and RMSE. Reversely, when sample size is large ($n = 100, 200$) $\tilde{\theta}^{(0)}$ performs slightly better than $\hat{\theta}^{(0)}$ in terms of both bias and RMSE.

Table 3.1: $\widehat{Bias}$ ($\widehat{RMSE}$) of estimations of $\theta = (\pi_1, \mu_1, \mu_2)$ with $n = 50$.

Component density	Parameters	M1	M2	M3	M4	M1'
Normal	$\mu_1 = -1$	.148 (.640)	.081 (.619)	.306 (.891)	.105 (.658)	.203(.688)
	$\mu_2 = 2$	.028 (.221)	.034 (.236)	.127 (.398)	.042 (.235)	.028(.230)
	$\pi = 0.15$	.017 (.085)	.020 (.093)	.075 (.168)	.027 (.101)	.020(.094)
	$\mu_1 = -1$	.057 (.431)	.004 (.407)	.003 (.403)	.006 (.410)	.069(.486)
	$\mu_2 = 2$	-.027 (.250)	.033 (.256)	.013 (.235)	.016 (.233)	-.071(.325)
	$\pi = 0.3$	.004 (.090)	.018 (.096)	.014 (.080)	.016 (.088)	.000(.124)
	$\mu_1 = -1$	.090 (.455)	-.000 (.389)	.015 (.399)	.021 (.438)	.298(.647)
	$\mu_2 = 2$	-.095 (.492)	.048 (.379)	-.009 (.341)	-.000 (.370)	-.333(.554)
	$\pi = 0.45$	-.007 (.114)	.016 (.111)	.012 (.119)	.015 (.130)	-.004(.211)
t_5	$\mu_1 = 0$	-.002 (1.919)	-.129 (1.651)	-.059 (1.410)	.059 (.948)	.128(1.032)
	$\mu_2 = 3$	.392 (1.394)	.479 (1.436)	.156 (.532)	.188 (.660)	.050(.036)
	$\pi = 0.15$	.055 (.189)	.097 (.258)	.072 (.181)	.064 (.183)	.044(.149)
	$\mu_1 = 0$	.120 (1.735)	-.140 (1.395)	-.146 (1.143)	-.046 (.861)	.094(.925)
	$\mu_2 = 3$	.310 (1.593)	.404 (1.463)	.022 (.351)	.170 (.905)	-.038(.581)
	$\pi = 0.3$	.020 (.170)	.055 (.208)	.015 (.119)	.034 (.187)	.026(.206)
	$\mu_1 = 0$	-.063 (2.248)	-.273 (1.853)	-.034 (.791)	.060 (.992)	.275(1.064)
	$\mu_2 = 3$	.030 (1.412)	.282 (1.127)	.050 (.533)	.290 (1.077)	-.207(.994)
	$\pi = 0.45$	-.024 (.182)	.032 (.229)	.022 (.163)	.066 (.229)	.019(.280)
Uniform	$\mu_1 = 0$	.173 (.275)	.132 (.256)	.242 (.266)	.108 (.267)	.104(.282)
	$\mu_2 = 1$	.201 (.207)	.252 (.193)	.276 (.172)	.193 (.193)	.069(.164)
	$\pi = 0.15$	.192 (.164)	.223 (.177)	.284 (.173)	.183 (.181)	.093(.171)
	$\mu_1 = 0$	.000 (.241)	-.029 (.221)	.081 (.260)	.001 (.247)	.041(.237)
	$\mu_2 = 1$	-.001 (.208)	.150 (.177)	.142 (.176)	.114 (.184)	-.053(.197)
	$\pi = 0.3$	.009 (.146)	.102 (.154)	.152 (.152)	.102 (.161)	-.011(.199)
	$\mu_1 = 0$	.037 (.201)	-.097 (.185)	-.006 (.239)	-.047 (.230)	.092(.219)
	$\mu_2 = 1$	-.114 (.211)	.109 (.168)	.057 (.199)	.066 (.189)	-.167(.217)
	$\pi = 0.45$	-.047 (.154)	.026 (.139)	.047 (.137)	.033 (.154)	-.049(.233)

Table 3.2: $\widehat{Bias}$ ($\widehat{RMSE}$) of estimations of $\theta = (\pi_1, \mu_1, \mu_2)$ with $n = 100$.

Component density	Parameters	M1	M2	M3	M4	M1'
Normal	$\mu_1 = -1$	.126 (.395)	.079 (.386)	.242 (.659)	.095 (.412)	.138(.431)
	$\mu_2 = 2$	.010 (.154)	.012 (.155)	.076 (.329)	.024 (.155)	.013(.156)
	$\pi = 0.15$	.011 (.051)	.012 (.055)	.049 (.123)	.017 (.056)	.016(.072)
	$\mu_1 = -1$	.040 (.265)	-.001 (.266)	.007 (.289)	.000 (.265)	.056(.318)
	$\mu_2 = 2$	-.006 (.184)	.008 (.171)	.005 (.175)	.006 (.174)	.001(.210)
	$\pi = 0.3$	.007 (.056)	.010 (.058)	.011 (.060)	.010 (.056)	.012(.086)
	$\mu_1 = -1$	.018 (.243)	-.005 (.259)	.016 (.299)	.014 (.298)	.270(.570)
	$\mu_2 = 2$	-.063 (.263)	-.026 (.273)	-.047 (.282)	-.032 (.281)	-.211(.455)
	$\pi = 0.45$	-.007 (.068)	.001 (.083)	.001 (.077)	.003 (.084)	.026(.184)
t_5	$\mu_1 = 0$	.129 (1.888)	-.017 (1.589)	.009 (1.491)	.079 (.695)	.209(.729)
	$\mu_2 = 3$	.459 (1.664)	.490 (1.654)	.144 (.697)	.159 (.800)	.015(.194)
	$\pi = 0.15$	.047 (.181)	.082 (.236)	.059 (.164)	.039 (.148)	.030(.120)
	$\mu_1 = 0$	-.169 (2.486)	-.464 (1.792)	-.508 (1.454)	-.247 (.782)	-.013(.696)
	$\mu_2 = 3$	.226 (1.627)	.227 (1.448)	-.081 (.322)	-.021 (.550)	-.080(.306)
	$\pi = 0.3$	-.030 (.148)	.009 (.190)	-.024 (.101)	-.013 (.100)	-.010(.115)
	$\mu_1 = 0$	-.131 (1.599)	-.229 (1.359)	-.142 (.689)	-.080 (.899)	.162(.919)
	$\mu_2 = 3$	.181 (1.523)	.244 (1.371)	-.060 (.385)	.065 (.998)	-.310(.805)
	$\pi = 0.45$	-.034 (.177)	.004 (.200)	-.016 (.116)	-.004 (.211)	-.012(.271)
Uniform	$\mu_1 = 0$	.156 (.220)	.139 (.228)	.245 (.225)	.060 (.180)	.070(.176)
	$\mu_2 = 1$	.196 (.207)	.211 (.191)	.258 (.176)	.108 (.156)	.022(.092)
	$\pi = 0.15$	.165 (.160)	.195 (.178)	.267 (.181)	.101 (.144)	.024(.081)
	$\mu_1 = 0$	-.007 (.175)	-.001 (.160)	.149 (.221)	-.002 (.147)	.029(.147)
	$\mu_2 = 1$	.038 (.162)	.116 (.149)	.098 (.156)	.053 (.134)	-.021(.144)
	$\pi = 0.3$	.016 (.105)	.080 (.131)	.158 (.149)	.043 (.115)	-.009(.124)
	$\mu_1 = 0$	-.027 (.212)	-.079 (.142)	.057 (.221)	.026 (.201)	.099(.204)
	$\mu_2 = 1$	-.023 (.190)	.088 (.131)	-.023 (.184)	.000 (.176)	-.134(.220)
	$\pi = 0.45$	-.028 (.122)	.020 (.109)	.035 (.116)	.031 (.122)	-.045(.200)

Table 3.3: $\widehat{Bias}$ ($\widehat{RMSE}$) of estimations of $\theta = (\pi_1, \mu_1, \mu_2)$ with $n = 200$.

Component density	Parameters	M1	M2	M3	M4	M1'
Normal	$\mu_1 = -1$	.110 (.281)	.063 (.285)	.176 (.535)	.067 (.285)	.108(.285)
	$\mu_2 = 2$	.024 (.090)	.017 (.092)	.056 (.219)	.024 (.094)	.026(.093)
	$\pi = 0.15$	.012 (.036)	.010 (.037)	.033 (.101)	.013 (.037)	.013(.036)
	$\mu_1 = -1$	.054 (.170)	.020 (.173)	.018 (.173)	.018 (.174)	.054(.172)
	$\mu_2 = 2$	.022 (.101)	.019 (.099)	.020 (.102)	.021 (.101)	.026(.104)
	$\pi = 0.3$	.012 (.041)	.012 (.041)	.012 (.041)	.012 (.041)	.013(.041)
	$\mu_1 = -1$	.010 (.135)	-.002 (.170)	-.019 (.279)	.000 (.225)	.175(.484)
	$\mu_2 = 2$	.004 (.126)	.017 (.231)	.026 (.276)	-.006 (.216)	-.182(.450)
	$\pi = 0.45$	.002 (.042)	.000 (.057)	.004 (.053)	-.000 (.066)	.001(.173)
t_5	$\mu_1 = 0$	.429 (2.581)	.053 (1.482)	-.266 (1.545)	.002 (.519)	.051(.502)
	$\mu_2 = 3$	.815 (2.565)	.753 (2.357)	.114 (.786)	.062 (.444)	.010(.128)
	$\pi = 0.15$	.043 (.190)	.084 (.244)	.025 (.134)	.018 (.092)	.010(.074)
	$\mu_1 = 0$	-.029 (3.310)	-.403 (2.209)	-.388 (1.466)	-.059 (.437)	.035(.356)
	$\mu_2 = 3$	.487 (2.397)	.501 (2.333)	-.046 (.284)	.025 (.277)	-.012(.240)
	$\pi = 0.3$	-.015 (.145)	.035 (.195)	-.011 (.085)	.010 (.085)	.005(.093)
	$\mu_1 = 0$	-.267 (2.751)	-.429 (1.801)	-.246 (1.201)	-.186 (.992)	.201(.903)
	$\mu_2 = 3$	.320 (2.141)	.373 (2.035)	-.074 (.564)	-.057 (.847)	-.366(.717)
	$\pi = 0.45$	-.054 (.182)	-.006 (.200)	-.021 (.143)	-.024 (.175)	-.022(.256)
Uniform	$\mu_1 = 0$	.152 (.219)	.095 (.203)	.279 (.224)	.007 (.113)	.031(.112)
	$\mu_2 = 1$	.170 (.191)	.148 (.188)	.253 (.186)	.030 (.080)	.003(.035)
	$\pi = 0.15$	.144 (.163)	.136 (.183)	.274 (.192)	.027 (.077)	.000(.040)
	$\mu_1 = 0$	-.008 (.099)	-.021 (.129)	.193 (.220)	-.006 (.124)	-.004(.078)
	$\mu_2 = 1$	.046 (.101)	.086 (.128)	.041 (.157)	.015 (.081)	.009(.062)
	$\pi = 0.3$	.020 (.071)	.050 (.106)	.147 (.144)	.020 (.071)	.000(.053)
	$\mu_1 = 0$	-.057 (.100)	-.075 (.067)	.111 (.203)	.067 (.203)	.061(.197)
	$\mu_2 = 1$	.041 (.093)	.081 (.067)	-.084 (.160)	-.063 (.166)	-.078(.201)
	$\pi = 0.45$	-.006 (.064)	.011 (.058)	.026 (.094)	.012 (.093)	-.020(.174)

From the comparison between M2 and M3 we can conclude that the initial estimates $\hat{\theta}^{(0)}$ and $\tilde{\theta}^{(0)}$ are very competitive.

In terms of computing time, M3 takes slightly longer time than M2 with average ratio 1.2. In another word, the proposed initial $\hat{\theta}^{(0)}$ is about 1.2 times slower than $\tilde{\theta}^{(0)}$.

M3 vs M4: When sample size is small ($n = 50$), the MHDE $\hat{\theta}^{(0)}$ with initial estimate $\hat{\theta}^{(0)}$ has similar bias no matter the modification is implemented or not, but the modification tends to generate larger RMSE. When sample size is large ($n = 100, 200$), the modification reduces significantly both the bias and RMSE for most cases. Especially when $n = 200$, the modification almost always gives much smaller bias and RMSE than when modification is not used.

In terms of computing time, M4 takes much longer time than M3 with average ratio 12.5. This is expected due to multiple initial estimates.

M4 vs M1': Our final proposed MHDE M4 is slightly better than the final proposed MPHDE in Wu et al. (2017) when sample size is $n = 100$, while the two estimations are very competitive for sample sizes $n = 50$ and 200.

In terms of computing time, M4 is about 7 times faster than M1'.

3.2 Robustness study

In this section, we study the robustness properties of the proposed estimation. The robustness could be against model distribution misspecification or against outliers. This section studies the robustness in the latter sense. So we deliberately contaminate the data with some outlying observations. Particularly we generate i.i.d. data from the following distribution:

$$\text{Case 4: } X \sim 98\% \cdot [\pi N(-1, 1) + (1 - \pi) N(2, 1)] + 2\% \cdot U(10, 20).$$

In another word, the population $\pi N(-1, 1) + (1 - \pi)N(2, 1)$ is contaminated with 2% outlying observations from $U(10, 20)$, the uniform distribution over interval $[10, 20]$. We apply estimation methods M1-M4 to Case 4 with varying sample size $n = 50, 100, 200$ and present the estimated bias and RMSE for each method in Table 3.4. Again we make pairwise comparisons of M1 vs M2, M2 vs M3, M3 vs M4, and M4 vs M1'.

M1 vs M2: When the data is contaminated, both M1 and M2 give very large bias for estimating any parameter especially μ_2 and μ_1 as both of them treat the second component $N(2, 1)$ as the first component and $U(10, 20)$ as the second component.

M2 vs M3: No matter for which sample size, the bias of M3 is always much smaller (at different magnitude) than M2. M3 obtains this superiority over M2 at the price of larger RMSE most time even though improved a bit when $n = 200$. From this comparison we see that the choice of initial estimate plays an important role in the MHDE for contaminated data. The proposed initial estimate $\hat{\theta}^{(0)}$ reduces bias dramatically compared with $\tilde{\theta}^{(0)}$. Even though $\hat{\theta}^{(0)}$ produces a larger RMSE than $\tilde{\theta}^{(0)}$, the loss in terms of RMSE is much smaller than the gain in terms of bias.

Here we observe something very different from the case of no data contamination (Case 1). For Case 1, M2 and M3 perform competitively when $n = 50$ but M3 performs worse than M2 for $n = 100$ and 200, even though not significantly, in terms of both bias and RMSE. Therefore, the initial estimator $\hat{\theta}^{(0)}$ may not be as efficient as $\tilde{\theta}^{(0)}$, but it seems to be more robust to outliers and thus more likely to lead to global minimizer, instead of local minimizer, of corresponding Hellinger distance than $\tilde{\theta}^{(0)}$. The large bias in M2 is due to the fact that the non-robust initial $\tilde{\theta}^{(0)}$ is likely to lead to a local minimizer.

M3 vs M4: No matter for which sample size, M4 significantly reduces the bias and RMSE for over half of the cases (9 cases at each sample size). Even though M3 gives smaller

Table 3.4: $\widehat{Bias}$ ($\widehat{RMSE}$) of estimations of $\theta = (\pi_1, \mu_1, \mu_2)$ for Case 4.

Normal + outliers	Parameters	M1	M2	M3	M4	M1'
	$\mu_1 = -1$	3.776 (2.357)	2.597 (.238)	.514 (1.141)	.062 (.670)	.126(.774)
	$\mu_2 = 2$	13.221 (2.881)	13.238 (2.883)	2.179 (4.948)	.005 (.236)	-.031(.240)
	$\pi = 0.15$	.672 (.283)	.800 (.000)	.146 (.303)	.011 (.101)	.010(.135)
	$\mu_1 = -1$	4.391 (3.316)	2.003 (.295)	-.029 (.488)	.154 (.710)	.149(.699)
$n = 50$	$\mu_2 = 2$	13.204 (2.908)	13.228 (2.898)	.080 (1.297)	-.185 (.376)	-.298(.476)
	$\pi = 0.3$	.338 (.398)	.650 (.000)	-.001 (.113)	.007 (.176)	.037(.175)
	$\mu_1 = -1$	3.804 (3.622)	1.600 (.235)	.177 (.647)	.726 (.730)	.707(.740)
	$\mu_2 = 2$	13.087 (3.081)	13.184 (2.968)	1.661 (4.549)	-.415 (1.389)	-.778(.636)
	$\pi = 0.45$	-.082 (.407)	.500 (.000)	.044 (.207)	.035 (.277)	-.028(.313)
	$\mu_1 = -1$	4.283 (3.976)	2.598 (.177)	.197 (.821)	.008 (.479)	.022(.522)
	$\mu_2 = 2$	13.470 (2.733)	13.053 (2.265)	.996 (3.552)	.079 (1.144)	-.027(.163)
	$\pi = 0.15$	.492 (.416)	.800 (.000)	.065 (.223)	.005 (.078)	-.002(.076)
	$\mu_1 = -1$	4.879 (4.462)	1.983 (.192)	-.018 (.317)	.039 (.444)	.137(.604)
$n = 100$	$\mu_2 = 2$	13.270 (2.540)	13.072 (2.337)	.035 (1.061)	-.100 (.288)	-.187(.376)
	$\pi = 0.3$	.142 (.435)	.650 (.000)	-.006 (.077)	-.013 (.092)	.009(.175)
	$\mu_1 = -1$	4.897 (5.006)	1.594 (.156)	.086 (.549)	.589 (.697)	.785(.622)
	$\mu_2 = 2$	13.027 (2.357)	12.909 (2.282)	.707 (3.179)	-.068 (2.396)	-.879(.511)
	$\pi = 0.45$	-.221 (.351)	.500 (.000)	.022 (.159)	.039 (.249)	-.051(.279)
	$\mu_1 = -1$	3.357 (2.888)	2.606 (.132)	.279 (.842)	.010 (.291)	.017(.297)
	$\mu_2 = 2$	13.710 (1.959)	13.260 (1.960)	1.294 (4.033)	.016 (.100)	.002(.107)
	$\pi = 0.15$	0.487 (.426)	.800 (.000)	.085 (.241)	.005 (.037)	.001(.038)
	$\mu_1 = -1$	3.998 (3.980)	1.963 (.142)	.024 (.279)	.016 (.316)	.105(.423)
$n = 200$	$\mu_2 = 2$	13.332 (1.766)	13.288 (1.987)	.105 (1.061)	-.015 (.148)	-.084(.265)
	$\pi = 0.3$	.166 (.443)	.650 (.000)	.009 (.076)	.007 (.072)	.011(.124)
	$\mu_1 = -1$	3.474 (3.942)	1.593 (.101)	-.004 (.375)	.428 (.676)	.932(.660)
	$\mu_2 = 2$	13.271 (1.569)	13.169 (1.875)	.066 (.982)	-.145 (1.897)	-1.025(.491)
	$\pi = 0.45$	-.301 (.279)	.500 (.000)	.000 (.091)	.006 (.223)	-.053(.315)

bias and RMSE for about half of the cases, the magnitude of bias and RMSE from the two estimations for those cases are still at the same level. Comparatively, in those cases when M4 performs better, the magnitude of bias and RMSE from M4 is much smaller than those from M3.

M4 vs M1': The final proposed MHDE M4 performs better than the final proposed MPHDE M1' in Wu et al. (2017), especially in terms of bias.

3.3 Real data analysis

In this section, we use the Old Faithful Geyser data to demonstrate the use of the proposed estimation methods. The Old Faithful Geyser data is publicly available in $\mathbb{R}$. It consists of 272 observations and 2 variables, Waiting Time between eruptions and Duration of eruptions in minutes, for the Old Faithful Geysar in the Yellowstone National Park and is of continuous measurement from August 1 to August 15, 1985.For this real data we will analyze the variable Waiting Time which was also used by Wu et al. (2017) to illustrate their MPHDE.

Figure 3.1 is the histogram of the variable Waiting Time. We can see from it that the data is approximately a mixture of two normal distributions. Assuming two-component location-shifted model with normal components, we can calculate the MLEs of the two location parameters and use them as a benchmark for comparison with other methods. Without assumption of normal components, we calculate both the proposed MHDE (the optimal M4) and the MPHD with modification technique as in Wu et al. (2017). The results are presented in Table 3.5. From Table 3.5 we can see that both the MHDE and MPHDE give close results to the parametric MLE even without the assumption of normal components. Nevertheless, the MHDE is a little bit closer to the MLE than the MPHDE.

To demonstrate the robustness property for this real data, we add five identical outliers '0' to the data set and the recalculated estimates are reported in Table 3.6. From Table 3.6 we can see that both the proposed MHDE and the MPHDE are very resistant to outliers in the sense that they

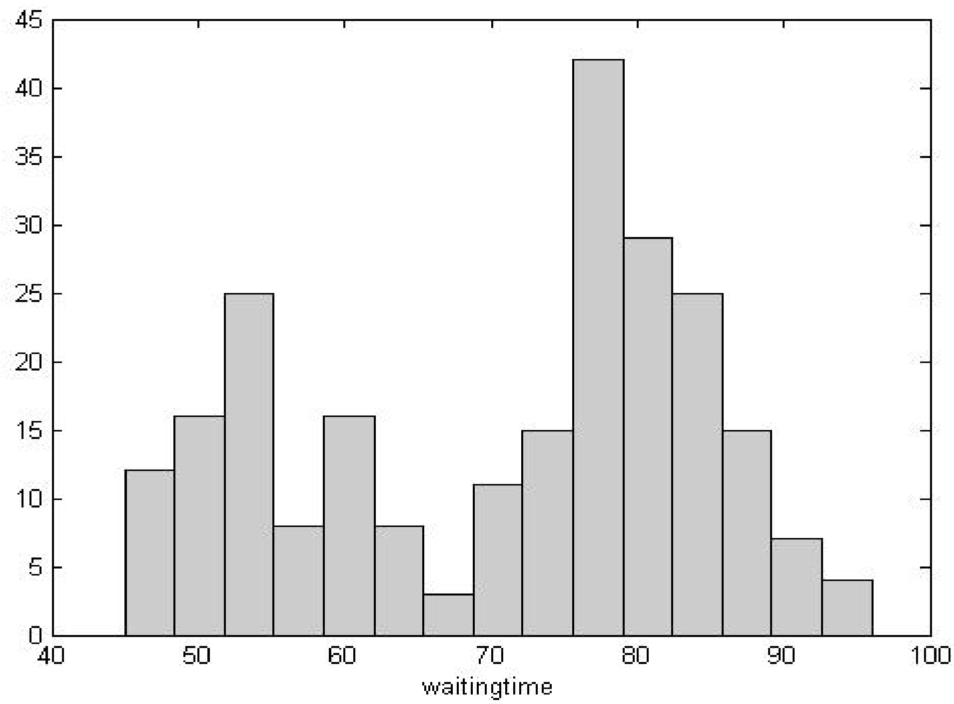

Figure 3.1: Histogram for the waiting time data of the Old Faithful Geyser.

Table 3.5: Estimation for the waiting time data of the Old Faithful Geyser.

Method	π	μ_1	μ_2
MHDE (M4)	0.358	54.70	80.05
MPHDE	0.359	54.90	80.30
MLE	0.361	54.61	80.09

don't change much before and after the data is contaminated by outlying observations. The MHDE and the MPHDE give very close estimates. However, the MLE is dramatically influenced by the outliers and thus not robust. The MLE treats the outliers as one mixture component located around 0 and all other real data as another component located around 70.

Table 3.6: Estimation for the contaminated waiting time data of the Old Faithful Geyser.

Method	π	μ_1	μ_2
MHDE (M4)	0.355	54.48	80.17
MPHDE	0.350	54.71	80.17
MLE	0.018	0.061	70.90

Chapter 4

CONCLUDING REMARKS

Over the last two decades, semiparametric mixture model receives increasing attention. In this thesis we consider a semiparametric two-component location-shifted mixture model (1.4). We propose to use the MHDE $\hat{\theta}$ to estimate the parameter θ. However, in the construction of MHDE, it is critical to give an appropriate estimation of the nuisance parameter f (an unknown function). Thus, we propose to use the bounded linear operator introduced by Bordes et al. (2006) to estimate the infinite-dimensional nuisance parameter. To facilitate the calculation of the MHDE, we develop an iterative algorithm and propose a novel initial estimation of the parameters of our interest. To improve performance, we adopt the modification technique used in Wu et al. (2017). Through the simulation studies we observe that

1. The proposed MHDE $\hat{\theta}$ is very competitive in terms of efficiency and is on average about 2.7 times computationally faster than the MPHDE $\tilde{\theta}$ in Wu et al. (2017).

2. The proposed initial $\hat{\theta}^{(0)}$ is very competitive in terms of efficiency but on average about 1.2 times computationally slower than the initial $\tilde{\theta}^{(0)}$ in Wu et al. (2017). In addition, $\hat{\theta}^{(0)}$ is a much more robust initial than $\tilde{\theta}^{(0)}$ when data is contaminated.

3. When data has no outliers, the modification technique reduces bias but increases RMSE for small sample size, while it reduces dramatically both the bias and RMSE for large sample size. For contaminated data, the modification technique reduces the magnitude of bias and RMSE.

4. Our proposed MHDE, proposed initial and the modification technique together (M4) is very competitive with the method proposed in Wu et al. (2017) no matter data is contaminated with outliers or not.

On the other hand, the proposed estimation has its potential limitations. First, Bordes et al. (2006) found that there does not exist an inversion formula like (2.5) for some three-component mixture models. That's to say, it is not trivial to extend the methodology proposed in this thesis for two-component mixture models to three-component ones. Second, even though the proposed initial estimate is very simple and intuitive, it is essentially for the case of f being unimodal. Though unimodal is commonly assumed and is often valid in practice, in case of multimodal f, the proposed initial may give unreliable estimate. Third, in the modification technique, we keep choosing random initial estimates from the interval formed by the range of data until we obtain one with corresponding Hellinger distance smaller than a predetermined small value or we reach the maximum 49 random initials. As the range of data gets wider, it is possible that no one among the 49 random initials satisfies the requirement that the corresponding Hellinger distance is smaller than a predetermined small value. To solve this problem, one may increase the allowed maximum number of random initials or, more efficiently, choose random initials within a smaller range around the centre of the data rather than over the whole range of the data. All these three aspects could be explored in the future work.

Bibliography

[1] Beren, R. (1977). Minimum Hellinger distance estimates for parametric models. *The Annals of Statistics*, **5**, 445-463.

[2] Benaglia, T., Chauveau, D. and Hunter D.R. (2009). An EM-like algorithm for semi- and non-parametric estimation in multivariate mixtures. *Journal of Computational and Graphical Statistics*, **18**, 505-526.

[3] Bordes, L., Chauveau, D. and Vandekerkhove, P. (2007). A stochastic EM algorithm for a semiparametric mixture model.*Computational Statistics and Data Analysis*, **51**, 5429-5443.

[4] Bordes, L., Mottelet, S. and Vandekerkhove, P. (2006). Semiparametric estimation of a two-component mixture model. *Annals of Statistics*, **34**,1204-1232.

[5] Bordes, L. and Vandekerkhove, P. (2010). Semiparametric two-component mixture model with a known component: An asymptotically normal estimator. *Mathematical Methods of Statistics*, **19**, 22-41.

[6] Butucea, C. and Vandekerkhove, P. (2013). Semiparametric mixtures of symmetric distributions. *Scandinavian Journal of Statistics*, **41**, 227-239.

[7] Chee,C and Wang,Y. (2013). Estimation of finite mixtures with symmetric components. *Statistics and Computing*, **2**, 233-249.

[8] Chen, G. and Wu, J. (2013). Molecular classification of acute leukemia. *Some Recent Advances in Mathematics and Statistics*, pages 60-74. World Scientific.

[9] Cutler, A. and Cordero-Braña, O.I. (1996). Minimum Hellinger distance estimation for finite mixture models. *Journal of the American Statistical Association*, **91**, 1716-1723.

[10] Dacunha-Castelle, D. and Gassiat, E. (1999). Testing the order of a model using locally conic parametrization: population mixtures and stationary ARMA processes. *Annals of Statistics*, **27**, 1178-1209.

[11] Frühwirth-Schnatter, S. (2006). *Finite mixture and markov switching models*. New York: Springer.

[12] Hosmer, D.W. (1973). A comparison of iterative maximum likelihood estimates of the parameters of a mixture of two normal distributions under three types of sample. *Biometrics*, **29**, 761-770.

[13] Hunter, D., Wang, S. and Hettmansperger, T.P. (2007). Inference for mixture symmetric distributions. *Annals of Statistics*, **35**, 224-251.

[14] Karlis, D. and Xekalaki, E. (1998). Minimum Hellinger distance estimation for poisson mixtures. *Computational Statistics* & *Data Analysis*, **29**, 81-103.

[15] Karunamuni, R.J. and Wu, J. (2011). One-step minimum Hellinger distance estimation. *Computational Statistics* & *Data Analysis*, **55**, 3148-3164.

[16] Karunamuni, R.J. and Wu, J. (2017). Efficient and robust tests for semiparametric models. *Annals of the Institute of Statistical Mathematics*. In Press.

[17] Keribin, C. (2000). Consistent Estimation of the Order of Mixture Models. *Sankhyā: The Indian Journal of Statistics, Series A (1961-2002)*, **62**, 49-66.

[18] Kottas, A. and Fellingham, G.W. (2012). Bayesian semiparametric modeling and inference with mixtures of symmetric distributions.*Statistics and Computing*, **22**, 93-106.

[19] Lindsay, B.G. (1995). *Mixture models: theory, geometry and applications*. NSF-CBMS Regional Conference Series in Probability and Statistics v5, Hayward, CA: Institute of Mathematical Statistics.

[20] Lu, Z., Hui, Y.V. and Lee, A.H. (2003). Minimum Hellinger distance estimation for finite mixtures of poisson regression models and its applications. *Biometrics*, **59**, 1016-1026.

[21] McLachlan, G.J. and Basford, K.E. (1988). *Mixture models:inference and applications to clustering*. New York: Marcel Dekker.

[22] McLachlan,G.J. and Peel, D. (2000). *Finite mixture models*. New York: Wiley.

[23] N'drin, J.A. and Hili, O. (2013). Parameter estimation of one-dimensional diffusion process by minimum Hellinger distance method. *Random Operators and Stochastic Equations*, **21**, 403-424.

[24] Prause, A., Steland, A. and Abujarad, M. (2016). Minimum Hellinger distance estimation for bivariate samples and time series with applications to nonlinear regression and copula-based models. *Metrika*, **79**, 425-455.

[25] Roeder, K. (1994). A graphical technique for determining the number of components in a mixture of normals. *Journal of American Statistical Association*, **89**, 487-495.

[26] Saraiva, E.F., Louzada, F. and Milan, L. (2014). Mixture models with an unknown number of components via a new posterior split-merge MCMC algorithm. *Applied Mathematics and Computation*, **244**, 959-975.

[27] Seo, B. (2017). The doubly smoothed maximum likelihood estimation for location-shifted semiparametric mixture. *Computational Statistics and Data Analysis*, **108**, 27-39.

[28] Simpson, D.G. (1987). Minimum Hellinger distance estimation for the analysis of count data. *Journal of the American statistical Association*, **82**, 802-807.

[29] Sriram, T. and Vidyashankar, A. (2000). Minimum Hellinger distance estimation for supercritical galton-watson processes. *Statistics & Probability Letters*, **50**, 331-342.

[30] Takada, T. (2009). Simulated minimum Hellinger distance estimation of stochastic volatility models. *Computational Statistics & Data Analysis*, **53**, 2390-2403.

[31] van der Vaart, A.W. (1996). Efficient maximum likelihood Estimation in semiparametric mixture models.*The Annals of Statistics*, **24**, 862-878.

[32] Woo, M.J. and Sriram, T. (2006). Robust estimation of mixture complexity. *Journal of the American Statistical Association*, **101**, 1475-1486.

[33] Woo, M.J. and Sriram, T. (2007). Robust estimation of mixture complexity for count data. *Computational Statistics & Data Analysis*, **51**, 4379-4392.

[34] Woodward, W.A., Whitney, P. and Eslinger, P.W. (1995). Minimum Hellinger distance estimation of mixture proportions. *Journal of Statistical Planning and Inference*, **48**, 303-319.

[35] Wu, J. and Karunamuni, R.J. (2009). On minimum Hellinger distance estimation. *Canadian Journal of Statistics*, **37**, 514-533.

[36] Wu, J. and Karunamuni, R.J. (2012). Efficient Hellinger distance estimates for semiparametric models. *Journal of Multivariate Analysis*, **107**, 1-23.

[37] Wu, J. and Karunamuni, R.J. (2015). Profile Hellinger distance estimation. *Statistics*, **49**, 711-740.

[38] Wu, J., Karunamuni, R., and Zhang, B. (2010). Minimum Hellinger distance estimation in a two-sample semiparametric model. *Journal of Multivariate Analysis*, **101**, 1102-1122.

[39] Wu, J.,Yao, W. and Xiang, S. (2017). Computation of an efficient and robust estimator in a semiparametric mixture model. *Journal of Statistical Computation and Simulation*, **87**, 2128-2137.

[40] Xiang, S. (2014). *Semiparametric mixture models*. PhD thesis, Kansas State University.

[41] Xiang, S., Yao, W. and Wu, J. (2014). Minimum profile Hellinger distance estimation for a semiparametric mixture model. *Canadian Journal of Statistics*, **42**, 246-267.

[42] Xiang, L., Yau, K.K., Hui, Y.V. and Lee, A.H. (2008). Minimum Hellinger distance estimation for k-component poisson mixture with random effects. *Biometrics*, **64**, 508-518.

[43] Yang, S. (1991). Minimum Hellinger distance estimation of parameter in the random censorship model. *The Annals of Statistics*, **19**, 579-602.

[44] Ying, Z. (1992). Minimum Hellinger-type distance estimation for censored data. *The Annals of Statistics*, **20**, 1361-1390.

[45] Zhu, Y., Wu, J. and Lu, X. (2013). Minimum Hellinger distance estimation for a two-sample semiparametric cure rate model with censored survival data. *Computational Statistics*, **28**, 2495- 2518.

[46] Zou, F., Fine, J.P. and Yandell, B.S. (2002). On empirical likelihood for a semiparametric mixture model. *Biometrika*, **89**, 61-75.