UNIVERSITY OF CALGARY

Statistical Inferences for Two-Component Semiparametric Location-Scale Mixture Models

by

Na Zhang

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE

DEGREE OF DOCTOR OF PHILOSOPHY

GRADUATE PROGRAM IN MATHEMATICS AND STATISTICS

CALGARY, ALBERTA

NOVEMBER, 2024

Abstract

Mixture models serve as a powerful statistical tool, particularly in capturing heterogeneous populations by representing them as a mixture of several distributions. These models are particularly useful in various fields, including genomics, economics, and social sciences, where data often arises from a combination of distinct subpopulations. Among the various mixture models, the two-component location-scale mixture model is of special interest due to its simplicity and flexibility in modeling diverse data structures.

Traditional methods of parameter estimation in mixture models, such as Maximum Likelihood Estimation (MLE), are widely used due to their desirable asymptotic properties such as asymptotic efficiency. However, MLE can be highly sensitive to model misspecification and the presence of outliers, often leading to biased or inefficient estimates. Recognizing these limitations, this thesis explores the use of Minimum Hellinger Distance Estimation (MHDE), a robust alternative estimation which offers a balance between efficiency and robustness, meaning that while MHDE may not be as efficient as MLE in perfectly specified models (i.e., when the model exactly fits the data), it remains sufficiently efficient while being far more robust to data that does not perfectly align with the model assumptions. The choice of MHDE is motivated by its robustness in the face of outliers and its ability to provide more reliable estimates when the underlying distribution deviates from the assumed model. Focusing on semiparametric mixtures introduces additional flexibility by allowing for an unspecified distribution, which enables the model to capture complex data structures without imposing strict parametric assumptions. In these semiparametric models, the emphasis is placed on estimating the unknown parameter vector, while the form of the mixing distribution remains unspecified. This approach strikes a balance between parametric precision and nonparametric flexibility, making it particularly useful in situations where the true distribution is unknown or deviates from common parametric forms.

This thesis mainly focuses on three primary objectives, which include minimum Hellinger

distance estimation for both parametric and semiparametric location-scale mixture models, and estimation for location-scale mixture when data are right-censored.

The first objective focuses on the estimation of the unknown parameters using minimum Hellinger distance, with a particular emphasis on the parametric location-scale mixtures. In this thesis, the Parametric Hellinger Distance Estimation (MHDEP) method is explored in depth. The Hellinger distance is defined in terms of the Hellinger integral, which was introduced by Ernst Hellinger in 1909. Chapter 2 delves into the methodology, theoretical asymptotic normality properties, simulation studies and real data analysis of MHDEP. This approach is chosen for its advanced robustness and competitive efficiency to be compared with the classical parametric likelihood-based estimations, especially in scenarios where traditional estimation methods may perform bad due to model misspecification or data irregularities.

The second objective focuses on Semiparametric Hellinger Distance Estimation (SEMIMHDE) for mixture models with unknown component distributions. While the full derivation of identifiability conditions is ongoing, identifiability remains crucial for ensuring reliable parameter estimation. It ensures that the parameters of interest, such as mixing proportions, location, and scale parameters can be uniquely determined from the data. We first constructed the SEMIMHDE by deriving a custom base function for each component in the semiparametric mixture model. Subsequently, we adapted Hellinger Minimization for Mixtures (HMIX) algorithm which is originally designed for parametric mixtures using MHDE, to accommodate the semiparametric setting. This modified HMIX algorithm allows the estimation of unknown component distributions. To assess the performance of SEMIMHDE, we conducted a series of simulation studies, evaluating its efficiency, robustness, and sensitivity to model misspecification compared to other parametric estimation methods. Finally, we applied SEMIMHDE to the Old Faithful Geyser dataset, demonstrating its practical applicability and illustrating how it can handle real-world data.

The third objective aims to refine and advance the current methodologies, specifically Kaplan-Meier-weighted MHDEP and SEMIMHDE for right-censored mixture is constructed, to provide a robust and comprehensive toolkit to analyze finite location-scale mixture models when data are right-censored. Right-censoring is a common issue in many practical applications, such as survival analysis, where the complete data for some observations is not available. Chapter 4 focuses on applying MHDEP and SEMIMHDE to various censoring rate scenarios from low to high rate to evaluate their finite sample performance. Simulation results show that both methods maintain good finite-sample performance even at high levels of censoring, demonstrating their robustness and reliability under different degrees of right-censoring. This study is anticipated to provide more reliable and versatile solutions for complex statistical modeling challenges, broadening the applicability of these methods to a wider range of practical problems. Additionally, both estimation methods were applied to a real right-censored dataset to assess their performance in scenarios where complete observations are unavailable.

In summary, this thesis makes significant contributions to the field of both parametric and semiparametric mixture models by addressing fundamental issues of efficiency, robustness and model misspecification using Minimum Hellinger distance estimations. The proposed MHDEP and SEMIMHDE not only advance the theoretical properties of these models but also provide practical tools for more accurate and reliable statistical analysis in various applied settings.

Acknowledgments

I am deeply grateful to everyone who has been a part of my PhD journey, offering essential support and guidance throughout this experience. Their contributions have played a crucial role in my development, both academically and personally, and I want to express my heartfelt gratitude for their unwavering encouragement and assistance.

Foremost, I would like to express my deepest gratitude to my supervisors. Dr. Jingjing Wu provided me with the inspiration for this work and started me on the path of my PhD journey. Dr. Antony Ware and Dr. Xuewen Lu have given me invaluable guidance and unwavering support in the latter part of my PhD journey. Their profound knowledge of statistics and mathematics has greatly enhanced my understanding and driven me to explore my research more deeply. Their patience in navigating the difficulties I encountered, combined with their encouragement to persevere, has been invaluable to my growth as a researcher. Beyond the technical aspects of my work, they offered professional advice that shaped my approach to writing and presenting my research, ensuring that my ideas were communicated effectively. Their encouragement to explore new ideas has been instrumental in the completion of this thesis. Under their supervision, I have not only developed skills in statistical analysis but also learned to approach problems with a more holistic and rigorous mindset. This journey has also been a significant period of personal development. The challenges and triumphs I experienced during my PhD have taught me the value of perseverance, adaptability, and continuous learning. I have gained a greater appreciation for collaboration, humility, and the importance of balancing hard work with self-care. My supervisors' support played a crucial role in fostering this growth, providing me not only with the tools to succeed academically but also with the confidence to tackle future challenges. I am truly fortunate to have had such exceptional supervisors.

I also wish to thank my PhD supervisory committee members, Dr. Gemai Chen and Dr. Mina Aminghafari, for their constructive feedback and continuous support. Additionally, I am grateful to my thesis defense committee members, Dr. Yongcheng Qi and Dr. Yunqi Ji, for their time and valuable input during the final stages of my research.

A special thank you to my parents, whose unwavering support and encouragement have been my constant source of strength throughout this journey.

Finally, I would like to appreciate the Department of Mathematics and Statistics at University of Calgary for providing financial support like teaching assistant opportunities and excellent research facilities such as computation labs, which have been essential to my study during PhD program. The peaceful environment of Calgary, with nearby beautiful parks like Kananaskis provincial park and Banff national park, offered me a tranquil space to reflect on my research challenges.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

In this thesis, we focus on statistical inferences for two-component semiparametric location-scale mixture models using the minimum Hellinger distance estimation (MHDE). This chapter provides an introduction to mixture models, the proposed estimation method MHDE, and outlines the research contributions presented in the subsequent three chapters. Identifiability for the semiparametric mixture is discussed at the end of this chapter.

1.1 Introduction

Mixture models have been utilized for over 150 years as an intuitively simple and practical tool for enriching the collection of probability distributions available for data modeling. They can be applied to a wide range of data types, including univariate or multivariate, continuous or categorical, cross-sectional, time series, and networks. Mixture analysis remains a highly active research topic in both statistics and machine learning, with continuous advancements in methodology and applications.

A mixture model is a probabilistic model that represents the presence of subpopulations within an overall population, without identifying the subpopulation to which an individual observation belongs. The m-component mixture model has the probability density function (p.d.f.)

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

where f_j is the j-th component density and $(\pi_1, \dots, \pi_m)$ are the mixture proportions with $\pi_j > 0$, $\sum_{j=1}^{m} \pi_j = 1$. When m is unknown, there are various articles that discuss the selection of m, such as Roeder (1994), Chen et al. (2001), and Kasahara and Shimotsu (2015). In this thesis, we assume that m is fixed and known to be two.

Depending on the form of the mixture components, mixture models can be categorized as parametric, nonparametric, or semiparametric. In the nonparametric framework, without the availability of training data, specific methods have been developed to estimate the mixture proportions. Huang et al. (2013) introduces a nonparametric finite mixture of regression models for analyzing US house price index data, addressing identifiability, proposing a kernel regression-based estimation procedure with asymptotic normality, and demonstrating its effectiveness through a modified EM algorithm, Monte Carlo simulations, and empirical analysis. Hall and Zhou (2003), by assuming that vectors of observations are conditionally independent but may have different marginal distributions, proposed a minimum distance estimator based on weighted-bootstrap estimation. Subsequently, Benaglia et al. (2009) proposed an EM-like algorithm extended from the work of Bordes et al. (2006). Other approaches to nonparametric estimation in mixture models rely on training samples observed directly from the components. For further reading on these approaches, interested readers are referred to Du Roy de Chaumaray and Marbac (2024), and Wu and Karunamuni (2009).

When the component densities f_j are assumed to belong to parametric families, the literature on parametric mixture models is extensive. Seminal works in this field include the monographs by Newton et al. (1986), McLachlan and Basford (1988), and McLachlan and Peel (2000). Several types of estimators have been proposed for these models, including maximum likelihood estimators (MLE) as discussed by Dick and Bowden (1973) and Redner and Walker (1984), minimum chi-square estimators as discussed in Day (1969), the method of moments in Lindsay and Basak (1993), Bayesian approaches in Roeder (1994), and techniques based on moment generating functions in Quandt and Ramsey (1978).

In recent years, identifiability has been established for specific finite parametric mixture models. Otiniano et al. (2015) demonstrated it for mixtures of skew-normal and skew-t distributions. Additionally, Aragam et al. (2020) established general conditions under which families of nonparametric mixture models are identifiable. They introduced a novel framework involving clustering over-fitted parametric (i.e., misspecified) mixture models, thereby advancing the theoretical understanding of mixture model identifiability.

When little is known about sub-populations, investigating parametric families becomes challenging, leading to a surge in research on semiparametric mixture models over the last decade. Xiang et al. (2019) provided a comprehensive overview of newly developed semiparametric mixture models. Pu and Arias-Castro (2020) introduced a semiparametric EM (SEM) algorithm for fitting mixture models, assuming symmetric and log-concave mixture components. Bordes et al. (2006) studied similar models where the second component is known to be symmetric. Helali et al. (2022) proposes a shrinkage method using Bernstein polynomials and a finite Gaussian mixture model to construct a semi-parametric density estimator that addresses boundary issues in density estimation, with demonstrated asymptotic properties and performance evaluation through simulations and real data. Xiang et al. (2014) introduces a new minimum profile Hellinger distance (MPHD) estimator for semiparametric mixture models, where one component has known distribution with possibly unknown parameters while the other component density and the mixing proportion are unknown. Wu and Zhou (2018) proposed a minimum Hellinger distance estimation (MHDE) for a two-component semiparametric mixture model with both of the two component unknown location-shifted symmetric distributions.

In recent years, while location-shifted mixtures have been extensively studied, there has been growing interest in location-scale mixture models. Hennig (2004) examines the breakdown points of maximum likelihood estimators for location-scale mixtures, showing that while alternatives like t-distribution mixtures offer some robustness, they do not significantly outperform normal mixtures in handling outliers, and suggests using improper uniform distributions to enhance robustness. Maleki and Mahmoudi (2017) present a study on maximum likelihood estimation for two-piece distributions derived from scale mixtures of normal distributions, using an EM-type algorithm to address both symmetric and asymmetric data modeling. Xiang et al. (2016) introduced a new estimation procedure for semiparametric mixture models where components are location shifts of symmetric distributions, themselves continuous normal scale mixtures.

1.2 Mixture Models

Given the focus of this thesis on the semiparametric location-scale mixture model, which extends naturally from the semiparametric location-shifted mixture model, methodologies and techniques developed for the latter can be used for the former. Accordingly, this section will conduct a comprehensive review of pertinent literature encompassing both location-shifted and location-scale mixture models.

1.2.1 Semiparametric Location-Shifted Mixture Models

Consider a semiparametric location-shifted mixture model defined as

$$h(x) = \sum_{j=1}^{m} \pi_j f(x - \mu_j), \tag{1.1}$$

where f represents an unknown even probability density function, $\mu_1, \ldots, \mu_m \in \mathbb{R}$ are the unknown location parameters, and the mixing proportions π_j satisfy $\pi_j \in (0,1)$ and $\sum_{j=1}^{m} \pi_j = 1$. When $m = 2$, this simplifies to the two-component semiparametric location-shifted mixture model:

$$h(x) = \pi_1 f(x - \mu_1) + \pi_2 f(x - \mu_2), \tag{1.2}$$

where $\theta = (\pi_1, \pi_2, \mu_1, \mu_2)^\top$ is the parameter vector of interest, π_1 and π_2 are two mixing proportions and μ_1 and μ_2 are two location parameters. The unknown density f is considered a nuisance parameter, complicating the estimation of θ. Bordes et al. (2006) established the identifiability of model (1.1) for $m = 2$, and Hunter et al. (2007) established the identifiability of model (1.1) for both $m = 2$ and $m = 3$. Following Hunter et al. (2007), model (1.2) is identifiable under $\theta \in \Theta_0$, where

$$\Theta_0 = \left\{ (\pi_1, \pi_2, \mu_1, \mu_2)^\top : \pi_1, \pi_2 \in (0, 0.5) \cup (0.5, 1), \pi_1 + \pi_2 = 1, \mu_1 < \mu_2, \mu_1, \mu_2 \in \mathbb{R} \right\}.$$

Various methods have been proposed to estimate the parameter θ in model (1.2). Hunter et al. (2007) introduced a generalized Hodges-Lehmann estimator. Bordes et al. (2006) employed a stochastic EM algorithm, while Benaglia et al. (2009) improved upon this with a deterministic algorithm that offers flexibility and applicability across multiple mixture components. Bordes et al. (2006) developed a distance-based estimator using a linear bounded operator and a contrast function, whereas Butucea and Vandekerkhove (2014) introduced a class of M-estimators based on Fourier analysis. Recently, the Minimum Hellinger Distance (MHD) method has gained popularity in semiparametric mixture models. Xiang et al. (2014) applied the MHDE to a mixture model with one specified component and unknown parameters. Wu et al. (2017) and Wu and Zhou (2018) extended this to MPHDE and MHDE for the two-component location-shifted mixture model (1.2). These approaches demonstrate that both MPHDE and MHDE offer competitive efficiency and superior robustness compared to

MLE.

1.2.2 Semiparametric Location-Scale Mixture Models

The k-component univariate location-scale mixture model is defined as:

$$h(x) \ = \ \sum_{i=1}^{k} \pi_i \frac{1}{\sigma_i} f(\frac{x-\mu_i}{\sigma_i}),$$

where, in addition to the previously explained parameters, $\sigma_1, \ldots, \sigma_k > 0$ are the scale parameters. Throughout this thesis we will focus on a two-component location-scale mixture model. Since the identifiability results in Hunter et al. (2007) for semiparametric two-component location-shifted mixtures require the component distribution f to be symmetric about zero, we similarly assume that f is symmetric about zero in the context of our two-component location-scale mixture model. The semiparametric two-component location-scale mixture model is a convolution of the unknown base function f with the unknown two-dimensional two-point mass function with mass π_1 and π_2 at points (μ_1, σ_1) and (μ_2, σ_2) respectively. In this thesis, we consider the case of $k = 2$, i.e., the class of two-component location-scale mixture models:

$$\mathcal{H} = \left\{ h \ = \ h_{\theta, f}(x) \ = \ \pi_1 \frac{1}{\sigma_1} f(\frac{x-\mu_1}{\sigma_1}) + \pi_2 \frac{1}{\sigma_2} f(\frac{x-\mu_2}{\sigma_2}) : \theta \in \Theta \right\}, \tag{1.3}$$

where $\theta = (\pi_1, \pi_2, \mu_1, \mu_2, \sigma_1, \sigma_2)^\top$, f is a probability density function with standard deviation 1 and symmetric about zero, and

$$\begin{aligned} \Theta = \Big\{ (\pi_1, \pi_2, \mu_1, \mu_2, \sigma_1, \sigma_2)^\top \in \mathbb{R}^6 : & \pi_1, \pi_2 \in (0, 0.5) \cup (0.5, 1), \quad \pi_1 + \pi_2 = 1, \\ & \mu_1 < \mu_2, \quad \sigma_1 \neq \sigma_2, \quad \sigma_1 > 0, \quad \sigma_2 > 0 \Big\}. \end{aligned} \tag{1.4}$$

This thesis focuses on model (1.3), where both the five-dimensional parameter θ and the infinite-dimensional function f are unknown and thus it is a semiparametric mixture model. In literature, scale mixture models are mixtures of scaled subpopulations with the mixing distribution often assumed unknown, and thus frequently fall in the class of semiparametric mixture models. Basu (1996) discussed the existence of a normal scale mixture with fixed variance and a percentile. With each component distribution known up to some location and scale parameters, Chen et al. (2017) investigated strong consistency of the maximum likelihood estimator (MLE) for finite mixtures with unknown point mass mixing distribution. Xiang et al. (2016) proposed a new estimation procedure for a class of semiparametric mixture models with components being location shifts of a symmetric distribution which itself is a continuous normal scale mixture. In all above works, the components are assumed known (usually normal) up to some unknown parameters (parametric part) while the mixing distribution is assumed unknown (nonparametric part). In this proposed research, we explore a different class of semiparametric mixture models, where the unknown distribution function is treated as a nuisance parameter. Our primary focus is on estimating five key parameters: the mixing proportion, two location parameters, and two scale parameters. This class of models is naturally called two-component semiparametric location-scale mixture models. To the best of my knowledge, almost all literature on location-scale mixture model (1.3) assume that f is known (e.g., Hennig (2004), Zitouni et al. (2018), Xiang et al. (2016)) and no work has been done for model (1.3) with unknown f for each component distribution. The reason for this gap is possibly due to the fact that each component distribution is unknown in the location-scale mixture model makes it difficult to even investigate its model identifiability. However, similar to the location-shifted model (1.2), we believe that the symmetry of f somehow guarantees the identifiability for a smaller parameter space than Θ in (1.4). Thus in this thesis, we will investigate the identifiability of model (1.3) for both parametric model and semiparametric model, construct efficient and robust estimators of the unknown parameters for both parametric and semiparametric location-scale mixture models, test estimation efficiency and robustness performance, and develop efficient and robust estimations based on location-scale mixture models when data are right-censored.

1.3 Estimations Based on Hellinger Distance

1.3.1 Minimum Hellinger distance estimation (MHDE)

In probability and statistics, the Hellinger distance is used to quantify the similarity between two probability distributions. It is a type of f-divergence. The Hellinger distance is defined in terms of the Hellinger integral, which was introduced by Ernst David Hellinger in 1909. Beran (1977) introduced the MHDE for parametric models in general. Suppose we observe i.i.d. 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 probability density functions (p.d.f.s) is defined as the L_2-distance between their respective square root densities. Specifically, the Minimum Hellinger Distance Estimator (MHDE) $\hat{\theta}$ of the unknown parameter θ is defined as:

$$
\begin{aligned}
\hat{\theta} &= \arg\min_{\theta\in\Theta} \left\| f_\theta^{1/2} - \hat{f}^{1/2} \right\|_{L_2}^2 \\
&= \arg\min_{\theta\in\Theta} \left(\frac{1}{2} \int \left(f_\theta^{1/2}(x) - \hat{f}^{1/2}(x) \right)^2 dx \right) \\
&= \arg\min_{\theta\in\Theta} \left(1 - \int \sqrt{f_\theta(x)\hat{f}(x)} dx \right).
\end{aligned}
$$

where $\|\cdot\|_{L_2}$ denotes the L_2-norm, f_θ is the model p.d.f. depending on the parameter θ, and $\hat{f}$ is an appropriate nonparametric empirical (e.g. kernel) density estimator based on the sample, and the Hellinger distance lies in interval [0,1]. Beran (1977) proved that the MHDE 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, however, MHDE for semiparametric models has not been well investigated until works by Wu and Karunamuni (2012).

Suppose we observe i.i.d. r.v.s $X_1, \ldots, X_n$ with p.d.f. $g_0 = f_{\theta,\eta}$ being a member of the general semiparametric model family $\mathcal{F} = \{f_{\theta,\eta} : \theta \in \Theta, \eta \in \mathcal{H}\}$, where Θ is a compact subset of $\mathbb{R}^p$ and $\mathcal{H}$ is an arbitrary set of infinite dimension. We assume $\mathcal{F}$ is identifiable in the sense that $\|f^{1/2}_{\theta_1,h_1} - f^{1/2}_{\theta_2,h_2}\| = 0$ implies $\theta_1 = \theta_2$ and $h_1 = h_2$. When an estimator $\hat{\eta}$ of η is available, Wu and Karunamuni (2012) proposed a plug-in MHDE of θ given by

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

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

1.3.2 Minimum Profile Hellinger distance estimation (MPHDE)

In the construction of MHDE for semiparametric models $f_{\theta,\eta}$, an appropriate estimator of the unknown nuisance parameter η is required which may be based on the current sample or other resource. If such an estimator of the each component density is not available, Wu and Karunamuni (2012) introduced the first time the MPHDE. The MPHDE is obtained by first profiling out the infinite-dimensional nuisance parameter of f and then minimizing the profiled Hellinger distance.

Consider the semiparametric model family $\mathcal{F} = \{f_{\theta,\eta} : \theta \in \Theta, \eta \in \mathcal{H}\}$ which is assumed identifiable. Assume that $X_1, \ldots, X_n$ are i.i.d. r.v.s with p.d.f. $f = f_{\theta,\eta} \in \mathcal{F}$. Suppose $\hat{f}$ is a nonparametric estimator of f based on the observed data. Then, the MPHDE of θ is given by

$$\hat{\theta} = \arg\min_{\theta \in \Theta} \left\{ \min_{h \in \mathcal{H}} \left\| f_{\theta,h}^{1/2} - \hat{f}^{1/2} \right\| \right\}. \tag{1.5}$$

Note that the arguments in above equation contain both parametric part θ and nonparametric part h. This optimization can be split into two steps via profiling. First, for each θ in Θ let

$$h_\theta(\hat{f}) = \arg\min_{h \in \mathcal{H}} \left\| f_{\theta,h}^{1/2} - \hat{f}^{1/2} \right\|,$$

and define the profiled Hellinger distance function

$$D(\theta) = \min_{h \in \mathcal{H}} \left\| f_{\theta,h}^{1/2} - \hat{f}^{1/2} \right\| = \left\| f_{\theta,h_\theta(\hat{f})}{}^{1/2} - \hat{f}^{1/2} \right\|.$$

Then MPHDE given in (1.5) is $\hat{\theta} = \arg\min_{\theta \in \Theta} D(\theta)$. Wu and Karunamuni (2012) have proved the efficiency and robustness of the MPHDE given in (1.5) for semiparametric models of general form. Wu and Karunamuni (2012) applied the MPHDE to a two-component location-shifted mixture model and developed an algorithm to calculate it. Their results showed that the MPHDE is very competitive when the components are normal and much better when the components are non-normal than some existing methods such as the method in Benaglia et al. (2009) and MLE.

1.4 Identifiability

It is well known that mixture models are generally non-identifiable. Thus before we explore methods to estimate the unknown parameters, it is necessary and important to discuss the identifiability of model (1.3). It does not make sense to estimate the parameters when they are actually not identifiable.

To the best of my knowledge, the majority of existing literatures on location-scale mixture models, specifically model (1.3), presumes that the density function f is known and predefined. This assumption is evident in several key studies, such as those by Hennig (2004), Zitouni et al. (2018), Xiang et al. (2016)). However, there appears to be a notable absence of research addressing the scenario where f is unknown.

The reason for this gap in the literature likely stems from the inherent complexity and challenges associated with dealing with an unknown density function within the context of location-scale mixture models. When the density function f is unknown, it introduces significant difficulties in determining the identifiability of the model. Identifiability refers to the ability to uniquely determine the model parameters from the observed data. In the case of location-scale mixture models, the inclusion of scale parameters further complicates the situation. Scale parameters add an additional layer of flexibility to the model, allowing it to accommodate a wider range of data characteristics and variability. This added flexibility makes the model more adaptable, but it also increases the complexity of investigating its identifiability. Specifically, the unknown density function f combined with the scale parameters introduces more degrees of freedom, making it challenging to disentangle the effects of these parameters and accurately determine the underlying structure of the model. The interplay between the location and scale parameters and the unknown density function means that there are more potential sources of variability and interaction that need to be accounted for in the model.

In univariate cases, all discussions on identifiability are for mixture models with symmetric components. For model (1.2),Bordes et al. (2006) and Hunter et al. (2007) established the identifiability for $m = 2$ and for both $m = 2$ and $m = 3$ respectively. For model (1.3) when f is unknown, to my knowledge, there is no clear discussion in literature on its identifiability. For model (1.3) when distribution of f is known, the identifiability has been well studied in literature such as gaussian distribution. Thus we propose to investigate both the sufficient and necessary conditions for model (1.3) being identifiable for semiparametric case. For model (1.3) being identifiable, by definition we need to prove that whenever $\pi_1 \frac{1}{\sigma_1} f(\frac{x-\mu_1}{\sigma_1}) + \pi_2 \frac{1}{\sigma_2} f(\frac{x-\mu_2}{\sigma_2}) = \tilde{\pi}_1 \frac{1}{\tilde{\sigma}_1} \tilde{f}(\frac{x-\tilde{\mu}_1}{\tilde{\sigma}_1}) + \tilde{\pi}_2 \frac{1}{\tilde{\sigma}_2} \tilde{f}(\frac{x-\tilde{\mu}_2}{\tilde{\sigma}_2})$ with $(\pi_1, \pi_2, \mu_1, \mu_2, \sigma_1, \sigma_2)^\top, (\tilde{\pi}_1, \tilde{\pi}_2, \tilde{\mu}_1, \tilde{\mu}_2, \tilde{\sigma}_1, \tilde{\sigma}_2)^\top \in \Theta$ and $f, \tilde{f} \in \mathcal{F}$, it must hold that $(\pi_1, \pi_2, \mu_1, \mu_2, \sigma_1, \sigma_2)^\top = (\tilde{\pi}_1, \tilde{\pi}_2, \tilde{\mu}_1, \tilde{\mu}_2, \tilde{\sigma}_1, \tilde{\sigma}_2)^\top$ and $f = \tilde{f}$.

For this purpose, we can first review the derivation of identifiability conditions for the simpler two-component location-shifted mixture model (1.2). Then we will try to check whether we can extend the techniques used to the location-scale mixture model (1.3) under our consideration. Both Bordes et al. (2006) and Hunter et al. (2007) established the identifiability of model (1.2) but used different methods. The former proved directly the identifiability with use of techniques such as characteristic function and linear independence of sine functions, while the latter proved somehow more general results for k-component with $k \geq 1$. Hunter et al. (2007) treated the k-component mixture as a convolution of a zero-symmetric p.d.f. with a mixture of point mass distribution functions at μ_i's. Hunter et al. (2007) first derived the sufficient and necessary conditions for the k-component mixture of point mass distribution functions, and then immediately proved the identifiability of the special two-component location-shifted mixture model (1.2). Following Bordes et al. (2006), some derivations for identifiability of the two-component location-scale mixture model (1.3) are given in Appendix A.

Although we have not yet determined the complete set of sufficient and necessary conditions for the identifiability of the semiparametric finite location-scale mixture model, our research will focus on a more focused parameter space (1.4) for investigation. This approach involves narrowing our analysis to a specific subset of the parameter space where identifiability is more likely to be achievable. We have addressed and excluded special cases that are known to result in non-identifiability as followed by equation (1.4), which can be treated as necessary identifiability conditions. These problematic scenarios often arise due to certain configurations of the parameters that lead to ambiguities or redundancies in the model, making it impossible to uniquely determine the parameters from the available data. By excluding these cases, we refine our focus to those situations where the model's identifiability is more tractable, thereby providing a more practical and insightful framework for analyzing the model's parameters.

1.5 Research Contributions

The proposed research encompasses four primary objectives, each integral to the advancement of our understanding and methodology in semiparametric location-scale mixture models.

The first objective is to establish the sufficient and necessary conditions for the identifiability of the two-component semiparametric location-scale mixture model (1.3). Identifiability is a fundamental prerequisite in statistical modeling, ensuring that the model parameters can be uniquely determined from the data. Without this assurance, subsequent analysis and inference may be flawed. Therefore, determining these conditions is essential before progressing with the proposed research, as it lays the groundwork for reliable and valid estimations.

The second and third objectives focus on the estimation of the unknown parameters using minimum distance techniques under parametric or semiparametric location-cale mixture model, with a particular emphasis on the Hellinger distance. The goal is to achieve robustness and efficiency in our estimations. The parametric Hellinger distance estimation (MHDEP) and semiparametric Hellinger distance estimation (SEMIMHDE) will be the primary methods explored. These methods are chosen for their ability to provide resilient and accurate parameter estimates, even in the presence of model misspecification or data irregularities. Chapter 2 will be dedicated to MHDEP, detailing its methodology, theoretical properties, and applications. Chapter 3 will focus on SEMIMHDE, presenting a comprehensive examination of its development, implementation, and advantages over traditional methods.

The final objective aims to refine and advance the current methodologies, specifically MHDEP and SEMIMHDE, to provide a more robust and comprehensive toolkit for analyzing finite location-scale mixture models, particularly in the context of right-censored data. Right-censoring is a common issue in many practical applications, and addressing it effectively requires sophisticated techniques. By enhancing MHDEP and SEMIMHDE, this research seeks to extend their utility and improve their performance in dealing with censored data, thereby offering more reliable and versatile solutions for complex statistical modeling challenges.

In summary, this research endeavors to not only establish minimum Hellinger distance estimation for semiparametric mixture models but also to develop and validate practical estimation techniques that can be widely applied in various fields requiring robust statistical analysis.

1.6 Appendix A

In this appendix, we use similar techniques as in Bordes et al. (2006) to study the identifiability of the two-component location-scale mixture model (1.3).

Assume a two-component location-scale mixture model has two representations $\pi_1 f\left(\frac{x-\mu_1}{\sigma_1}\right)+\pi_2 f\left(\frac{x-\mu_2}{\sigma_2}\right)$ and $\pi_1' f'\left(\frac{x-\mu_1'}{\sigma_1'}\right)+\pi_2' f'\left(\frac{x-\mu_2'}{\sigma_2'}\right)$, where $f, f' \in \mathcal{F}$. Let F and F' denote the cumulative distribution functions (c.d.f.s) of f and f', respectively. Then,

$\pi_1 F(\frac{x-\mu_1}{\sigma_1})+(1-\pi_1)F(\frac{x-\mu_2}{\sigma_2}) = \pi_1' F'(\frac{x-\mu_1'}{\sigma_1'})+(1-\pi_1')F'(\frac{x-\mu_2'}{\sigma_2'}), \quad x \in \mathbb{R}$. Let Φ and Φ' denote the characteristic functions of f and f', respectively. Then the characteristic functions of the two mixture models satisfy $\pi_1 \exp(it\mu_1)\Phi(\sigma_1 t) + \pi_2 \exp(it\mu_2)\Phi(\sigma_2 t) = \pi_1' \exp(it\mu_1')\Phi'(\sigma_1' t) + \pi_2' \exp(it\mu_2')\Phi'(\sigma_2' t)$, denoted as $A(t) = B(t)$, say, for notation simplicity. Note that both Φ and Φ' are real functions symmetric about zero by the symmetry of $f, f' \in \mathcal{F}$. $\overline{A(t)}$ denotes the complex conjugate of the function $A(t)$, replacing t with $-t$ in above equality gives $\overline{A(t)} = \overline{B(t)}$. As a result, $A(t)\overline{B(t)} = \overline{A(t)}B(t) = \overline{A(t)\overline{B(t)}}$, implying the imaginary part of $\overline{A(t)}B(t)$ is zero, i.e.

$$\begin{aligned}&\pi_1\pi_1' \sin\left[(\mu_1' - \mu_1)t\right] \Phi(\sigma_1 t)\Phi'(\sigma_1' t) + \pi_1\pi_2' \sin\left[(\mu_2' - \mu_1)t\right] \Phi(\sigma_1 t)\Phi'(\sigma_2' t)\\ &+ \pi_2\pi_1' \sin\left[(\mu_1' - \mu_2)t\right] \Phi(\sigma_2 t)\Phi'(\sigma_1' t) + \pi_2\pi_2' \sin\left[(\mu_2' - \mu_2)t\right] \Phi(\sigma_2 t)\Phi'(\sigma_2' t) = 0.\end{aligned} \tag{1.6}$$

Let $\alpha_1 = \mu_1' - \mu_1$, $\alpha_2 = \mu_2' - \mu_2$, $\eta = \mu_2 - \mu_1$, $\beta_1 = \pi_1\pi_1'\Phi(\sigma_1 t)\Phi(\sigma_1' t)$, $\beta_2 = \pi_1\pi_2'\Phi(\sigma_1 t)\Phi'(\sigma_2' t)$, $\beta_3 = \pi_2\pi_1'\Phi(\sigma_2 t)\Phi'(\sigma_1' t)$ and $\beta_4 = \pi_2\pi_2'\Phi(\sigma_2 t)\Phi'(\sigma_2' t)$, then (1.6) is reduced to $\beta_1 \sin[\alpha_1 t] + \beta_2 \sin\left[(\alpha_2 + \eta)t\right] + \beta_3 \sin\left[(\alpha_1 - \eta)t\right] + \beta_4 \sin[\alpha_2 t] = 0, \quad x \in \mathbb{R}$.

1. Assume $\alpha_1 = \alpha_2 = 0$. Then (1.6) is reduced to $(\beta_2 - \beta_3)\sin[\eta t] = 0$ for all $t \in \mathbb{R}$. Since $\eta \neq 0$, we have $\pi_1\pi_2'\Phi(\sigma_1 t)\Phi'(\sigma_2' t) - \pi_2\pi_1'\Phi(\sigma_2 t)\Phi'(\sigma_1' t) = 0$ for any $t \in \mathbb{R}\backslash\{k\pi/\eta : k \in \mathbb{N}\}$. By the fact that all characteristic functions are uniformly continuous, we have

$$\pi_1\pi_2'\Phi(\sigma_1 t)\Phi'(\sigma_2' t) - \pi_2\pi_1'\Phi(\sigma_2 t)\Phi'(\sigma_1' t) = 0, \quad t \in \mathbb{R}. \tag{1.7}$$

Since $\Phi(0) = \Phi'(0) = 1$, plugging $t = 0$ into (1.7) gives $\pi_1\pi_2' - \pi_2\pi_1' = 0$ and further $\pi_i = \pi_i'$, $i = 1, 2$. By $\pi_1, \pi_2 \neq 0$, (1.7) is further reduced to

$$\Phi(\sigma_1 t)\Phi'(\sigma_2' t) = \Phi(\sigma_2 t)\Phi'(\sigma_1' t), \quad t \in \mathbb{R}. \tag{1.8}$$

Define r.v.s $X \sim f$ and independently $Y \sim f'$. Then the left and right sides of (1.8) are the characteristic functions of $Z_1 := \sigma_1 X + \sigma_2' Y$ and $Z_2 := \sigma_2 X + \sigma_1' Y$ respectively. By (1.8), Z_1 and Z_2 follow exactly the same p.d.f.. Since Z_1 and Z_2 have the same p.d.f, they have the same expected value and the same second moment ($\mathbb{E}[Z_1^2] = \mathbb{E}[Z_2^2]$). Therefore:

$$\mathbb{V}ar(Z_1) = \mathbb{E}[Z_1^2] - (\mathbb{E}[Z_1])^2 = \mathbb{E}[Z_2^2] - (\mathbb{E}[Z_2])^2 = \mathbb{V}ar(Z_2).$$

To find $\mathbb{V}ar(Z_1)$, we use the formula for the variance of a linear combination of independent random variables:

$$\mathbb{V}ar(Z_1) = \mathbb{V}ar(\sigma_1 X + \sigma_2' Y).$$

Since X and Y are independent, the variance of their linear combination is:

$$\mathbb{V}ar(Z_1) = \sigma_1^2 \mathbb{V}ar(X) + (\sigma_2')^2 \mathbb{V}ar(Y).$$

Assuming that X and Y are identically distributed (meaning $\mathbb{V}ar(X) = \mathbb{V}ar(Y)$), let σ^2 be their common variance:

$$\mathbb{V}ar(X) = \mathbb{V}ar(Y) = \sigma^2.$$

Thus:

$$\mathbb{V}ar(Z_1) = \sigma_1^2 \sigma^2 + (\sigma_2')^2 \sigma^2 = (\sigma_1^2 + (\sigma_2')^2)\sigma^2.$$

Similarly, for Z_2:

$$\mathbb{V}ar(Z_2) = \mathbb{V}ar(\sigma_2 X + \sigma_1' Y).$$

Again, using the independence of X and Y:

$$\mathbb{V}ar(Z_2) = \sigma_2^2 \mathbb{V}ar(X) + (\sigma_1')^2 \mathbb{V}ar(Y).$$

Substitute the common variance σ^2:

$$\mathbb{V}ar(Z_2) = \sigma_2^2 \sigma^2 + (\sigma_1')^2 \sigma^2 = (\sigma_2^2 + (\sigma_1')^2)\sigma^2.$$

Then $\mathbb{V}ar(Z_1) = \sigma_1^2 + (\sigma_2')^2 = \sigma_2^2 + (\sigma_1')^2 = \mathbb{V}ar(Z_2)$.

Let $\gamma = \sigma_2 - \sigma_1$ and $\gamma' = \sigma_2' - \sigma_1'$. The statement we have proved, "$\mathbb{V}ar(Z_1) = \sigma_1^2 + (\sigma_2')^2 =$

$\sigma_2^2 + (\sigma_1')^2 = \mathbb{V}ar(Z_2)$", implies:

$$
\begin{aligned}
0 &= (\sigma_2^2 - \sigma_1^2) - \left[(\sigma_2')^2 - (\sigma_1')^2\right] \\
&= \gamma(\sigma_2 + \sigma_1) - \gamma'(\sigma_2' + \sigma_1') \\
&= \gamma(2\sigma_1 + \gamma) - \gamma'(2\sigma_1' + \gamma').
\end{aligned}
$$

Now, assume $\gamma = \gamma'$. Substituting this assumption simplifies the equation to:

$$
\begin{aligned}
0 &= \gamma\,(2\sigma_1 + \gamma) - \gamma\,(2\sigma_1' + \gamma) \\
&= \gamma\,[2(\sigma_1 - \sigma_1')]\,.
\end{aligned}
$$

Since $\gamma \neq 0$, it follows that $\sigma_1 = \sigma_1'$.

By the fact that $\gamma \neq 0$, immediately we have $\sigma_1 = \sigma_1'$ and $\sigma_2 = \sigma_2'$. Now with $Z_1 = \sigma_1 X + \sigma_2 Y$ and $Z_2 = \sigma_2 X + \sigma_1 Y$, we will show that $\Phi = \Phi'$. Assume $\Phi \neq \Phi'$, then there exists a $k \in \mathbb{N}$ such that $\mathbb{E}[X^i] = \mathbb{E}[Y^i]$ for $i = 0, \ldots, k-1$ and $\mathbb{E}[X^k] \neq \mathbb{E}[Y^k]$. Obviously $k \geq 3$ by $\mathbb{E}[X] = \mathbb{E}[Y] = 0$ and $\mathbb{V}ar(X) = \mathbb{V}ar(Y) = 1$. Since Z_1 and Z_2 follow exactly the same p.d.f., we have $\mathbb{E}[Z_1^k] = \mathbb{E}[Z_2^k]$. However this equality contradicts with the fact that $\sigma_2 \neq \sigma_1$ and $\mathbb{E}[Y^k] \neq \mathbb{E}[X^k]$. Therefore $\Phi = \Phi'$, i.e. $f = f'$.

To summarize, we have shown in this part that if $\alpha_1 = \alpha_2 = 0$ and $\gamma = \gamma'$, then $\pi_i = \pi_i'$ and $\sigma_i = \sigma_i'$ for $i = 1, 2$ and $f = f'$.

2. In this part, we tried to show that neither $\alpha_1 \neq 0$, $\alpha_2 \neq 0$ nor $\gamma \neq \gamma'$ is admissible. However, there is no evidence to show that the conditions are not admissible. Thus we could not follow the same technique to complete the identifiability for two-component semiparametric location-scale mixture and we leave this for our future research. □

Chapter 2

MHDE for Parametric Location-Scale Mixture Model

2.1 Introduction

Mixture models have long been a powerful tool in statistical modeling, providing a flexible framework for representing complex data distributions. Among these, parametric location-scale mixture models are particularly prominent due to their ability to describe data as a combination of subpopulations, each characterized by its own location and scale parameters. This chapter focuses on the minimum Hellinger distance estimation (MHDE) for parametric location-scale mixture models, a robust approach that ensures efficiency and robustness in parameter estimation.

2.1.1 Parametric Location-Scale Mixture Models

Parametric location-scale mixture models assume that the observed data can be represented as a mixture of several subpopulations, each following a known distribution defined by location and scale parameters. These models are widely used in various fields such as biology, finance, and engineering due to their flexibility in modeling heterogeneous data.

The general form of a parametric location-scale mixture model is given by:

$$f(x;\theta) = \sum_{k=1}^{K} \pi_k f_k \left(\frac{x-\mu_k}{\sigma_k} \right),$$

where π_k are the mixing proportions, μ_k and σ_k are the location and scale parameters of the k-th component, respectively, and f_k are the component density functions.

Historically, significant contributions to the development and application of these models include the work of Newton et al. (1986) on statistical analysis of finite mixture distributions, and McLachlan and Peel (2000) on finite mixture models, which provided comprehensive treatments of theoretical and practical aspects of mixture modeling. Recent advances in the study of location-scale mixture models have focused on improving the flexibility and applicability of these models in various contexts. For instance, Frühwirth-Schnatter and Pyne (2010) provided a comprehensive review of Bayesian methods for finite mixture models, emphasizing the importance of prior selection and computational efficiency.

2.1.2 Estimation Techniques for Parametric Location-Scale Mixture Models

Estimating the parameters of parametric location-scale mixture models is challenging due to the complexity of the likelihood function, which often exhibits multiple local maxima. Traditional estimation methods primarily focus on the maximum likelihood estimation (MLE) approach.

The Expectation-Maximization (EM) algorithm, introduced by Dempster et al. (1977), is a widely used method for MLE in mixture models. The EM algorithm iteratively maximizes the likelihood function by alternating between an expectation step (E-step), which computes the expected value of the log-likelihood given the current parameter estimates, and a maximization step (M-step), which updates the parameter estimates to maximize this expected log-likelihood.

Despite its popularity, the EM algorithm has limitations, particularly its sensitivity to initial values and convergence to local maxima. Consequently, robust estimation techniques have been developed to address these issues.

One such technique is the minimum Hellinger distance estimation (MHDE), first introduced by Beran (1977) for parametric models. The Hellinger distance, a measure of the discrepancy between two probability distributions, provides a robust alternative to likelihood-based methods. The MHDE minimizes the Hellinger distance between the empirical distribution and the model distribution, yielding estimators that are asymptotically efficient and robust against model misspecification.

Recent developments in MHDE for mixture models include the work of Kang (2020) that explores the limitations of linear mixed models with normal errors in handling outliers in longitudinal data, proposing a robust Bayesian estimation method using minimum Hellinger distance that demonstrates superior performance in terms of biases, mean squared errors, and standard errors in both simulated and real-world datasets.Wu and Karunamuni (2012) extended MHDE to semiparametric models, further establishing its applicability in more complex settings. Additionally, Cutler and Cordero-Brana (1996) showed the effectiveness of MHDE in finite mixture models, providing a foundation for further research in this area.

2.2 Motivation

This chapter aims to develop and evaluate the MHDEP for parametric location-scale mixture models. We will begin by formulating the MHDEP and discussing its theoretical properties, including consistency and asymptotic normality. We will then present an algorithm for implementing the MHDEP, addressing practical considerations such as initialization and convergence.

The performance of the MHDEP will be assessed through simulation studies, comparing it with traditional MLE and other robust estimation methods. Finally, we will apply the MHDEP to real-world data, illustrating its practical advantages and limitations.

By the end of this chapter, we aim to demonstrate that MHDEP provides a robust and efficient alternative to traditional estimation methods for parametric location-scale mixture models, offering valuable insights and tools for both theoretical and applied researchers.

2.3 Model and Estimation

2.3.1 Parametric location-scale mixture model

In this section, we introduce the Minimum Hellinger Distance Estimator (MHDEP) for the parameters of a finite parametric location-scale mixture model. The parametric k-component mixture model is an essential framework in statistical modeling, providing a flexible method to capture heterogeneous data distributions. This model assumes that the observed data x are generated from a mixture of k different subpopulations, each represented by a component density f with its own location and scale parameters.

The general form of a parametric k-component location-scale mixture model can be expressed as:

$$h(x) \; = \; \sum_{i=1}^{k} \pi_i \frac{1}{\sigma_i} f\left(\frac{x-\mu_i}{\sigma_i}\right),$$

where f is a known probability density function for each component, and the parameters of interest are the mixture proportions π_i, the location parameters μ_i, and the scale parameters σ_i. The parameter vector $\theta = (\pi_i, \mu_i, \sigma_i)^T$ encapsulates all the unknowns that we aim to estimate.

In this thesis, we focus on the simpler but practically significant case of a two-component parametric mixture model, i.e., $k = 2$. This model is defined as:

$$H_1 = \left\{ h = h_\theta(x) = \pi_1 \frac{1}{\sigma_1} f\left(\frac{x-\mu_1}{\sigma_1}\right) + \pi_2 \frac{1}{\sigma_2} f\left(\frac{x-\mu_2}{\sigma_2}\right) : \theta \in \Theta \right\}, \tag{2.1}$$

where H_1 is parametric two-component location-scale mixture family, $\theta = (\pi_1, \pi_2, \mu_1, \mu_2, \sigma_1, \sigma_2)^\top$ represents the vector of unknown parameters, and f is a member of location-scale family of probability density functions defined on the real line. The parameter space Θ is defined with the following constraints to ensure the model's identifiability and practical applicability:

$$\begin{aligned} \Theta = \Big\{ \theta \in \mathbb{R}^6 : 0 < \pi_1, \pi_2 < 1, \quad \pi_1 + \pi_2 = 1, \quad \pi_1, \pi_2 \in (0, 0.5) \cup (0.5, 1), \\ \mu_1 < \mu_2, \quad \mu_1, \mu_2 \in [L_\mu, R_\mu], \quad \sigma_1 \neq \sigma_2, \quad \sigma_1 > 0, \quad \sigma_2 > 0, \\ \sigma_1, \sigma_2 \in [L_\sigma, R_\sigma], \quad L_\mu, R_\mu, R_\sigma < \infty, \quad L_\sigma > 0 \Big\}. \end{aligned} \tag{2.2}$$

In this formulation:

π_1 and π_2 are the mixture proportions, ensuring that the sum of the proportions equals one, $\pi_1 + \pi_2 = 1$, reflecting the complete partitioning of the data.

μ_1 and μ_2 are the location parameters for the two components, constrained by $\mu_1 < \mu_2$ to ensure the model's identifiability.

σ_1 and σ_2 are the scale parameters, constrained to be positive and distinct, ensuring that each component has a unique variance structure.

L_μ and R_μ define the range within which the location parameters lie, while L_σ and R_σ define the range for the scale parameters.

The goal is to estimate the parameter vector θ using the Minimum Hellinger Distance Estimation (MHDEP) approach. The Hellinger distance is a measure of the dissimilarity between two probability distributions, and it is particularly robust against outliers and model misspecification. By minimizing the Hellinger distance between the empirical density function and the model density, we aim to achieve robust and efficient parameter estimates that are less sensitive to deviations from the model assumptions.

Different estimations have been studied for finite parametric location-scale mixture models, such as likelihood-based estimation and distance-based estimation. It is well-known that the likelihood for a finite mixture of normal distributions is unbounded, thus the ordinary likelihood estimator is not consistent. To avoid this problem, some researchers have turned to estimators with constraints on parameter spaces, see Redner (1981), Hathaway (1985), and penalized Maximum Likelihood Estimation (PMLE) from Chen et al. (2008). Except for putting constraints on parameter space, Yao (2010) proposed a profile log-likelihood method and a graphical way to find the maximum interior mode. Zhang and Chen (2012) proposed a Minimum Wasserstein Distance Estimator (MWDE) for finite location-scale mixtures and found the MWDE to be consistent. They compared the MWDE with PMLE and showed that MWDE is less efficient but obviously more robust than PMLE. Recent advancements have further enhanced the estimation techniques for parametric mixture models. For instance, Lambert et al. (2022) proposes new methods for variational inference based on gradient flows in the Bures–Wasserstein space, offering strong theoretical guarantees for Gaussian approximations in log-concave settings. Wang (2023) discusses the challenges of regression analysis with heterogeneous populations, emphasizing the need for clustering when subpopulations are unidentified, and critiques the use of infinite Bayesian mixture models like Dirichlet process mixtures for their inconsistency in estimating the number of clusters.

Identifiability for finite parametric location-scale mixture models has been extensively investigated. The most commonly used finite mixture is the normal location-scale mixture model. Teicher (1960) proved that finite normal location-scale mixtures are identifiable. Teicher (1961) discussed the identifiability of parametric mixtures such as Type 3 and uniform distributions under certain conditions. In more recent years, researchers have continued to explore identifiability conditions for various mixture models. Otiniano et al. (2015) extended the identifiability results to mixtures of skew-normal distributions, providing new insights into the behavior of these models under different parameter settings. The work by Holzmann et al. (2006) extended the identifiability results to mixtures of elliptical distributions. They showed that under certain conditions on the characteristic generator, the parameters of finite mixtures of elliptical distributions can be identified. A more general class of distributions that extend beyond parametric models is the class of elliptical distributions. Elliptical distributions are a family of multivariate distributions characterized by the fact that their contours of equal density form ellipsoids. Notably, the normal distribution is a special case of this class, but elliptical distributions also include others such as Student's t-distribution and Cauchy distribution. Elliptical distributions are often described in terms of their characteristic functions, which generalize the moment generating function. The characteristic function of an elliptical distribution can be expressed in terms of a characteristic generator that governs the distribution's properties. For elliptical distributions, these characteristic generators have been studied extensively, leading to deeper insights into identifiability. The key result from Holzmann et al. (2006) hinges on Schoenberg's theorem, which states that a function $\phi_G(u)$ is a characteristic generator if and only if it can be written as a specific Laplace transform of a probability measure G. The theorem is based on a condition that relates the Laplace transform $L_G(u)$ of the mixing measure G. The critical identifiability condition is that for a scaling factor $a > 1$,

$$\lim_{u \to \infty} \frac{L_G(au)}{L_G(u)} = 0,$$

which ensures that the mixtures generated by the characteristic function $\phi_G(u)$ are identifiable. In the case of normal distributions (a special case of elliptical distributions), this condition is satisfied as long as the means and variances of the mixture components are distinct. Here we will show how to apply the identifiability condition from Holzmann et al. (2006) to our two-component location-scale normal mixture model. In the context of this thesis, we consider a two-component location-scale mixture model where each component follows a normal distribution for the parametric case studies. The model can be expressed as:

$$h_{\theta,f}(x) = \pi_1 \frac{1}{\sigma_1} f\left(\frac{x-\mu_1}{\sigma_1}\right) + \pi_2 \frac{1}{\sigma_2} f\left(\frac{x-\mu_2}{\sigma_2}\right),$$

where

$$\theta = (\pi_1, \pi_2, \mu_1, \mu_2, \sigma_1, \sigma_2)^\top$$

are the unknown parameters, $\pi_1 + \pi_2 = 1$, and f is the probability density function of the normal distribution. This is a specific case of the broader class of elliptical distributions, as normal distributions are a special case within this family. Therefore, the results from the Holzmann et al. (2006) apply directly to this model. To apply the identifiability theorem to this model, we need to check the conditions under which the two components of the mixture are distinguishable. The normal distribution has a well-defined characteristic function, and in this case, the characteristic generator $\phi_G(u)$ corresponds to the normal distribution's characteristic function.

From Holzmann et al. (2006) identifiability is guaranteed as long as the Laplace transform $L_G(u)$ satisfies the decay condition mentioned earlier. For the normal distribution, this condition is naturally satisfied as long as the two components of the mixture have distinct means ($\mu_1 \neq \mu_2$) and distinct scales ($\sigma_1 \neq \sigma_2$). Equivalently, if for any $\mu_2 > \mu_1$, $\lim_{t\to\infty} \phi(\mu_2 t)/\phi(\mu_1 t) = 0$, for any $\sigma_1 > \sigma_2 > 0$, $\lim_{t\to\infty} \phi(\sigma_1 t)/\phi(\sigma_2 t) = 0$, then the finite normal location-scale mixture model is identifiable. From now on, we assume that the identifiability condition mentioned above is satisfied for the subsequent parametric simulation studies.

2.3.2 Construction of MHDEP

Let $X_1, \ldots, X_n$ be independent observations with sample size of n

$$X_1, \ldots, X_n \overset{\text{i.i.d}}{\sim} h_\theta(x), \tag{2.3}$$

where $h_\theta(x)$ is the finite parametric location-scale mixture model in (2.1).

In order to employ the MHD technique of Beran (1977) which has proved the theoretical properties for general parametric model, for parametric MHDE, as each component density is completely known, we will study the properties of $\hat{\theta}_{MHDE}$ under specific parametric two-component location-scale mixture model. The proposed estimator of θ is that value $\hat{\theta}_{MHDE}$ in parameter space Θ which minimizes the Hellinger distance between h_θ and $\hat{h}$, where $\hat{h}$ is a suitable nonparametric estimator of the density of the parametric two-component location-scale mixture based on data $X_1, \ldots, X_n$.

Here we use the following adaptive nonparametric kernel density estimator from Scott (1992) of h_θ based on data $X_1, \ldots, X_n$:

$$\hat{h}(x) = \frac{1}{nb_n s_n} \sum_{j=1}^{n} K(\frac{x - X_j}{b_n s_n}), \tag{2.4}$$

where K is a symmetric density function, a sequence of bandwidths b_n are positive constants satisfy $b_n \to 0$ as $n \to \infty$ and s_n is a robust scale statistic (this statistic generally estimates the scale parameter of the respective distribution). A range of kernel functions are commonly used such as uniform, triangular, bi-weight, triweight, Epanechnikov, normal and others. The kernel density estimator (2.4) is the same as the one used in Beran (1977) for density estimator.

The minimum Hellinger distance estimator $\hat{\theta}_{MHDE}$ is defined as $T(\hat{h})$, where $\hat{h}$ is the kernel density estimator (2.4) and T is a functional. Let H denote the set of all densities with respect to Lebesgue measure on the real line, $\|\cdot\|$ denotes the L_2 norm, then the constructed estimator is defined as

$$\hat{\theta}_{MHDEP} = T(\hat{h}) = \arg\min_{\theta \in \Theta} \left\| h_\theta^{\frac{1}{2}} - \hat{h}^{\frac{1}{2}} \right\| = \min_{\theta \in \Theta} \left(\int \left(h_\theta^{1/2}(x) - \hat{h}^{1/2}(x) \right)^2 dx \right)^{1/2}. \tag{2.5}$$

Since our proposed estimator is also a distance-based estimator, we will study the asymptotic properties of the Minimum Hellinger Distance Estimator (MHDEP) and compare it with other existing estimations for parametric finite mixture models. The robustness and efficiency of MHDE make it a promising tool for analyzing complex datasets, especially those with potential model misspecification or outliers.

In the subsequent sections, we will introduce the construction of the MHDEP, discuss the theoretical underpinnings of MHDEP, the estimation procedure, and the asymptotic properties of the estimators. We will also illustrate the application of MHDEP to the two-component parametric location-scale mixture model through simulations and real data examples, demonstrating its practical advantages and effectiveness.

2.4 Theoretical asymptotic properties of MHDEP for parametric mixture model

Since Beran (1977) has tested the asymptotic properties such as consistency and asymptotic normality for MHD of general parametric models, in order to employ the technique from Beran (1977), we have the following two corollaries regarding the consistency and asymptotic normality for (2.5) under finite parametric location-scale mixture model. Corollary 1 studies the consistency and Corollary 2 studies the asymptotic normality.

Corollary 1. ***Consistency of MHDEP***

Suppose

1. *For almost every x, $h_\theta(x)$ is continuous in θ.*

2. *K is absolutely continuous and has compact support; w' is bounded.*

3. $\lim_{n\to\infty} b_n = 0$*,* $\lim_{n\to\infty} n^{1/2} b_n = \infty$*.*

4. *As $n \to \infty$, $s_n \xrightarrow{p} s$ a positive finite constant depending on h_θ.*

5. *The density f for each component is uniformly continuous.*

6. *T is a functional continuous at h_θ in the Hellinger metric.*

Then $\hat{\theta}_{MHDE} \xrightarrow{p} \theta$ as $n \to \infty$ where θ is the true parameter.

Explanation for condition 1 in Corollary 1:

By condition 1 "Assume $h_\theta(x)$ is continuous in θ" in Corollary 1, it is equivalent to the condition in Theorem 1 in Beran (1977) that "for almost every x, $f_\theta(x)$ is continuous in θ." As the parameter vector of interest θ in our model is $\theta = (\pi_1, \mu_1, \sigma_1, \pi_2, \mu_2, \sigma_2)^T$, it is equivalent to prove for almost every x, $h_\theta(x)$ is continuous in $\theta = (\pi_1, \mu_1, \sigma_1, \pi_2, \mu_2, \sigma_2)^T$.

Explanation for condition 5 in Corollary 1:

By condition 5 "The density f for each component is uniformly continuous." in Corollary 1, it is equivalent to the condition (ii) in Theorem 3 in Beran (1977) that "g is uniformly continuous.". The following is to investigate what we can conclude from this condition "The density f for each component is uniformly continuous."

To prove that $h_\theta(x)$ is uniformly continuous, we need to show that for every $\varepsilon > 0$, there exists a $\delta > 0$ such that for all $x, y \in \mathbb{R}$,

$$|x - y| < \delta \implies |h_\theta(x) - h_\theta(y)| < \varepsilon.$$

Step 1: Express $h_\theta(x)$

The function $h_\theta(x)$ is defined as:

$$h_\theta(x) = \frac{\pi_1}{\sigma_1} f\left(\frac{x-\mu_1}{\sigma_1}\right) + \frac{\pi_2}{\sigma_2} f\left(\frac{x-\mu_2}{\sigma_2}\right),$$

where $f(\cdot)$ is a continuous density function.

Step 2: Uniform Continuity of $f(\cdot)$

Since $f(\cdot)$ is continuous and defined on the real line, it is also uniformly continuous. Thus, for every $\varepsilon > 0$, there exists a $\delta > 0$ such that:

$$|u - v| < \delta \implies |f(u) - f(v)| < \varepsilon.$$

Step 3: Uniform Continuity of Scaled and Shifted $f(\cdot)$

Let us consider the scaled and shifted arguments of $f(\cdot)$, specifically:

$$f\left(\frac{x-\mu_1}{\sigma_1}\right) \quad \text{and} \quad f\left(\frac{x-\mu_2}{\sigma_2}\right).$$

Using the uniform continuity of $f(\cdot)$, we choose $\delta_1 = \delta\sigma_1$ for the term involving μ_1 and $\delta_2 = \delta\sigma_2$ for the term involving μ_2. Then:

$$|x - y| < \delta_1 \implies \left|\frac{x-\mu_1}{\sigma_1} - \frac{y-\mu_1}{\sigma_1}\right| < \delta \implies \left|f\left(\frac{x-\mu_1}{\sigma_1}\right) - f\left(\frac{y-\mu_1}{\sigma_1}\right)\right| < \frac{\varepsilon}{2}.$$

Similarly:

$$|x - y| < \delta_2 \implies \left|\frac{x-\mu_2}{\sigma_2} - \frac{y-\mu_2}{\sigma_2}\right| < \delta \implies \left|f\left(\frac{x-\mu_2}{\sigma_2}\right) - f\left(\frac{y-\mu_2}{\sigma_2}\right)\right| < \frac{\varepsilon}{2}.$$

Step 4: Combine Results for $h_\theta(x)$

Now, consider the difference:

$$h_\theta(x) - h_\theta(y) = \frac{\pi_1}{\sigma_1}\left[f\left(\frac{x-\mu_1}{\sigma_1}\right) - f\left(\frac{y-\mu_1}{\sigma_1}\right)\right] + \frac{\pi_2}{\sigma_2}\left[f\left(\frac{x-\mu_2}{\sigma_2}\right) - f\left(\frac{y-\mu_2}{\sigma_2}\right)\right].$$

Taking absolute values and applying the triangle inequality:

$$|h_\theta(x) - h_\theta(y)| \le \left|\frac{\pi_1}{\sigma_1}\left[f\left(\frac{x-\mu_1}{\sigma_1}\right) - f\left(\frac{y-\mu_1}{\sigma_1}\right)\right]\right| + \left|\frac{\pi_2}{\sigma_2}\left[f\left(\frac{x-\mu_2}{\sigma_2}\right) - f\left(\frac{y-\mu_2}{\sigma_2}\right)\right]\right|.$$

Step 5: Use the Bounds

Using the bounds from Step 3:

$$\left|f\left(\frac{x-\mu_1}{\sigma_1}\right) - f\left(\frac{y-\mu_1}{\sigma_1}\right)\right| < \frac{\varepsilon}{2} \quad \text{and} \quad \left|f\left(\frac{x-\mu_2}{\sigma_2}\right) - f\left(\frac{y-\mu_2}{\sigma_2}\right)\right| < \frac{\varepsilon}{2}.$$

Thus:

$$|h_\theta(x) - h_\theta(y)| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

Step 6: Conclusion

Since for every $\varepsilon > 0$, we can find a $\delta > 0$ such that $|x-y| < \delta \implies |h_\theta(x) - h_\theta(y)| < \varepsilon$, it follows that $h_\theta(x)$ is uniformly continuous.

□

Conditions 2,3,4,6 are exactly the same as the conditions in Theorem 3 in Beran (1977). Thus from above proof, all the 6 conditions in Corollary 1 are equivalent to the conditions in Theorem 3 in Beran (1977).

□

Corollary 2 ***Asymptotic Normality of MHDEP*** *Suppose*

1. *K is symmetric about 0 and has compact support.*

2. *K is twice absolutely continuous: K is bounded.*

3. *For every $t \in int(\Theta)$, $f'(x)$ exists, $f(x)$ and $f'(x)$ are continuous in x.*

4. *For every $t \in int(\Theta)$, $\int \frac{(f')^2}{f}(x)dx < \infty$, $\int \frac{(f')^2}{f}(x) \times x^2 dx < \infty$.*

5. *f has compact support H.*

6. *$f > 0$ on the compact support H; f is twice absolutely continuous and f" is bounded.*

7. *$\lim_{n\to\infty} n^{1/2} b_n = \infty$, $\lim_{n\to\infty} n^{(1/2)} b_n^2 = 0$.*

8. *There exists a positive finite constant s depending on the parametric mixture h_θ such that $\sqrt{(n)}(s_n - s)$ is bounded in probability.*

Then the limiting distribution of $n^{\frac{1}{2}}[\theta_{MHDE} - \theta]$ is

$$N\left(0, \left[\int \frac{1}{(h_\theta(x) h_\theta^T(x))^{\frac{1}{2}}} \cdot \frac{\partial h_\theta(x)}{\partial \theta} \cdot \frac{\partial h_\theta^T(x)}{\partial \theta}\, dx\right]^{-1}\right).$$

Explanation for condition 3 in Corollary 2.

By condition 3 "For every $t \in int(\Theta)$, $f'(x)$ exists, $f(x)$ and $f'(x)$ are continuous in x" in Corollary 2, we aim to prove it is equivalent to the condition (i) in LEMMA 1 in Beran (1977) that "For every $x \notin N$ (a Lebesgue null set) and for every θ in some neighborhood of t, $s_\theta(x)$ has first partial derivatives with respect to θ which are continuous in θ at $\theta = t$." In our model, assuming $f(x)$ and $f'(x)$ are continuous in x, we suppose $h_\theta^{-1/2}(x)$ has first partial derivatives with respect to θ which are continuous in $\theta = (\pi_1, \pi_2, \mu_1, \mu_2, \sigma_1, \sigma_2)^T$.

Explanation for condition 6 in Corollary 2.

Since in Beran (1977) $g > 0$, in our model g is our $h_\theta(x)$, thus we set $f > 0$. $h_\theta(x)$ is twice absolutely continuous is equivalent to f is twice absolutely continuous.

Conditions 1,2,4,5,7,8 are exactly the same as the conditions in Theorem 3 in Beran

(1977).

2.5 Simulation Study

2.5.1 Efficiency Study

For the efficiency study, in order to compare the parametric estimations with semiparametric estimations, we build upon the same scenario discussed in Chapter 3 for the semiparametric case. Specifically, we extend our analysis to include various estimation methods for the parametric case, such as the Minimum Hellinger Distance Estimator for Parametric Models (MHDEP), Hellinger Minimization Mixtures (HMIX) algorithm, and Maximum Likelihood Estimation (MLE). These parametric estimation methods will be systematically compared with the Semiparametric Minimum Hellinger Distance Estimator (SEMIMHDE) within the same framework. This comparison aims to provide a comprehensive understanding of the relative efficiencies of these estimators.

Tables 3.16 through 3.33 will present detailed efficiency studies across three different cases, showcasing both parametric and semiparametric estimations. Each table will be meticulously analyzed to elucidate the performance characteristics of MHDEP, HMIX, and MLE relative to SEMIMHDE. The comparative analysis will highlight the strengths and potential limitations of each estimation method under various scenarios. These findings will be thoroughly discussed in Chapter 3, providing a nuanced interpretation of the results and offering insights into the practical implications of choosing between parametric and semiparametric estimation techniques in real-world applications.

2.5.2 Robustness Study

Following the efficiency study, our attention shifts towards examining the robustness of the Minimum Hellinger Distance Estimator (MHDE) for parametric location-scale mixture models. Robustness is a critical attribute of any estimator, as it indicates the estimator's resilience to deviations from model assumptions and the presence of outliers.

To provide a comprehensive comparison, we will conduct a robustness study that parallels the one detailed in Chapter 3 for semiparametric estimations. This approach ensures consistency and allows for a clear and direct comparison between parametric and semiparametric estimation methods. Specifically, we will evaluate the robustness of parametric estimation techniques such as the Minimum Hellinger Distance Estimator for Parametric Models (MHDEP), the HMIX algorithm, and Maximum Likelihood Estimation (MLE). These parametric methods will be juxtaposed with the Semiparametric Minimum Hellinger Distance Estimator (SEMIMHDE) within the same tables to facilitate a direct comparison.

Tables 3.34 through 3.37 will meticulously present the robustness studies across the case with outliers, including both parametric and semiparametric estimations. These tables will serve as a detailed record of the robustness performance of each estimator. The ensuing analysis will be presented in Chapter 3, where we will delve into the comparative performance, highlighting the robustness of each method. This analysis will shed light on the practical implications and provide insights into the scenarios where one estimator may be preferred over others due to its robustness properties. The findings from this robustness study will contribute to a deeper understanding of the strengths and limitations of the MHDEP, HMIX, MLE, and SEMIMHDE estimators in real-world applications.

2.5.3 Model Misspecification Study

Following the efficiency and robustness study, our attention shifts towards examining the model misspecification of the Parametric Minimum Hellinger Distance Estimator (MHDEP) for parametric location-scale mixture models. For this study, we aasume our data is generated from a uniform distribution and MHDEP is wrongly assume the data follows a normal distribution. We aims to see how MHDEP performs when the distribution is under wrong assumptions. The simulation results for this part is also presented in Chapter 3 in Table 3.38 and Table 3.39 to compare together with the semiparametric MHDE (SEMIMHDE).

2.6 Real Data Analysis

2.6.1 Hydropower Data Introduction

In this section, we demonstrate the effectiveness of the proposed parametric Minimum Hellinger Distance Estimation (MHDEP) by applying it to a real-world dataset from the hydropower data. The dataset comprises hourly inflow measurements collected from January 1, 2010, to October 23, 2013, resulting in a total of 24 measurements per day over the entire observation period, which are inferred from observations of the reservoir's surface height and the outflow of water from the reservoir. Due to the nature of these measurements, the data contains significant noise and errors, which are reflected in numerous negative values, likely the result of imprecise instrumentation or environmental factors influencing the recording process.

A notable feature of the dataset is its distinct seasonal pattern, which results in markedly different behaviour during the winter months compared to the spring and summer seasons. In particular, during the spring and early summer, snowmelt contributes substantially to inflows, leading to a period of heightened variability in the measurements. This seasonal variation in the inflow data strongly suggests a mixture distribution, as the data appears to be a composite of at least two distinct underlying processes: one governing the lower inflows typical in winter, and the other reflecting the higher inflows during periods of snowmelt.

To further investigate this, we begin by plotting a histogram of the raw inflow measurements which is shown in Figure 2.1. The histogram clearly shows a bimodal distribution, with two distinct peaks in the density. This bimodal structure provides a strong indication that a two-component location-scale mixture model is appropriate for describing the data. The presence of these two modes suggests the existence of two subpopulations within the dataset, possibly corresponding to different hydrological conditions (e.g., dry and wet seasons).

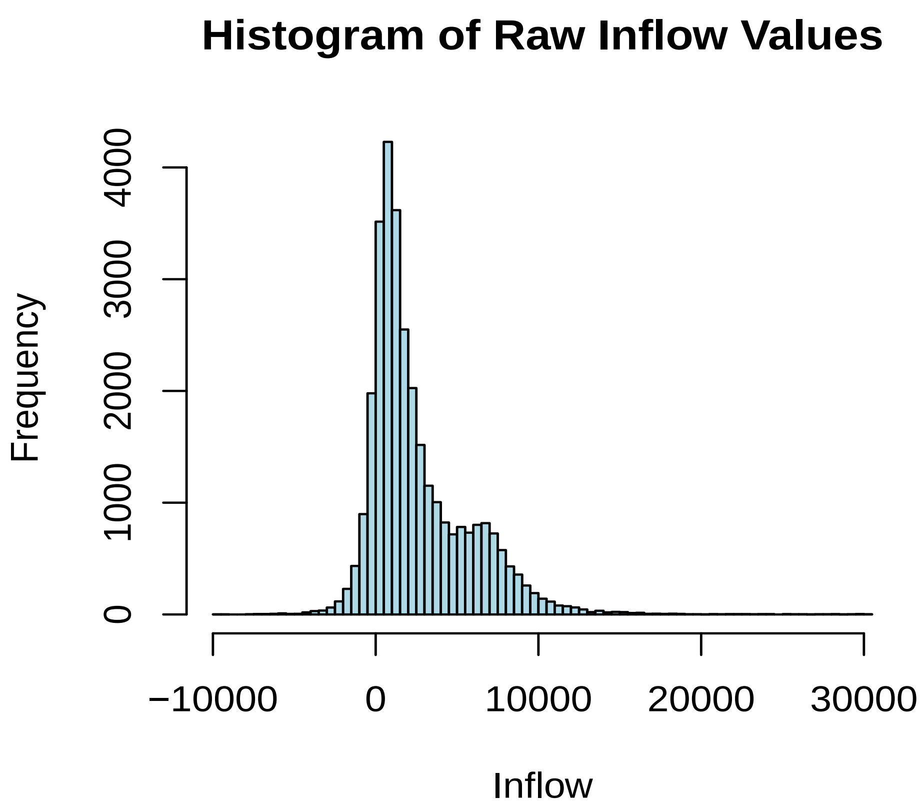


Figure 2.1: Histogram of Raw Inflow Values

Given the characteristics of the dataset, our goal is to apply the proposed two-component location-scale normal mixture model using the MHDEP method. The mixture model is particularly well-suited to capturing the complex patterns in the data, as it allows for the estimation of two separate normal components, each with its own mean and variance. These components can then be interpreted as representing distinct inflow regimes, thereby providing a clearer understanding of the inflow dynamics under different environmental conditions. We will use the parametric MHDEP to estimate the parameters of the mixture model, including the mixing proportions, means, and variances of the two components. Finally, we will overlay the MHDEP-fitted density curve on the histogram of the modified inflow data to visually assess how well the proposed method captures the underlying structure of the real dataset. This comparison will provide an intuitive understanding of the performance of the MHDEP and its ability to accurately model the observed distribution, despite the presence of potential errors and seasonal variations in the data.

2.6.2 Imputing New Values for Negative Data

Due to the inherent measurement errors in the hydropower inflow data, around 18% of the recorded inflow values are negative. These negative values do not reflect actual inflow measurements but are rather the result of truncation, which must be addressed before applying our parametric model. We consider a scenario in which the original dataset contains measurement errors, potentially resulting in negative values. However, prior to data analysis, these negative values have been censored and replaced with zeros, likely due to an overly cautious preprocessing step. Below is the algorithm designed to correct these errors and generate a more reliable dataset for analysis:

1. **Fitting a Partial Normal Distribution to Negative Values**

   Begin by fitting a normal distribution to the negative portion of the inflow data, which represents the measurement errors. Let p be the fraction of negative values in the dataset. To model these errors, assume the negative values are generated by a normal distribution with mean $\mu > 0$ and standard deviation $\sigma = \mu/z_p$, where z_p is the p-th quantile of the standard normal distribution. Estimate the value of μ using Maximum Likelihood Estimation (MLE).

2. **Simulating Replacement Values for Negative Data**

   Once the normal distribution is fitted, simulate replacement values for the negative measurements. Generate a sample by selecting a uniform random variable from the interval $[0, p]$, and then apply the inverse cumulative density function (CDF) of the fitted normal distribution to this value, effectively simulating from the tail of the distribution to replace the negative values.

3. **Estimating the Mixture Distribution**

   Estimate the mixture distribution of the full dataset, assuming the data follows a two-component normal location-scale mixture model. Apply the Expectation-Maximization (EM) algorithm to estimate the parameters: mixing proportions, means, and standard deviations for both components of the mixture.

4. **Simulating Negative Data from the Mixture Distribution**

   Using the estimated mixture model, simulate new values to replace the negative observations. Let the estimated mixture density from step 3 be $af_1(x) + (1-a)f_2(x)$, where $f_1(x)$ and $f_2(x)$ represent the two components of the mixture. Define p_1 and p_2 as the areas under $f_1(x)$ and $f_2(x)$ below zero, respectively. The conditional probability of a negative value being generated from $f_1(x)$ is given by:

$$q = \frac{ap_1}{ap_1 + (1-a)p_2}.$$

Draw a sample from a binomial distribution with parameter q to determine whether the replacement value will be drawn from $f_1(x)$ or $f_2(x)$. If the sample is from $f_1(x)$, generate a uniform random variable in $[0, p_1]$, and apply the inverse CDF of $f_1(x)$. If the sample is from $f_2(x)$, repeat the process using p_2 and the inverse CDF of $f_2(x)$.

5. **Iterative Process**

   Repeat steps 3 and 4 as necessary, using the updated estimates to iteratively refine the replacement of negative values. Continue the process until the algorithm converges to a stable solution, resulting in a complete and corrected dataset.

2.6.3 Results Obtained from MHDEP on New Inflow Data

After applying the five-step data preprocessing algorithm to modify the negative values, we obtain a new, modified dataset that can now be treated as complete data. This new dataset is serves as the foundation for evaluating the performance of the parametric Minimum Hellinger Distance Estimation (MHDEP).

The parameter estimates obtained through MHDEP for the new dataset are as follows:

$$\hat{\theta}_{\text{MHDEP}} = (0.459, 0.541, 0.987, 1.338, 3.889, 3.832).$$

To visually assess the goodness-of-fit, we superimpose the MHDEP-fitted density plot onto the histogram of the modified inflow data, as shown in Figure 2.2. This allows us to observe how well the MHDEP model captures the underlying distribution of the modified data. From Figure 2.2, it is evident that the MHDEP-fitted density aligns closely with the observed distribution of the modified inflow data. The two modes in the histogram are well captured by the model, reflecting the mixture nature of the dataset. The excellent fit suggests that the MHDEP is effectively accommodating the heterogeneity in the data, particularly in handling the potential measurement errors and seasonal variability that the hydropower inflows exhibit. The smoothness of the fitted density indicates that the mixture components have been appropriately identified, and the negative values, once modified, do not adversely affect the estimation process. This demonstrates the robustness of the MHDEP in modeling such real-world data with complexities like error and truncation. In Figure 2.2, the parametric MHDEP (red curve) is compared with the nonparametric Kernel Density Estimate (KDE, green curve) to assess the fit of each method. The figure indicates that both the MHDEP and KDE provide a close approximation to the data. Notably, the MHDEP exhibits a higher peak, which suggests a better alignment with the scale of the first component compared to the KDE. This enhanced peak suggests that MHDEP may capture finer details in the distribution's scale, particularly in regions where the first component dominates.

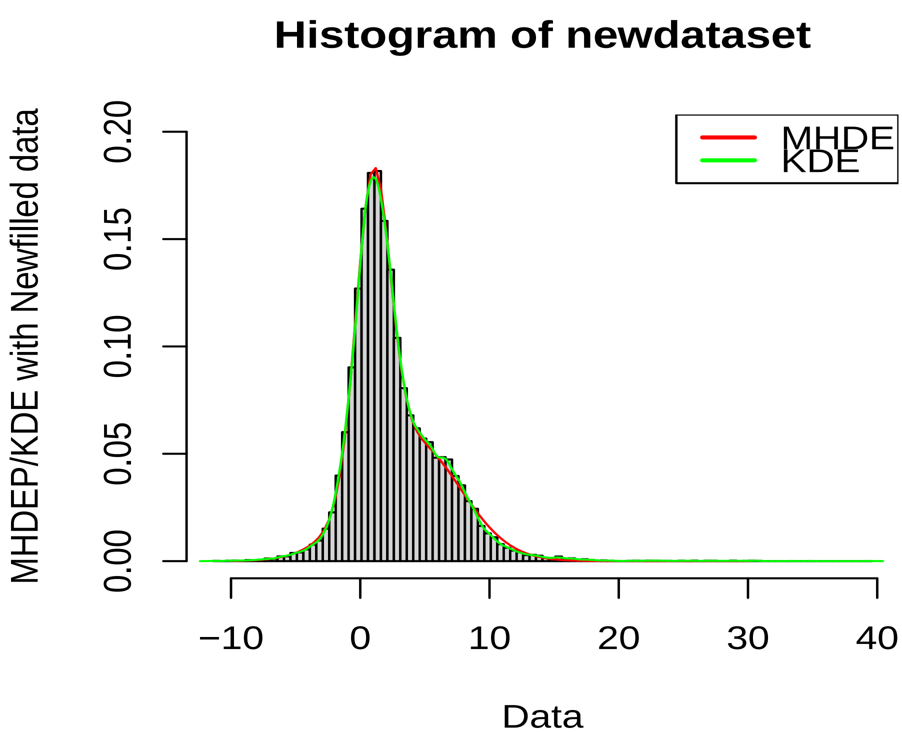


Figure 2.2: MHDEP and KDE Fitted Density Plot with Histogram of Infilled Inflow Values

Chapter 3

MHDE for Semiparametric Location-Scale Mixture Model

3.1 Introduction

The estimation of parameters in semiparametric location-scale mixture models has been an active area of research due to its wide-ranging applications in fields such as biology, economics, and engineering. In biology, semiparametric location-scale mixture models can be used to analyze gene expression data. In RNA sequencing, for instance, gene expression levels often follow a mixture distribution where one component represents non-expressed genes (background noise), and another represents expressed genes. In income distribution analysis, semiparametric location-scale mixture models can be used to capture the heterogeneity in income levels across different subpopulations (e.g., low-income vs. high-income earners). These models combine parametric and nonparametric components, providing flexibility in modeling complex data structures where pure parametric or nonparametric approaches may fall short. This section reviews the development of semiparametric location-scale mixture models and the various estimation techniques that have been proposed, focusing on maximum likelihood estimation (MLE), the EM algorithm, the HMIX algorithm and the minimum Hellinger distance estimator (MHDE).

The concept of semiparametric mixture models dates back to the mid-20th century, with significant contributions from researchers aiming to address the limitations of purely parametric models. Lindsay (1995) laid the groundwork for the theoretical framework of these models, highlighting the advantages of combining parametric and nonparametric elements. The semiparametric approach allows for the flexibility of nonparametric models while retaining the interpretability and efficiency of parametric models.

One of the earliest and most widely used methods for estimating parameters in semiparametric models is maximum likelihood estimation (MLE). MLE seeks to find parameter values that maximize the likelihood function given the observed data. The book Bickel et al. (1993) discusses various semiparametric models and the use of MLE, focusing on the efficient estimation of parameters and the role of influence functions.

The expectation-maximization (EM) algorithm, introduced by Dempster et al. (1977), has been extensively applied to mixture models, including semiparametric ones. The EM algorithm iteratively estimates the parameters by alternating between the expectation step (E-step), which computes the expected log-likelihood, and the maximization step (M-step), which maximizes this expected log-likelihood. Bordes et al. (2006) adapted the EM algorithm for semiparametric mixtures, showing its effectiveness in handling incomplete data and latent variables. Despite its popularity, the EM algorithm can suffer from slow convergence and sensitivity to initial values.

The minimum Hellinger distance estimator (MHDE) has emerged as a robust alternative to MLE and EM for semiparametric models. Beran (1977) first introduced the Hellinger distance in the context of density estimation, and since then, it has been utilized for various statistical problems due to its robustness properties. For semiparametric mixture models, MHDE minimizes the Hellinger distance between the empirical distribution and the model distribution, providing parameter estimates that are robust to model misspecification and resistant to the influence of outliers.

Recent research has focused on improving the efficiency and robustness of estimation techniques for semiparametric location-scale mixture models. Wu and Karunamuni (2012) introduced a profiled minimum Hellinger distance estimator (MPHDE), which profiles out the nuisance parameter to reduce dimensionality and computational complexity. Zhang and Chen (2012) proposed a minimum Wasserstein distance estimator (MWDE) for semiparametric models, demonstrating its robustness compared to traditional likelihood-based estimations such as MLE and EM methods.

Moreover, advancements in computational methods and algorithms have facilitated the practical implementation of these estimators. Techniques such as stochastic approximation Kiefer and Wolfowitz (1952) and variational inference Jordan et al. (1999) have been employed to enhance the convergence properties and scalability of estimation procedures for large datasets.

The semiparametric location-scale mixture model remains a vital tool in statistical modeling, offering a balance between flexibility and interpretability. Estimation techniques such as MLE, EM, and MHDE have evolved significantly over the years, each contributing unique strengths and facing specific challenges. The ongoing development of robust and efficient estimators, coupled with advances in computational methodologies, continues to expand the applicability and performance of semiparametric mixture models in various scientific domains.

In the following sections, we will introduce the construction of the minimum Hellinger distance estimator (MHDE) for semiparametric location-scale mixture models, we modified the HMIX algorithm under parametric mixture to semiparametric mixture, simulation studies such as efficiency studies, robustness studies and distribution studies, highlighting MHDE's advantages and comparing it with other established estimation methods.

3.2 Motivation

The Minimum Hellinger Distance Estimation (MHDE) for semiparametric location-scale mixture models is motivated by the need for robust and efficient estimation techniques in the presence of complex data structures. Semiparametric models combine the flexibility of nonparametric models with the interpretability and efficiency of parametric models, making them particularly useful in diverse real-world applications. This section elucidates the key motivations behind employing MHDE in this context.

Parametric location-scale mixture models, though powerful, often suffer from restrictive assumptions about the underlying data distribution. These assumptions can lead to model misspecification, which in turn affects the accuracy and reliability of parameter estimates. Semiparametric models alleviate this issue by allowing for greater flexibility in modeling the data distribution, accommodating a wider range of data behaviors without being constrained by strict parametric forms.

One of the primary motivations for using MHDE is its robustness to outliers and model misspecification. Traditional estimation methods, such as Maximum Likelihood Estimation (MLE), are highly sensitive to deviations from model assumptions and the presence of outliers. MHDE, on the other hand, minimizes the Hellinger distance between the empirical and theoretical distributions, providing a more robust criterion that reduces the influence of anomalous data points and enhances the reliability of the estimates.

The semiparametric framework aims to strike a balance between efficiency and robustness. While fully nonparametric models offer robustness, they often suffer from inefficiency particularly in small sample sizes, because they do not make use of distributional assumptions to reduce variance and high computational cost especially as the sample size grows, due to the need for more parameters or higher smoothing factors. Parametric models, although efficient, lack robustness. Semiparametric models leverage the strengths of both approaches, and MHDE further enhances this balance by ensuring that the estimation procedure is both statistically efficient and robust to deviations from model assumptions.

Semiparametric location-scale mixture models are particularly suitable for applications involving heterogeneous data, such as biostatistics, finance, and machine learning. These models can capture complex data structures that are not well-represented by simple parametric forms. MHDE facilitates the application of these models by providing a reliable estimation technique that can handle the intricacies of real-world data, making it a valuable tool for practitioners.

The recent advancements in computational methods have made the implementation of MHDE more feasible. Techniques such as stochastic approximation, variational inference, and modern optimization algorithms have significantly reduced the computational burden associated with semiparametric estimation. These advancements allow for the practical application of MHDE in large-scale data settings, further motivating its use.

Finally, the development and application of MHDE for semiparametric location-scale mixture models contribute to the broader field of statistical methodology. By exploring and validating the theoretical properties of MHDE, this research adds to the understanding of robust and efficient estimation techniques, providing a foundation for future methodological advancements and practical applications.

In summary, the motivation for employing MHDE in semiparametric location-scale mixture models is driven by the need for robust, efficient, and flexible estimation methods that can handle complex and heterogeneous data. The integration of MHDE into the semiparametric framework addresses the limitations of traditional parametric models, enhances robustness, and leverages recent computational advancements, making it a compelling approach for modern statistical analysis.

3.3 Model and Estimation

In this section, we propose a Minimum Hellinger Distance Estimation (MHDE) of the parameter θ in a semiparametric two-component location-scale mixture model. In this thesis, we consider the case of $k = 2$, i.e. the class of semi-parametric two-component location-scale mixture models

$$H_2 = \left\{ h \;=\; h_{\theta,f}(x) \;=\; \pi_1 \frac{1}{\sigma_1} f(\frac{x-\mu_1}{\sigma_1}) + \pi_2 \frac{1}{\sigma_2} f(\frac{x-\mu_2}{\sigma_2}) : \theta \in \Theta \right\}, \tag{3.1}$$

where H_2 is the semiparametric two-component location-scale mixture family, $\theta = (\pi_1, \pi_2, \mu_1, \mu_2, \sigma_1, \sigma_2)^\top$, f is a probability density function with standard deviation 1 and symmetric about zero and

$$\begin{aligned}\Theta = \Big\{ &(\pi_1, \pi_2, \mu_1, \mu_2, \sigma_1, \sigma_2)^\top \in \mathbb{R}^6 : \pi_1, \pi_2 \in (0, 0.5) \cup (0.5, 1), \quad \pi_1 + \pi_2 = 1, \\ &\mu_1 < \mu_2, \quad \sigma_1 \neq \sigma_2, \quad \sigma_1 > 0, \quad \sigma_2 > 0 \Big\}. \end{aligned} \tag{3.2}$$

The parameter space $\Theta \times \mathcal{F}$ are chosen in such a way as in (3.2) so that reduced models and obviously non-identifiable mixtures can be excluded. When $\sigma_1 = \sigma_2$, (3.1) is reduced to the two-component location-shifted mixture model (1.2) which has been examined separately in literature. Without the restriction $\mu_1 < \mu_2$, clearly a permutation applied to $(\pi_i, \mu_i, \sigma_i)^\top$, $i = 1, 2$, does not change the mixture $h_{\theta,f}$ (label-switching problem) and thus $h_{\theta,f}$ is not identifiable. When $\pi_i = 0$ or 1, (3.1) has only one component and thus is not a mixture model. When $\pi_i = 0.5$, Hunter et al. (2007) has presented serval examples for two-component location-shifted model to be non-identifiable. Without the constraint of the each component density function has standard deviation 1, we have $\pi_1 \frac{1}{\sigma_1} f(\frac{x-\mu_1}{\sigma_1}) + \pi_2 \frac{1}{\sigma_2} f(\frac{x-\mu_2}{\sigma_2}) = \pi_1 \tilde{f}(x - \mu_1) + \pi_2 \frac{1}{\sigma_2/\sigma_1} \tilde{f}(\frac{x-\mu_2}{\sigma_2/\sigma_1})$ with $\tilde{f}(x) = \frac{1}{\sigma_1} f(\frac{x}{\sigma_1})$. Thus both $(\pi_1, \pi_2, \mu_1, \mu_2, \sigma_1, \sigma_2)^\top$ with f and $(\pi_1, \pi_2, \mu_1, \mu_2, 1, \sigma_2/\sigma_1)^\top$ with $\tilde{f}$ give exactly the same mixture model and thus non-identifiable. The symmetry assumption on f is based on the identifiability condition from Bordes et al. (2006) for two-component semiparametric location-shifted mixture, thus we also assume the symmetry condition satisfied for model (3.1).

In this thesis, we assume f for the two component is the same, and each component varies from different location and scale parameters. Recall that the construction of MHDE in semiparametric models requires an appropriate base function f. Thus we propose an estimation of the nonparametric nuisance parameter f based on which an MHDE is then constructed. We present an iterative algorithm to facilitate the calculation of the MHDE. Since $\mathcal{F}$ and Θ are constructed by excluding certain non-identifiable cases, these constraints in $\mathcal{F}$ and Θ can be viewed as necessary conditions for identifiability. However, the complete set of sufficient and necessary conditions for the identifiability of (3.1) has yet to be fully established.

3.3.1 Base function f in terms of mixture

For the two-component location-shifted mixture model (1.2), Bordes et al. (2006) recovered the base function f in terms of the mixture h and location parameters μ_i's. Now for the two-component location-scale mixture model (3.1), we will recover below the base function f with similar technique from Bordes et al. (2006). Replacing x with $\sigma_1 x + \mu_1$ in (3.1) gives

$$h(\sigma_1 x + \mu_1) = \frac{\pi_1}{\sigma_1} f(x) + \frac{\pi_2}{\sigma_2} f\left(\frac{\sigma_1}{\sigma_2} x + \frac{1}{\sigma_2}(\mu_1 - \mu_2)\right),$$

and thus

$$f(x) = \frac{\sigma_1}{\pi_1} h(\sigma_1 x + \mu_1) - \frac{\sigma_1}{\pi_1} \frac{\pi_2}{\sigma_2} f\left(\frac{\sigma_1}{\sigma_2} x + \frac{1}{\sigma_2}(\mu_1 - \mu_2)\right). \tag{3.3}$$

In (3.3), replacing x with $\frac{\sigma_1}{\sigma_2}x + \frac{1}{\sigma_2}(\mu_1 - \mu_2)$ gives

$$\begin{aligned} f\left(\frac{\sigma_1}{\sigma_2}x + \frac{1}{\sigma_2}(\mu_1 - \mu_2)\right) = &\frac{\sigma_1}{\pi_1}h\left(\frac{\sigma_1^2}{\sigma_2}x + \frac{\sigma_1}{\sigma_2}(\mu_1 - \mu_2) + \mu_1\right) \\ &- \frac{\sigma_1\pi_2}{\sigma_2\pi_1}f\left(\frac{\sigma_1^2}{\sigma_2^2}x + \frac{\sigma_1}{\sigma_2^2}(\mu_1 - \mu_2) + \frac{1}{\sigma_2}(\mu_1 - \mu_2)\right). \end{aligned}$$

Plugging the above equation into (3.3) gives

$$\begin{aligned} f(x) = &\frac{\sigma_1}{\pi_1}h(\sigma_1 x + \mu_1) - \left(\frac{\sigma_1}{\pi_1}\right)^2 \frac{\pi_2}{\sigma_2}h\left(\frac{\sigma_1^2}{\sigma_2}x + \frac{\sigma_1}{\sigma_2}(\mu_1 - \mu_2) + \mu_1\right) \\ &+ \left(\frac{\sigma_1\pi_2}{\pi_1\sigma_2}\right)^2 f\left(\frac{\sigma_1^2}{\sigma_2^2}x + \frac{\sigma_1}{\sigma_2^2}(\mu_1 - \mu_2) + \frac{1}{\sigma_2}(\mu_1 - \mu_2)\right). \end{aligned} \tag{3.4}$$

By mathematical induction we can easily show the following lemma.

Lemma 1. Under model (3.1), for any $m \geq 1$,

$$\begin{aligned} f(x) = &\frac{\sigma_1}{\pi_1}\sum_{i=0}^{m-1}\left(-\frac{\pi_2\sigma_1}{\pi_1\sigma_2}\right)^i h\left(\sigma_1\left(\frac{\sigma_1}{\sigma_2}\right)^i x + \sum_{j=0}^{i}\left(\frac{\sigma_1}{\sigma_2}\right)^j(\mu_1 - \mu_2) + \mu_2\right) \\ &+ \left(-\frac{\pi_2\sigma_1}{\pi_1\sigma_2}\right)^m f\left(\left(\frac{\sigma_1}{\sigma_2}\right)^m x + \frac{1}{\sigma_2}\sum_{j=0}^{m-1}\left(\frac{\sigma_1}{\sigma_2}\right)^j(\mu_1 - \mu_2)\right). \end{aligned} \tag{3.5}$$

Proof of Lemma 1. When $m = 1$, (3.5) is reduced to (3.3). When $m = 2$, (3.5) is reduced to (3.4). Assume (3.5) holds when $m = t$, i.e.

$$\begin{aligned} f(x) = &\frac{\sigma_1}{\pi_1}\sum_{i=0}^{t-1}\left(-\frac{\pi_2\sigma_1}{\pi_1\sigma_2}\right)^i h\left(\sigma_1\left(\frac{\sigma_1}{\sigma_2}\right)^i x + \sum_{j=0}^{i}\left(\frac{\sigma_1}{\sigma_2}\right)^j(\mu_1 - \mu_2) + \mu_2\right) \\ &+ \left(-\frac{\pi_2\sigma_1}{\pi_1\sigma_2}\right)^t f\left(\left(\frac{\sigma_1}{\sigma_2}\right)^t x + \frac{1}{\sigma_2}\sum_{j=0}^{t-1}\left(\frac{\sigma_1}{\sigma_2}\right)^j(\mu_1 - \mu_2)\right). \end{aligned}$$

We replace x with $(\frac{\sigma_1}{\sigma_2})^t x + \frac{1}{\sigma_2}\sum_{j=0}^{t-1}(\frac{\sigma_1}{\sigma_2})^j(\mu_1-\mu_2)$ in (3.3) to get

$$\begin{aligned}
f\left(\left(\frac{\sigma_1}{\sigma_2}\right)^t x + \frac{1}{\sigma_2}\sum_{j=0}^{t-1}\left(\frac{\sigma_1}{\sigma_2}\right)^j(\mu_1-\mu_2)\right) &= \frac{\sigma_1}{\pi_1}h\left(\sigma_1\left(\frac{\sigma_1}{\sigma_2}\right)^t x + \frac{\sigma_1}{\sigma_2}\sum_{j=0}^{t-1}\left(\frac{\sigma_1}{\sigma_2}\right)^j(\mu_1-\mu_2)+\mu_1\right) \\
&- \frac{\sigma_1\pi_2}{\sigma_2\pi_1}f\left(\left(\frac{\sigma_1}{\sigma_2}\right)^{t+1} x + \frac{\sigma_1}{\sigma_2^2}\sum_{j=0}^{t-1}\left(\frac{\sigma_1}{\sigma_2}\right)^j(\mu_1-\mu_2)+\frac{1}{\sigma_2}(\mu_1-\mu_2)\right).
\end{aligned}$$

Plugging above equation into the one for $m=t$ gives

$$\begin{aligned}
f(x) &= \frac{\sigma_1}{\pi_1}\sum_{i=0}^{t-1}\left(-\frac{\pi_2\sigma_1}{\pi_1\sigma_2}\right)^i h\left(\sigma_1\left(\frac{\sigma_1}{\sigma_2}\right)^i x + \sum_{j=0}^{i}\left(\frac{\sigma_1}{\sigma_2}\right)^j(\mu_1-\mu_2)+\mu_2\right) \\
&\quad + \frac{\sigma_1}{\pi_1}\left(-\frac{\pi_2\sigma_1}{\pi_1\sigma_2}\right)^t h\left(\sigma_1\left(\frac{\sigma_1}{\sigma_2}\right)^t x + \sum_{j=0}^{t}\left(\frac{\sigma_1}{\sigma_2}\right)^j(\mu_1-\mu_2)+\mu_1\right) \\
&\quad + \left(-\frac{\pi_2\sigma_1}{\pi_1\sigma_2}\right)^{t+1} f\left(\left(\frac{\sigma_1}{\sigma_2}\right)^{t+1} x + \frac{1}{\sigma_2}\sum_{j=0}^{t}\left(\frac{\sigma_1}{\sigma_2}\right)^j(\mu_1-\mu_2)+\frac{1}{\sigma_2}(\mu_1-\mu_2)\right) \\
&= \frac{\sigma_1}{\pi_1}\sum_{i=0}^{t}\left(-\frac{\pi_2\sigma_1}{\pi_1\sigma_2}\right)^i h\left(\sigma_1\left(\frac{\sigma_1}{\sigma_2}\right)^i x + \sum_{j=0}^{i}\left(\frac{\sigma_1}{\sigma_2}\right)^j(\mu_1-\mu_2)+\mu_2\right) \\
&\quad + \left(-\frac{\pi_2\sigma_1}{\pi_1\sigma_2}\right)^{t+1} f\left(\left(\frac{\sigma_1}{\sigma_2}\right)^{t+1} x + \frac{1}{\sigma_2}\sum_{j=0}^{t}\left(\frac{\sigma_1}{\sigma_2}\right)^j(\mu_1-\mu_2)\right).
\end{aligned}$$

This completes the proof. □

With this lemma, we can recover the base function f from the mixture in (3.1). The result is summarized in the following theorem.

Theorem 1. *Under model (3.1), we have almost everywhere (a.e.) that*

$$f(x) = \begin{cases} \frac{\sigma_1}{\pi_1}\sum_{i=0}^{\infty}\left(-\frac{\pi_2\sigma_1}{\pi_1\sigma_2}\right)^i h\left(\sigma_1\left(\frac{\sigma_1}{\sigma_2}\right)^i x + \sum_{j=0}^{i}\left(\frac{\sigma_1}{\sigma_2}\right)^j(\mu_1-\mu_2)+\mu_2\right), & \text{if } \pi_1 > \pi_2, \\ \frac{\sigma_2}{\pi_2}\sum_{i=0}^{\infty}\left(-\frac{\pi_1\sigma_2}{\pi_2\sigma_1}\right)^i h\left(\sigma_2\left(\frac{\sigma_2}{\sigma_1}\right)^i x + \sum_{j=0}^{i}\left(\frac{\sigma_2}{\sigma_1}\right)^j(\mu_2-\mu_1)+\mu_1\right), & \text{if } \pi_2 > \pi_1. \end{cases} \tag{3.6}$$

Proof of Theorem 1. First, note that the results of Lemma 1 hold when the labels 1 and 2 are swapped. Without loss of generality, we will consider the case where $\pi_1 > \pi_2$, and the case where $\pi_2 > \pi_1$ follows similarly.

From Lemma 1, for all $m \in \mathbb{N}$ and $x \in \mathbb{R}$, the function $f(x)$ can be expressed as:

$$f(x) = \alpha_m(x) + \beta_m(x),$$

where

$$\alpha_m(x) = \frac{\sigma_1}{\pi_1} \sum_{i=0}^{m-1} \left(-\frac{\pi_2\sigma_1}{\pi_1\sigma_2}\right)^i h\left(\sigma_1 \left(\frac{\sigma_1}{\sigma_2}\right)^i x + \sum_{j=0}^{i} \left(\frac{\sigma_1}{\sigma_2}\right)^j (\mu_1 - \mu_2) + \mu_2\right),$$

and

$$\beta_m(x) = \left(-\frac{\pi_2\sigma_1}{\pi_1\sigma_2}\right)^m f\left(\left(\frac{\sigma_1}{\sigma_2}\right)^m x + \frac{1}{\sigma_2} \sum_{j=0}^{m-1} \left(\frac{\sigma_1}{\sigma_2}\right)^j (\mu_1 - \mu_2)\right).$$

Now, we analyze the behavior of $\beta_m(\cdot)$ as $m \to \infty$. It follows that:

$$\|\beta_m(\cdot)\|_1 \to 0 \quad \text{as} \quad m \to \infty,$$

because of the combined effects of the exponential decay in the coefficient $\left(-\frac{\pi_2\sigma_1}{\pi_1\sigma_2}\right)^m$ and the scaling behavior of the argument of the function f. This ensures that the contribution of $\beta_m(\cdot)$ vanishes as $m \to \infty$. Therefore, it follows that:

$$\|f(\cdot) - \alpha_m(\cdot)\|_1 \to 0 \quad \text{as} \quad m \to \infty.$$

Thus, in L_1,

$$f = lim_{m\to\infty}\alpha_m(\cdot),$$

so that, for almost all x,

$$f(x) = \frac{\sigma_1}{\pi_1} \sum_{i=0}^{\infty} \left(-\frac{\pi_2 \sigma_1}{\pi_1 \sigma_2} \right)^i h \left(\sigma_1 \left(\frac{\sigma_1}{\sigma_2} \right)^i x + \sum_{j=0}^{i} \left(\frac{\sigma_1}{\sigma_2} \right)^j (\mu_1 - \mu_2) + \mu_2 \right).$$

Thus, the result follows. $\square$

3.3.2 Construction of SEMIMHDE and computing algorithm

With the derived base function f given in equation (3.6), if we can estimate h based on data from the mixture along with some preliminary estimates of the parameters, then we can estimate f according to (3.6). Given a random sample $X_1, \ldots, X_n$ from the mixture (3.1), we can estimate $h = h_{\theta,f}$ non-parametrically by kernel density estimator

$$\hat{h}(x) = \frac{1}{nb_n} \sum_{j=1}^{n} K\left(\frac{x - X_j}{b_n}\right), \tag{3.7}$$

where K is a kernel probability density function (p.d.f.), which is a non-negative function, and b_n is a bandwidth parameter satisfying the conditions $b_n \to 0$ and $nb_n \to \infty$ as $n \to \infty$. Common choices for K include the uniform, triangular, biweight, triweight, Epanechnikov, and normal kernels. The normal kernel, $K(x) = \phi(x)$ where ϕ is the standard normal density function, is frequently used due to its advantageous mathematical properties.

Given a consistent initial estimate $\hat{\theta}^{(0)}$, we substitute $\hat{\theta}^{(0)}$ and $\hat{h}$ into (3.6) to obtain an estimate of f, denoted by $\hat{f}$. Note the initial estimate will be introduced in the simulation section. Since (3.6) involves an infinite sum, we apply the criterion that the summation is terminated once the absolute difference between two consecutive terms is smaller than 10^{-6}. The MHDE of the parameter vector θ is then defined as:

$$\hat{\theta}_{MHD} = \arg\min_{\theta \in \Theta} \left\| h_{\theta,\hat{f}}^{\frac{1}{2}} - \hat{h}^{\frac{1}{2}} \right\|, \tag{3.8}$$

where Θ denotes the parameter space and $\hat{h}$ denotes the kernel density estimator which are given in (3.2) and (3.7) respectively. The objective is to minimize the Hellinger distance between $h_{\theta,\hat{f}}$ and $\hat{h}$, defined as:

$$\begin{aligned} d_H^2(h_{\theta,\hat{f}}, \hat{h}) &= \min\left(\left\|h_{\theta,\hat{f}}^{\frac{1}{2}} - \hat{h}^{\frac{1}{2}}\right\|_2^2\right) \\ &= \min\left(\frac{1}{2}\int\left(\sqrt{h_{\theta,\hat{f}}(x)} - \sqrt{\hat{h}(x)}\right)^2 dx\right) \\ &= \min\left(1 - \int\sqrt{h_{\theta,\hat{f}}(x)\hat{h}(x)}dx\right). \end{aligned} \tag{3.9}$$

To minimize the Hellinger distance in (3.9), we equivalently maximize:

$$g(\theta) = \int\sqrt{h_{\theta,\hat{f}}(x)\hat{h}(x)}dx, \tag{3.10}$$

subject to the constraints $\pi_i > 0$ for $i = 1, 2$, $\mu_1 < \mu_2$, and $\sum_{i=1}^{2}\pi_i = 1$. If $\hat{\theta}_{MHD} = (\hat{\pi}_1, \hat{\mu}_1, \hat{\mu}_2, \hat{\sigma}_1, \hat{\sigma}_2)$, the necessary conditions for $\hat{\theta}_{MHD}$ to maximize $g(\theta)$ in the interior of the parameter space are:

$$\frac{\partial}{\partial\phi_i}g(\hat{\theta}) = 0, \quad i = 1, 2, \tag{3.11}$$

and

$$\frac{\partial}{\partial\pi_i}\left(g(\hat{\theta}) + \lambda\left(1 - \sum_{i=1}^{2}\pi_i\right)\right) = 0, \quad i = 1, 2, \tag{3.12}$$

where $\phi_i = (\mu_i, \sigma_i)$ and λ is the Lagrange multiplier, an auxiliary variable used to incorporate the constraint that the sum of the mixing proportions π_i equals 1.

However, as the maximum of $g(\theta)$ may occur on the boundary of the parameter space (e.g., when $\pi_1 \to 0$ or $\mu_1 \to \mu_2$), we must also check the behavior of $g(\theta)$ at the boundary points of the parameter space. In such cases, the first-order conditions (partial derivatives equal to zero) may not apply. Therefore, it is important to evaluate the objective function both in the interior and at the boundary of the parameter space to ensure that the global maximum is identified.

To find $\hat{\theta}_{MHD}$, we solve the following equations for θ:

$$\sqrt{\pi_i} \int \sqrt{\frac{\pi_i f(x|\phi_i)}{\sum_{j=1}^{k} \pi_j f(x|\phi_j)}} \frac{\partial}{\partial \phi_i} \sqrt{f(x|\phi_i)} \sqrt{\hat{h}(x)} dx = 0, \tag{3.13}$$

and

$$\frac{1}{2\sqrt{\pi_i}} \int \sqrt{\frac{\pi_i f(x|\phi_i)}{\sum_{j=1}^{k} \pi_j f(x|\phi_j)}} \sqrt{f(x|\phi_i)} \sqrt{\hat{h}(x)} dx - \lambda = 0, \tag{3.14}$$

for $i = 1, 2$. Explicit solutions to these equations are generally intractable. However, following the rationale behind the HMIX algorithm, equation (3.10) can be rewritten as:

$$\begin{aligned} g(\theta) &= \sum_{i=1}^{k} \sqrt{\pi_i} \int \sqrt{\frac{\pi_i f(x|\phi_i)}{\sum_{j=1}^{k} \pi_j f(x|\phi_j)}} \sqrt{f(x|\phi_i) \hat{h}(x)} dx \\ &= \sum_{i=1}^{k} \sqrt{\pi_i} \int \sqrt{a_i(x) f(x|\phi_i) \hat{h}(x)} dx \\ &= H(A, \theta), \end{aligned} \tag{3.15}$$

where

$$A = (a_1(x), \ldots, a_k(x)), \quad a_i(x) = \frac{\pi_i f(x|\phi_i)}{\sum_{j=1}^{k} \pi_j f(x|\phi_j)}, \tag{3.16}$$

for $i = 1, 2$. Treating A as fixed and maximizing $H(A, \theta)$ subject to parameter constraints gives:

$$\phi_i^\star(x) = \frac{\pi_i f(x|\phi_i)}{\sum_{j=1}^k \pi_j f(x|\phi_j)}. \tag{3.17}$$

We refer to this iterative optimization procedure as the Hellinger Minimization for Mixtures (HMIX) algorithm from Cutler and Cordero-Brana (1996).

For semiparametric mixtures, where the component distributions are not fully specified, we replace the density function $f(x|\phi_i)$ with a base function derived in Theorem 1 (3.6). The resulting Modified Semiparametric Hellinger Minimization for Semiparametric Mixtures (MHMIX) algorithm is outlined as follows:

MHMIX Algorithm:

Given an initial estimate $\theta^c = (\pi_1^c, \mu_1^c, \mu_2^c, \sigma_1^c, \sigma_2^c)^T$ and a function $g(\theta^c)$:

1. For $i = 1, 2$, set
$$a_i^c(x) = \frac{\pi_i^c f(x \mid \phi_i^c)}{\sum_{j=1}^k \pi_j^c f(x \mid \phi_j^c)},$$
and choose $\phi_i^\star$ to maximize the objective function
$$H_i(\phi_i) = \int \sqrt{a_i^c(x) f(x \mid \phi_i) \hat{h}(x)}\, dx.$$

2. Update $\pi_i^\star$ as:
$$\pi_i^\star = \frac{H_i^2(\phi_i^\star)}{\sum_{j=1}^k H_j^2(\phi_j^\star)}, \quad \text{for } i = 1, 2.$$

3. Compute
$$h(\theta^\star) = H(A^\star, \theta^\star) = \sum_{i=1}^k \sqrt{\pi_i^\star} \int \sqrt{a_i^\star(x) f(x \mid \phi_i^\star) \hat{h}(x)}\, dx,$$
where
$$a_i^\star(x) = \frac{\pi_i^\star f(x \mid \phi_i^\star)}{\sum_{j=1}^k \pi_j^\star f(x \mid \phi_j^\star)}.$$

4. If $|g(\theta^\star) - g(\theta^c)| < \epsilon$, terminate the algorithm. Otherwise, set $\theta^c = \theta^\star$ and repeat from

Step 1.

3.3.3 Kernel Density Estimations and Constraints on Parameters

Given that the second component of the objective function in (3.10), denoted as $\hat{h}$, represents a non-parametric kernel density estimate based on the observed data, it is crucial to carefully consider the choice of bandwidth parameter in this estimation process. The selection of an appropriate kernel and bandwidth significantly influences the estimator's performance, as it determines the smoothness and accuracy of the resulting density estimate. In this context, we explore various kernel functions, such as the Gaussian, Epanechnikov, and biweight kernels, which are commonly employed due to their desirable mathematical properties and empirical performance. Additionally, we examine methods for bandwidth selection, including cross-validation and plug-in approaches, to optimize the bias-variance trade-off in the density estimation. The outcomes of these choices are not only pivotal in the context of our theoretical framework but also directly inform the settings used in our simulation studies, where we aim to evaluate the practical implications and robustness of our estimation method under different conditions.

Given that our proposed semiparametric Minimum Hellinger Distance Estimation (SEMIMHDE) method is based on the Hellinger Minimization Mixtures (HMIX) algorithm as described in Cutler and Cordero-Brana (1996), we seek to compare our self-defined parametric MHDE (MHDEP) with the parametric HMIX estimates presented in Cutler and Cordero-Brana (1996). To ensure a parallel and consistent comparison, we first review the kernel density function and bandwidth selection methodologies utilized by Cutler and Cordero-Brana (1996). The following section provides a review of the kernel and bandwidth selection for parametric finite mixture models, drawing on the methodologies outlined in Cutler and Cordero-Brana (1996).

In the context of kernel density estimation as represented by (3.7), Parzen (1962) established that the optimal bandwidth, b_n, which minimizes the integrated mean squared error between the estimator $\hat{h}$ and the true density h, is given by:

$$b_n = \varphi(\omega)\beta(h)n^{-1/5}, \tag{3.18}$$

where

$$\varphi(\omega) = \left[\frac{\int \omega^2(y)dy}{(\int \omega(y)y^2dy)^2}\right]^{-1/5},$$

and

$$\beta(h) = \left[\int \left(\frac{\partial^2 h(x)}{\partial x^2}\right)^2 dx\right]^{-1/5}.$$

For our simulation studies involving parametric estimations, we employ the Epanechnikov kernel, which is known for its efficiency and optimal properties in density estimation. As discussed in Chiu (1991), kernel density estimates using the Epanechnikov kernel and the bandwidth formula (3.18) provide a robust framework for analysis of parametric estimations. Specifically, for a mixture of two normal distributions, the function $\beta(h)$ can be expressed as:

$$\begin{aligned}\beta(h) = \Bigg\{\int & \frac{\pi_1^2}{2\sigma_1^5\sqrt{\pi_1}}\phi(x;\mu_1,\sigma_1^2/2)(z_1^2-1)^2 \\ & + \frac{(1-\pi_1)^2}{2\sigma_2^5\sqrt{\pi_1}}\phi(x;\mu_2,\sigma_2^2/2)(z_2^2-1)^2 \\ & + \frac{2\pi_1(1-\pi_1)}{2\pi_1\sigma_1^3\sigma_2^3}\exp\left\{-\frac{1}{2}(z_1^2+z_2^2)\right\}(z_1^2-1)(z_2^2-1)\Bigg\}^{-1/5},\end{aligned} \tag{3.19}$$

where $\phi(x;\mu,\sigma^2)$ denotes the normal density function with mean μ and variance σ^2, and $z_i = (x-\mu_i)/\sigma_i$ for $i = 1, 2$. For the Epanechnikov kernel, the constant φ is approximately 1.71877.

In the semi-parametric case, the distribution of each component is not explicitly specified, which necessitates a different approach compared to the parametric methods outlined in Cutler and Cordero-Brana (1996), where the distributions are typically assumed to be normal. The parametric selection criteria, including the specific choices of kernel function and bandwidth, are tailored for known distributions like the normal distribution and may not be directly applicable to semi-parametric models where component distributions are unknown or unspecified. To address this, we adopt a Gaussian kernel function for the kernel density estimation in the semi-parametric setting. The Gaussian kernel is chosen due to its favorable properties, including smoothness and infinite support, which make it a robust choice for non-parametric density estimation. The bandwidth b_n in this context is determined using a data-driven approach, calculated as follows:

$$b_n = n^{-1/5} \times S_n, \tag{3.20}$$

where n represents the sample size. The scale parameter S_n is a crucial element in this formula, as it adjusts the bandwidth based on the spread of the data. According to the recommendations by Wu and Karunamuni (2012) which studied the MHD under semiparametric mixture, S_n is computed using:

$$S_n = \frac{1}{0.674} \times \text{median}(|X_{\text{mix}} - \text{median}(X_{\text{mix}})|), \tag{3.21}$$

where X_{mix} denotes the observed mixture data. This method uses the median absolute deviation (MAD) from the median, which is a robust measure of statistical dispersion, particularly resistant to outliers in the data. This approach to bandwidth selection ensures that the kernel density estimator adapts to the underlying data structure, providing a flexible and reliable estimation method in the absence of specific distributional assumptions for the mixture components.

To ensure the sufficient identifiability of the location-scale mixture model, we have imposed specific constraints on the parameters, in line with the settings described in equation (3.2). The constraints are as follows:

1. The mixing proportions satisfy:

$$0 < \pi_1 < 1 \quad \text{and} \quad 0 < \pi_2 < 1,$$

with the additional conditions from Bordes et al. (2006) for identifiability of location-shifted mixture:

$$\pi_1 \neq 0.5 \quad \text{and} \quad \pi_2 \neq 0.5.$$

2. The mean parameters μ_1 and μ_2 are constrained by the range of the observed data, denoted as X_{mix}. Specifically, $(X_{\text{mix}})_{\text{min}}$ represents the minimum value of the observed data, and $(X_{\text{mix}})_{\text{max}}$ represents the maximum value. Therefore, the constraints on μ_1 and μ_2 can be written as:

$$(X_{\text{mix}})_{\text{min}} < \mu_1 < (X_{\text{mix}})_{\text{max}},$$

$$(X_{\text{mix}})_{\text{min}} < \mu_2 < (X_{\text{mix}})_{\text{max}}.$$

This ensures that both μ_1 and μ_2 lie within the range of the observed data, which helps maintain realistic parameter estimates based on the data at hand.

3. To avoid label switching and ensure identifiability, the means are ordered such that:

$$\mu_1 < \mu_2.$$

4. The scale parameters σ_1 and σ_2 are strictly positive and distinct:

$$\sigma_1 > 0,$$

$$\sigma_2 > 0.$$

$$\sigma_1 \neq \sigma_2.$$

Although these conditions do not constitute the full set of sufficient and necessary conditions for the identifiability of two-component semiparametric location-scale mixtures, they remain essential constraints on the parameters. These constraints are crucial for uniquely identifying the components of the mixture model and ensuring that the estimation process produces meaningful and interpretable parameter estimates. The constrained parameter space Θ as defined in (1.4) has shown the non-identifiable example without the constrained conditions, where the conditions can be treated as necessary conditions of identifiability.

3.3.4 Initial Estimate Selection

In this section, we discuss two approaches for determining initial estimates in the context of our estimation procedure: Woodward's initial estimates as introduced in Woodward et al. (1984) and our proposed method, named by "Truncated Initials". The HMIX algorithm from Cutler and Cordero-Brana (1996), which is conceptually similar to the Expectation-Maximization (EM) algorithm, involves solving a sequence of weighted one-component likelihood problems. As with the EM algorithm, establishing convergence can be challenging, and there is no guarantee of reaching a global maximum. Consequently, the choice of initial estimates is crucial. In particular, when dealing with contaminated data, it is essential to utilize robust initial estimates. Woodward et al. (1984) proposed a method for obtaining robust starting values, which we have employed in certain aspects of our simulation studies.

Woodward's Initial Estimates

Woodward's approach for obtaining initial estimates involves an ad hoc quasi-clustering technique that is straightforward to implement. For the purposes of this discussion, we assume, without loss of generality, that $\mu_1 \leq \mu_2$. Woodward's method restricts the possible initial values for the first mixing proportion π_1 to discrete values in the set $\{0.1, 0.2, \ldots, 0.9\}$, denoted by $\tilde{\pi_1}$. For each candidate value of π_1, the sample is divided into two subsamples: $Y_1, Y_2, \ldots, Y_{n_1}$ and $Y_{n_1+1}, Y_{n_1+2}, \ldots, Y_n$, where Y_i is the i-th order statistic and n_1 is np rounded to the nearest integer. The optimal value of $\tilde{\pi_1}$ is determined by maximizing the expression $\tilde{\pi_1}(n-\tilde{\pi_1})(m_1-m_2)^2$, where n is the sample size and m_j denotes the sample median of the jth subsample. Since m_1 is an approximately unbiased estimator of the population mean in the case of symmetric distributions, it serves as the initial estimator for μ_1. Our preliminary estimate of σ_1^2 is derived from the observations to the left of m_1 and is given by:

$$\tilde{\sigma_{1(0)}}^2 = \left(\frac{m_1 - {r_1}^{(0.25)}}{0.6745}\right)^2, \tag{3.22}$$

where ${r_j}^{(0.25)}$ represents the 25th percentile from the jth cluster. Similarly, m_2 is used as the initial estimate for μ_2, and thus $\tilde{\sigma_{2(0)}}^2$ is given by:

$$\tilde{\sigma_{2(0)}}^2 = \left(\frac{{r_2}^{(0.75)} - m_2}{0.6745}\right)^2. \tag{3.23}$$

Truncated Initials

While Woodward's initial estimates generally perform well in many simulation studies, particularly in efficiency studies, their robustness can be compromised in the presence of contaminated data. This limitation becomes evident when the contaminated data significantly deviates from the original dataset, as illustrated in Case 6, Table 3.14, where the bias in the estimated parameters from Woodward's initials is notably large. Therefore, finding a robust initial estimation method is crucial before conducting simulation studies.

We propose a novel method for initial estimation, termed "Truncated Initials", which is designed to enhance robustness. This method is based on a percentile approach to detect potential outliers. Specifically, observations falling outside the interval defined by the ath and bth percentiles are flagged as potential outliers. For example, if our data is contaminated by 10%, we set $a = 5$ and $b = 95$, thereby identifying the 10 percent of data (5 percent on each tail) as potentially contaminated.

The Truncated Initials method involves combining this percentile-based outlier detection with Woodward's approach to generate robust initial estimates. Unlike Woodward's method, which operates solely on the original data irrespective of contamination, the Truncated Initials method adjusts for contamination by truncating the data at the specified percentiles and then applying Woodward's procedure on the truncated dataset.

For the ensuing simulation studies, we employ both initial estimation methods—Woodward's and Truncated Initials—to evaluate their performance in terms of efficiency and robustness. In the distribution study, however, we use only Woodward's initials due to the alterations in data range introduced by the Truncated Initials method.

3.4 Simulation Study

In this section, we present the design and outcomes of our simulation studies, which are categorized into three main aspects: efficiency, robustness, and model misspecification studies. Each aspect addresses different facets of the estimation methods, providing a comprehensive evaluation of their performance under various conditions.

Efficiency Studies

Efficiency studies are conducted to assess the performance of estimation methods under ideal conditions. Here, we consider four cases, each representing a different mixture of normal distributions with varying degrees of overlap between the components. The overlap is defined as the area under the two normal distributions' probability density functions that is common to both components, indicating the difficulty of distinguishing between the components.

The cases are as follows:

Case 1: $0.25 \cdot N(0, \sqrt{2}) + 0.75 \cdot N(2.573, 1)$.

This case involves a mixture of 25% from a normal distribution $N(0, \sqrt{2})$ and 75% from $N(2.573, 1)$. The overlap between the components is 0.10, indicating a non-obvious separation between the distributions. The sample sizes considered are 100, 400, and 800. This setting replicates a scenario where one component is more prevalent than the other, allowing us to observe how the methods perform with imbalanced data.

Case 2: $0.5 \cdot N(0, \sqrt{2}) + 0.5 \cdot N(3.066, 1)$.

In this case, the data comprises a balanced mixture with 50% from $N(0, \sqrt{2})$ and 50% from $N(3.066, 1)$, also with an overlap of 0.10. The sample sizes for this case are 100, 400, and 800. The balanced nature of the mixture provides insight into the methods' performance when neither component dominates. This case is only use for the two types of initials comparison, since $\pi_1 = 0.5$ and $\pi_2 = 0.5$ are not allowed in our constraint parameter space, thus Case 2 is not tested in our simulation studies.

Case 3: $0.75 \cdot N(0, \sqrt{2}) + 0.25 \cdot N(2.964, 1)$.

This case features a mixture of 75% from $N(0, \sqrt{2})$ and 25% from $N(2.964, 1)$, with the same 0.10 overlap. The sample sizes are 100, 400, and 800. This case examines the methods' efficiency when one component is significantly less prevalent, posing a challenge for accurate estimation of the smaller component. The first mixing proportion is larger than 0.5, thus we will use the second case formula in (3.6). This is a kind of close to Case 1, but we can test our base function when $\pi_1 > \pi_2$.

Case 4: $0.25 \cdot N(0, \sqrt{2}) + 0.75 \cdot N(4.203, 1)$.

This scenario involves a mixture of 25% from $N(0, \sqrt{2})$ and 75% from $N(4.203, 1)$ with a smaller overlap of 0.03, indicating a greater degree of separation between the components. Sample sizes of 100, 400, and 800 are used. This case is particularly interesting for studying the methods' ability to handle well-separated distributions, which might simplify the estimation problem. These cases replicate the settings used in a key paper published in Cutler and Cordero-Brana (1996), providing a basis for comparison with previously established results.

Robustness Studies

Robustness studies examine the resilience of the estimation methods in the presence of contaminated data, representing real-world scenarios where outliers or unexpected data points can occur. We focus on two cases where 5% of the data is contaminated, either by being near to or far from the main data distribution.

Case 5: $95\%(0.25 \cdot N(0, \sqrt{2}) + 0.75 \cdot N(4.203, 1)) + 5\% \cdot N(8, 1)$.

In this near-contaminated scenario, the main data is generated from the same distribution as in Case 4, but with an additional 5% contamination from a normal distribution $N(8, 1)$. The overlap with the main distribution remains at 0.03. This setup helps evaluate how estimation methods handle data that is moderately deviant from the expected distribution.

Case 6: $95\%(0.25 \cdot N(0, \sqrt{2}) + 0.75 \cdot N(4.203, 1)) + 5\% \cdot N(20, 1)$.

This case features far contamination, with the 5% contamination component being from

$N(20, 1)$, which is significantly distant from the main data distribution generated as in Case 4. This scenario is particularly challenging, testing the robustness of the methods against severe contamination.

For both Case 5 and Case6, sample sizes of 100, 400, and 800 are examined, with 500 repetitions to ensure the stability and reliability of the results.

Model Misspecification Study

In the model misspecification studies, we assess the performance of the semi-parametric Minimum Hellinger Distance Estimator (SEMIMHDE) in comparison to parametric estimation methods when the data does not conform to the assumption of normality.

Case 7: $0.25 \cdot U(\mathbf{mean} = 0, \mathbf{SD} = \sqrt{2}) + 0.75 \cdot U(\mathbf{mean} = 2.573, \mathbf{SD} = 1)$.

This case involves a mixture of uniform distributions: 25% from $U(-\sqrt{6}, \sqrt{6})$ with a mean of 0 and variance of 2, and 75% from $U(0.841, 4.305)$ with a mean of 2.573 and variance of 1. This setup challenges the methods to perform well without the assumption of normality, testing the flexibility and adaptability of the semi-parametric MHDE. The focus in this study is to determine whether the semi-parametric MHDE can outperform traditional parametric methods, particularly when the underlying component distributions deviate from normality. This evaluation helps to understand the practical advantages of semi-parametric approaches in real-world data scenarios.

Case 8: $0.25G(\text{mean} = 2, \text{SD} = 2) + 0.75G(\text{mean} = 6, \text{SD} = 6)$.

Here, $G(\text{mean} = 2, \text{SD} = 2)$ has a shape parameter $k = 1$ and scale parameter $\theta = 2$, while $G(\text{mean} = 6, \text{SD} = 6)$ has a shape parameter $k = 1$ and scale parameter $\theta = 6$. In addition to Case 7 for Uniform distribution, we are also interested in the data generated from Gamma distribution but parametric estimations assumed Normal distribution. Since the signal noise ratio for Case 8 is: $\text{SNR} = \frac{\mu}{\sigma} = \frac{2}{2} = 1$, thus the Gamma distribution is skewed which is distinct from Normal distribution.

3.4.1 Comparison of Initial Estimates for Cases of Efficiency Study

Case 1 : Comparison of Initial Estimates $0.25 * N(0, \sqrt{2}) + 0.75 * N(2.573, 1)$, overlap=0.10, sample size=100, reps=500

Table 3.1: Case 1 Estimated Bias based on Woodward's and Truncated Initials, Overlap=0.10, sample size =100, reps=500

	π_1	μ_1	σ_1	μ_2	σ_2
Woodward's Initials Bias	$-.007$	$-.347$	$-.443$	.042	$-.063$
Truncated Initials Bias	.031	.280	$-.587$	.043	$-.250$

Table 3.2: Case 1 Estimated Bias based on Woodward's and Truncated Initials, Overlap=0.10, sample size =400, reps=500

	π_1	μ_1	σ_1	μ_2	σ_2
Woodward's Initials Bias	$-.030$	$-.318$	$-.272$	$-.018$	.002
Truncated Initials Bias	$-.020$	$-.123$	$-.351$	.035	$-.028$

Table 3.3: Case 1 Estimated Bias based on Woodward's and Truncated Initials, Overlap=0.10, sample size =800, reps=500

	π_1	μ_1	σ_1	μ_2	σ_2
Woodward's Initials Bias	$-.039$	$-.329$	$-.213$	$-.034$	.006
Truncated Initials Bias	$-.049$	$-.298$	$-.271$	$-.033$	.007

From Tables 3.1 to 3.3, we observe that the initial estimates obtained using Woodward's initials and Truncated Initials are similar for Case 1.

Case 2 : Comparison of Initial Estimates $0.5 * N(0, \sqrt{2}) + 0.5 * N(3.066, 1)$, overlap=0.10, sample size=100, reps=500

Table 3.4: Case2 Initials Comparison, Overlap=0.10, sample size =100, reps=500

	π_1	μ_1	σ_1	μ_2	σ_2
Woodward's Initials Bias	$-.089$	$-.431$	$-.356$	$-.126$	$-.006$
Truncated Initials Bias	$-.074$	$-.165$	$-.531$	$-.226$	$-.176$

Table 3.5: Case2 Initials Comparison, Overlap=0.10, sample size =400, reps=500

	π_1	μ_1	σ_1	μ_2	σ_2
Woodward's Initials Bias	$-.090$	$-.361$	$-.231$	$-.156$	$-.057$
Truncated Initials Bias	$-.108$	$-.350$	$-.299$	$-.166$	$.065$

Table 3.6: Case2 Initials Comparison, Overlap=0.10, sample size =800, reps=500

	π_1	μ_1	σ_1	μ_2	σ_2
Woodward's Initials Bias	$-.093$	$-.353$	$-.199$	$-.170$	$.068$
Truncated Initials Bias	$-.091$	$-.265$	$-.166$	$-.146$	$.056$

Tables 3.4 to 3.6 indicate that there is no significant difference between the initial estimates obtained using Woodward's and Truncated Initials for Case 2. It is important to note that in Case 2, the mixing proportions are equal ($\pi_1 = \pi_2$). Our derived base function only accommodates scenarios where $\pi_1 < \pi_2$ or $\pi_1 > \pi_2$. Consequently, we did not conduct efficiency studies for Case 2, as the SEMIMHDE method is not applicable in this context.

Case 3 : Comparison of Initial Estimates $0.75 * N(0, \sqrt{2}) + 0.25 * N(2.964, 1)$, overlap=0.10, sample size=100, reps=500

Table 3.7: Case 3 Initials Comparison, Overlap=0.10, sample size =100, reps=500

	π_1	μ_1	σ_1	μ_2	σ_2
Woodward's Initials Bias	−.166	−.460	−.306	−.427	.021
Truncated Initials Bias	−.173	−.393	−.517	−.740	−.146

Table 3.8: Case 3 Initials Comparison, Overlap=0.10, sample size =400, reps=500

	π_1	μ_1	σ_1	μ_2	σ_2
Woodward's Initials Bias	−.155	−.399	−.234	−.450	.141
Truncated Initials Bias	−.166	−.350	−.210	−.444	.140

Table 3.9: Case 3 Initials Comparison, Overlap=0.10, sample size =800, reps=500

	π_1	μ_1	σ_1	μ_2	σ_2
Woodward's Initials Bias	−.148	−.366	−.193	−.438	.154
Truncated Initials Bias	−.146	−.316	−.236	−.413	.139

Tables 3.7 to 3.9 reveal that the estimates obtained using Woodward's and Truncated Initials are very similar for Case 3.

Case 4 : Comparison of Initial Estimates $0.25 * N(0, \sqrt{2}) + 0.75 * N(4.203, 1)$, overlap=0.03, sample size=100, reps=500

Table 3.10: Case 4 Initials Comparison, Overlap=0.03, sample size =100, reps=500

	π_1	μ_1	σ_1	μ_2	σ_2
Woodward's Initials Bias	−.166	−.460	−.306	−.427	.021
Truncated Initials Bias	−.173	−.393	−.517	−.740	−.146

Table 3.11: Case 4 Initials Comparison, Overlap=0.03, sample size =400, reps=500

	π_1	μ_1	σ_1	μ_2	σ_2
Woodward's Initials Bias	.006	−.005	−.070	.014	−.008
Truncated Initials Bias	.008	.147	−.132	.053	−.028

Table 3.12: Case 4 Initials Comparison, Overlap=0.03, sample size =800, reps=500

	π_1	μ_1	σ_1	μ_2	σ_2
Woodward's Initials Bias	.010	.040	−.006	.019	−.013
Truncated Initials Bias	−.030	−.144	−.169	−.031	.014

Tables 3.10 to 3.12 indicate that, in the case with a smaller overlap, Woodward's initial estimates perform slightly better than the Truncated Initials.

3.4.2 Comparison of Initial Estimates for Cases of Robustness Study

Case 5: Robustness with 5% Contamination For this case, the data is comprised of 95% from the mixture $0.25 \cdot N(0, \sqrt{2}) + 0.75 \cdot N(4.203, 1)$ and 5% from $N(8, 1)$, with an overlap of 0.03. The sample size is 100, and the simulation was repeated 500 times. Initial estimates are derived based on the contaminated data.

Table 3.13: Case 5: Comparison of Initial Estimates, Overlap=0.03, Sample Size=100, Repetitions=500

	π_1	μ_1	σ_1	μ_2	σ_2
Woodward's Initials Bias	0.019	-0.009	-0.179	0.145	0.090
Truncated Initials Bias	0.003	0.424	-0.444	0.052	-0.088

From Table 3.13, it is observed that when the contaminated data is relatively close to the original data distribution, Woodward's initial estimates exhibit slightly better performance compared to the Truncated Initials. Consequently, for Case 5 in the robustness studies, we will rely on Woodward's initial estimates based on the non-contaminated data.

Case 6: Robustness with 5% Substantial Contamination In this scenario, the data consists of 95% from the mixture $0.25 \cdot N(0, \sqrt{2}) + 0.75 \cdot N(4.203, 1)$ and 5% from $N(20, 1)$, with an overlap of 0.03. The sample size is 100, and the simulation was conducted with 500 repetitions. Initial estimates are derived based on the contaminated data, noting that this case includes larger outliers compared to Case 5.

As shown in Table 3.14, for Case 6, where the contamination is more pronounced and the data includes outliers far from the original data distribution, the Truncated Initials demonstrate substantially better performance compared to Woodward's Initials. Specifically, for the parameter μ_2, Woodward's initial estimates are significantly influenced by the

Table 3.14: Case 6: Comparison of Initial Estimates, Overlap=0.03, Sample Size=100, Repetitions=500

	π_1	μ_1	σ_1	μ_2	σ_2
Woodward's Initials Bias	0.371	2.052	0.762	7.405	1.748
Truncated Initials Bias	0.012	0.471	-0.428	0.248	-0.087

contaminated data $5\% \cdot N(20, 1)$. Therefore, for robustness studies involving substantial contamination, we recommend using the Truncated Initials.

3.4.3 Comparison of Initial Estimates for Cases of Model Misspecification Study

Case 7: Comparison of Initial Estimates In this case, the data is generated from the mixture $0.25 \cdot U(\textbf{mean} = 0, \textbf{SD} = \sqrt{2}) + 0.75 \cdot U(\textbf{mean} = 2.573, \textbf{SD} = 1)$ with a sample size of 100 and 500 repetitions. The analysis focuses on evaluating the performance of the initial estimates derived from Woodward's method and the proposed Truncated Initials.

Table 3.15: Case 7: Comparison of Initial Estimates, Sample Size=100, Repetitions=500

	π_1	μ_1	σ_1	μ_2	σ_2
Woodward's Initials Bias	-0.019	-0.599	-0.438	-0.005	-0.213
Truncated Initials Bias	0.142	0.535	-0.694	0.262	-0.211

From the results presented in Table 3.15, it is evident that Woodward's initial estimates outperform the Truncated Initials for Case 7. The Truncated Initials method, which involves removing data from the tails, significantly affects the distribution's shape. This alteration can lead to less accurate initial estimates for parameters. Therefore, for distribution studies, where preserving the full data range is crucial for accurate estimation, Woodward's initial estimates are preferred. This approach maintains the integrity of the data distribution and provides more reliable parameter estimates.

3.4.4 Efficiency Studies

For the efficiency studies, we will concentrate on three specific cases: Case 1, Case 3, and Case 4, as outlined in the previous section. Case 2 is excluded from this analysis due to the fact that its mixing proportion is set at 0.5. This particular proportion is not covered by our base function, as specified in Equation (3.6). The rationale for excluding Case 2 is based on the fact that our base function is designed to handle mixing proportions within a specific range, and Case 2 falls outside this range. By focusing on Cases 1, 3, and 4, we ensure that our efficiency studies are conducted within the parameters for which our base function is applicable, thereby maintaining the relevance and accuracy of our analysis.

Case 1 use Woodward's Initials $0.25 * N(0, \sqrt{2}) + 0.75 * N(2.573, 1)$, initials estimates are based on Woodward's initials, overlap=0.10, sample size=100/400/800, reps=500.

Table 3.16: Case 1 Efficiency Study Estimations Comparison, Using Woodward's Initials Overlap=0.10, sample size =100, reps=500

Estimations		π_1	μ_1	σ_1	μ_2	σ_2
MLE	Bias	.005	−.253	−.193	−.012	−.018
	Variance	.022	.744	.148	.052	.027
	MSE	.022	.808	.186	.052	.028
HMIX	Bias	−.014	−.275	−.238	−.019	.038
	Variance	.014	.544	.098	.045	.022
	MSE	.014	.620	.155	.045	.023
MHDEP	Bias	−.004	−.319	−.267	−.019	.033
	Variance	.024	.814	.132	.068	.034
	MSE	.024	.916	.203	.069	.035
SEMIMHDE	Bias	−.063	.042	−.686	−.146	−.230
	Variance	.007	.546	.053	.053	.068
	MSE	.011	.548	.524	.074	.091

Table 3.17: Case 1 Efficiency Study Estimations Comparison, Using Woodward's Initials, Overlap=0.10, sample size =400, reps=500

Estimations		π_1	μ_1	σ_1	μ_2	σ_2
MLE	Bias	$-.001$	$-.095$	$-.079$	$-.019$	$-.003$
	Variance	.008	.346	.067	.016	.007
	MSE	.008	.355	.073	.017	.007
HMIX	Bias	$-.031$	$-.246$	$-.134$	$-.041$	.051
	Variance	.003	.158	.030	.012	.005
	MSE	.004	.219	.048	.014	.008
MHDEP	Bias	$-.012$	$-.016$	$-.109$	$-.027$	.043
	Variance	.007	.303	.054	.016	.006
	MSE	.007	.329	.065	.017	.008
SEMIMHDE	Bias	$-.054$	$-.291$	$-.374$	$-.062$	$-.003$
	Variance	.002	.145	.034	.011	.014
	MSE	.005	.230	.174	.015	.014

Table 3.18: Case 1 Efficiency Study Estimations Comparison, Using Woodward's Initials, Overlap=0.10, sample size =800, reps=500

Estimations		π_1	μ_1	σ_1	μ_2	σ_2
MLE	Bias	.002	$-.032$	$-.028$	$-.013$	$-.000$
	Variance	.005	.199	.039	.009	.004
	MSE	.005	.200	.040	.009	.004
HMIX	Bias	$-.032$	$-.232$	$-.107$	$-.043$	.047
	Variance	.001	.070	.016	.005	.002
	MSE	.008	.124	.027	.007	.004
MHDEP	Bias	$-.011$	$-.111$	$-.063$	$-.022$	.035
	Variance	.004	.163	.031	.008	.003
	MSE	.004	.176	.035	.009	.005
SEMIMHDE	Bias	$-.050$	$-.253$	$-.292$	$-.064$	$-.019$
	Variance	.002	.096	.021	.006	.010
	MSE	.005	.016	.106	.010	.010

From the analysis of Table 3.16 through Table 3.18, which pertains to Case 1 with a larger overlap, it is evident that the HMIX method exhibits the highest efficiency where HMIX shows smallest variance and bias. Among the methods evaluated, HMIX outperforms the others in terms of efficiency for this case. In comparison, the MLE and the MHDEP demonstrate similar levels of efficiency. Both methods perform comparably, though they do not achieve the efficiency of HMIX. On the other hand, the semiparametric approach, SEMIMHDE, shows the lowest efficiency among the methods considered. This result highlights the limitations of the SEMIMHDE in handling cases with larger overlaps, where the efficiency of the semiparametric method lags behind that of the parametric methods. Overall, these findings underscore the relative strengths and weaknesses of the different estimation methods in the context of cases with larger overlap, with HMIX emerging as the most efficient and SEMIMHDE as the least efficient.

Given that the SEMIMHDE exhibits noticeably poorer performance compared to the parametric estimations, we are motivated to investigate whether alternative initial choices might improve efficiency. To address this, we will explore different initialization strategies in the upcoming section. Specifically, we will present a comparative analysis of efficiency for Case 1, focusing on the performance of truncated initials. This examination aims to determine if selecting a different initial can enhance the efficiency of the SEMIMHDE and potentially address its observed shortcomings.

Case 1 use Truncated Initials $0.25 * N(0, \sqrt{2}) + 0.75 * N(2.573, 1)$, initials estimates are based on Truncated initials, overlap=0.10, sample size=100/400/800, reps=500.

Table 3.19: Case 1 Efficiency Study Estimations Comparison, Using Truncated Initials Overlap=0.10, sample size =100, reps=500

Estimations		π_1	μ_1	σ_1	μ_2	σ_2
MLE	Bias	.013	−.204	−.176	−.003	−.025
	Variance	.023	.766	.151	.054	.028
	MSE	.023	.808	.182	.054	.029
HMIX	Bias	.038	.039	−.133	.042	.005
	Variance	.018	.646	.095	.047	.022
	MSE	.019	.647	.112	.048	.022
MHDEP	Bias	.003	−.275	−.250	−.008	.025
	Variance	.025	.819	.124	.068	.036
	MSE	.025	.894	.187	.068	.036
SEMIMHDE	Bias	.012	.166	−.570	−.036	.188
	Variance	.013	.615	.058	.057	.061
	MSE	.013	.643	.384	.058	.097

Table 3.20: Case 1 Efficiency Study Estimations Comparison, Using Truncated Initials, Overlap=0.10, sample size =400, reps=500

Estimations/Parameters		π_1	μ_1	σ_1	μ_2	σ_2
MLE	Bias	.003	−.072	−.070	−.017	−.002
	Variance	.008	.356	.069	.016	.007
	MSE	.008	.361	.074	.016	.007
HMIX	Bias	−.000	−.041	−.064	−.008	.033
	Variance	.004	.235	.042	.010	.004
	MSE	.004	.237	.046	.010	.005
MHDEP	Bias	−.011	−.151	−.108	−.026	.041
	Variance	.007	.309	.055	.016	.006
	MSE	.007	.332	.066	.017	.008
SEMIMHDE	Bias	.007	.090	−.261	−.015	−.077
	Variance	.008	.305	.040	.012	.024
	MSE	.008	.313	.109	.012	.030

Table 3.21: Case 1 Efficiency Study Estimations Comparison, Using Truncated Initials, Overlap=0.10, sample size =800, reps=500

Estimations/Parameters		π_1	μ_1	σ_1	μ_2	σ_2
MLE	Bias	.003	−.029	−.027	−.012	−.001
	Variance	.005	.201	.040	.009	.004
	MSE	.005	.201	.040	.009	.004
HMIX	Bias	−.008	−.078	−.052	−.017	.033
	Variance	.003	.143	.026	.006	.002
	MSE	.003	.149	.028	.007	.003
MHDEP	Bias	−.011	−.109	−.063	−.022	.034
	Variance	.004	.165	.032	.008	.003
	MSE	.004	.177	.036	.009	.005
SEMIMHDE	Bias	−.014	−.026	−.204	−.033	−.052
	Variance	.003	.172	.035	.009	.010
	MSE	.003	.172	.076	.010	.013

The results presented in Table 3.19 through Table 3.21 indicate that HMIX achieves the highest efficiency among the methods evaluated. In comparison, the Parametric MHDE (MHDEP) approach and the Maximum Likelihood Estimation (MLE) method exhibit similar performance levels, though they do not reach the efficiency observed for the HMIX algorithm. On the other hand, the semiparametric SEMIMHDE approach demonstrates the lowest efficiency, lagging behind the parametric methods.

A comparative analysis of these results with those presented in Table 3.19 through Table 3.21 reveals an interesting observation: the SEMIMHDE method exhibits smaller bias when using truncated initials as opposed to Woodward's initials. Specifically, the use of truncated initials appears to mitigate the bias associated with the SEMIMHDE estimates. This improvement is particularly evident when examining the results for μ_1. When employing Woodward's initials, an increase in sample size from 100 to 400, as seen in Table 3.16 and Table 3.17, results in a paradoxical increase in bias for the SEMIMHDE estimates of μ_1. This inconsistency suggests a potential robustness issue with Woodward's initials. In contrast, the use of truncated initials does not exhibit this inconsistency. The SEMIMHDE estimates remain stable and consistent across varying sample sizes when using truncated initials, indicating a more robust performance. This robustness is critical for ensuring reliable parameter estimates, particularly in cases involving larger overlap.

Therefore, based on these findings, we can confidently conclude that truncated initials provide a more robust alternative for Case 1 efficiency studies with larger overlap. The improved consistency and reduced bias associated with truncated initials make them a preferable choice for enhancing the reliability and accuracy of the SEMIMHDE method in such scenarios.

Case 3 use Woodward's Initials $0.75 * N(0, \sqrt{2}) + 0.25 * N(2.964, 1)$, initials estimates are based on Woodward's initials, overlap=0.10, sample size=100/400/800, reps=500.

Table 3.22: Case 3 Efficiency Study Estimations Comparison, Using Woodward's Initials Overlap=0.10, sample size =100, reps=500

Estimations		π_1	μ_1	σ_1	μ_2	σ_2
MLE	Bias	−.091	−.219	−.116	−.260	.054
	Variance	.033	.232	.065	.494	.102
	MSE	.041	.280	.078	.561	.105
HMIX	Bias	−.121	−.301	−.111	.378	.147
	Variance	.022	.151	.037	.348	.069
	MSE	.037	.242	.049	.491	.091
MHDEP	Bias	−.079	−.213	−.080	−.158	.045
	Variance	.040	.316	.073	.564	.093
	MSE	.046	.361	.080	.589	.095
SEMIMHDE	Bias	−.133	−.212	−.405	−.700	−.159
	Variance	.013	.152	.123	.355	.063
	MSE	.031	.197	.287	.845	.088

Table 3.23: Case 3 Efficiency Study Estimations Comparison, Using Woodward's Initials, Overlap=0.10, sample size =400, reps=500

Estimations		π_1	μ_1	σ_1	μ_2	σ_2
MLE	Bias	−.023	−.055	−.027	−.057	−.004
	Variance	.011	.079	.020	.163	.039
	MSE	.012	.082	.021	.166	.039
HMIX	Bias	−.109	−.266	−.085	−.388	.189
	Variance	.007	.043	.009	.118	.024
	MSE	.019	.114	.016	.269	.060
MHDEP	Bias	−.026	−.063	.002	−.062	.050
	Variance	.010	.075	.018	.154	.030
	MSE	.011	.079	.018	.157	.033
SEMIMHDE	Bias	−.147	−.354	−.254	−.560	.062
	Variance	.007	.041	.028	.148	.026
	MSE	.029	.167	.093	.462	.030

Table 3.24: Case 3 Efficiency Study Estimations Comparison, Using Woodward's Initials, Overlap=0.10, sample size =800, reps=500

Estimations		π_1	μ_1	σ_1	μ_2	σ_2
MLE	Bias	−.010	−.023	−.011	−.026	−.003
	Variance	.004	.033	.010	.066	.016
	MSE	.004	.033	.010	.067	.016
HMIX	Bias	−.100	−.247	−.079	−.366	.186
	Variance	.004	.027	.006	.066	.014
	MSE	.014	.088	.012	.200	.049
MHDEP	Bias	−.014	−.033	.014	−.037	.046
	Variance	.004	.033	.009	.066	.014
	MSE	.004	.034	.010	.067	.016
SEMIMHDE	Bias	−.150	−.365	−.204	−.543	.138
	Variance	.004	.028	.017	.073	.015
	MSE	.027	.161	.059	.368	.034

From the analysis of Table 3.22 through Table 3.24, which pertains to Case 3 where $\pi_1 > \pi_2$, it is evident that the MLE exhibits the best performance among the estimation methods evaluated. Following MLE, the Parametric MHDE (MHDEP) demonstrates the next best performance, while the HMIX algorithm ranks third. The semiparametric SEMIMHDE method performs the worst in this scenario.

SEMIMHDE has large bias on location parameters μ_1 and μ_2, especially on μ_2. This heightened bias can be attributed to the smaller mixing proportion of the second component in Case 3, which makes it more challenging to accurately estimate the second location parameter. The difficulty in estimating μ_2 underscores the limitations of the SEMIMHDE method in situations where the components of the mixture have significantly different mixing proportions.

Case 3 use Truncated Initials $0.75 * N(0, \sqrt{2}) + 0.25 * N(2.964, 1)$, initials estimates are based on Truncated initials, overlap=0.10, sample size=100/400/800, reps=500.

Table 3.25: Case 3 Efficiency Study Estimations Comparison, Using Truncated Initials Overlap=0.10, sample size =100, reps=500

Estimations		π_1	μ_1	σ_1	μ_2	σ_2
MLE	Bias	−.086	−.206	−.114	−.239	.046
	Variance	.034	.232	.066	.482	.100
	MSE	.041	.274	.079	.539	.102
HMIX	Bias	−.097	−.252	−.091	−.299	.121
	Variance	.021	.160	.037	.314	.067
	MSE	.031	.224	.046	.403	.082
MHDEP	Bias	−.073	−.193	−.070	−.142	.041
	Variance	.038	.283	.070	.540	.093
	MSE	.043	.320	.075	.560	.095
SEMIMHDE	Bias	−.120	−.230	−.305	−.593	−.125
	Variance	.013	.112	.119	.344	.065
	MSE	.028	.165	.212	.695	.080

Table 3.26: Case 3 Efficiency Study Estimations Comparison, Using Truncated Initials, Overlap=0.10, sample size =400, reps=500

Estimations		π_1	μ_1	σ_1	μ_2	σ_2
MLE	Bias	−.023	−.054	−.027	−.055	−.005
	Variance	.011	.079	.020	.163	.039
	MSE	.012	.082	.021	.166	.039
HMIX	Bias	−.088	−.218	−.064	−.311	.160
	Variance	.007	.042	.010	.111	.023
	MSE	.014	.089	.014	.207	.048
MHDEP	Bias	−.026	−.063	.003	−.062	.052
	Variance	.010	.075	.018	.153	.030
	MSE	.010	.079	.018	.157	.033
SEMIMHDE	Bias	−.130	−.333	−.211	−.479	.054
	Variance	.006	.037	.028	.126	.024
	MSE	.023	.141	.072	.356	.027

Table 3.27: Case 3 Efficiency Study Estimations Comparison, Using Truncated Initials, Overlap=0.10, sample size =800, reps=500

Estimations		π_1	μ_1	σ_1	μ_2	σ_2
MLE	Bias	−.010	−.022	−.011	−.026	−.003
	Variance	.004	.033	.010	.066	.016
	MSE	.004	.033	.010	.066	.016
HMIX	Bias	−.082	−.205	−.062	−.304	.159
	Variance	.004	.025	.006	.060	.013
	MSE	.011	.067	.010	.152	.038
MHDEP	Bias	−.014	−.033	.014	−.036	.046
	Variance	.005	.033	.009	.066	.014
	MSE	.004	.034	.010	.067	.016
SEMIMHDE	Bias	−.126	−.312	−.165	−.455	.100
	Variance	.004	.025	.019	.075	.017
	MSE	.020	.123	.046	.282	.027

From all 3 cases efficiency studies, by comparing the results of using Woodward's initials and truncated initials, we can see the results of using truncated initials has better results. Thus our proposed truncated initial estimate is robust for different cases efficiency studies.

From the analysis of Table 3.25 through Table 3.27, it is evident that the performance ranking from best to worst among the estimation methods is as follows: Parametric Minimum Hellinger Distance Estimation (MHDEP), Maximum Likelihood Estimation (MLE), HMIX, and then semiparametric SEMIMHDE. This ranking mirrors the results obtained when using Woodward's initials.

Notably, all four estimation methods exhibit smaller biases when using truncated initials compared to Woodward's initials. Consequently, we will adopt the truncated initials for the efficiency studies in Case 3. Examining the efficiency studies across all three cases and comparing the results obtained using Woodward's initials with those using truncated initials, it is apparent that the truncated initials yield superior results. Thus, our proposed truncated initial estimate demonstrates improved performance across different case efficiency studies.

Case 4 use Woodward's Initials The data distribution defined as:

$$0.25 \cdot N(0, \sqrt{2}) + 0.75 \cdot N(4.203, 1).$$

For this case, the initial estimates are derived using Woodward's initials. The overlap between the components is 0.03, which is smaller compared to the overlap observed in Case 1. The sample sizes considered for this analysis are 100, 400, and 800, with each scenario subjected to 500 repetitions. This smaller overlap in Case 4, as compared to Case 1, allows for a more distinct separation between the components of the mixture model. By evaluating the performance across different sample sizes, we can assess the scalability and robustness of Woodward's initials in providing reliable parameter estimates under varying data conditions.

Table 3.28: Case 4 Efficiency Study Estimations Comparison, Using Woodward's Initials, Overlap=0.03, sample size =100, reps=500

Estimations		π_1	μ_1	σ_1	μ_2	σ_2
MLE	Bias	.005	−.002	−.025	.006	−.005
	Variance	.003	.240	.092	.022	.012
	MSE	.003	.240	.093	.022	.012
HMIX	Bias	−.004	−.009	−.071	−.009	.050
	Variance	.003	.221	.071	.021	.014
	MSE	.003	.221	.076	.022	.016
MHDEP	Bias	−.003	−.005	−.070	−.009	.049
	Variance	.003	.246	.078	.022	.014
	MSE	.003	.246	.083	.022	.017
SEMIMHDE	Bias	− .024	.188	−.382	−.079	−.123
	Variance	.003	.312	.060	.030	.032
	MSE	.004	.347	.206	.037	.047

Table 3.29: Case 4 Efficiency Study Estimations Comparison, Using Woodward's Initials, Overlap=0.03, sample size =400, reps=500

Estimations		π_1	μ_1	σ_1	μ_2	σ_2
MLE	Bias	.001	.020	−.004	.002	−.001
	Variance	.001	.069	.029	.006	.003
	MSE	.001	.070	.029	.006	.003
HMIX	Bias	−.003	.009	−.018	−.007	.014
	Variance	.008	.060	.023	.006	.004
	MSE	.008	.060	.023	.006	.006
MHDEP	Bias	− .002	.011	−.017	−.006	.040
	Variance	.001	.065	.024	.006	.004
	MSE	.001	.065	.025	.006	.005
SEMIMHDE	Bias	− .012	.085	−.157	−.037	−.023
	Variance	.002	.151	.046	.018	.016
	MSE	.002	.158	.071	.020	.017

Table 3.30: Case 4 Efficiency Study Estimations Comparison, Overlap=0.03, Using Woodward's Initials, sample size =800, reps=500

Estimations		π_1	μ_1	σ_1	μ_2	σ_2
MLE	Bias	.001	.007	−.001	.000	−.001
	Variance	.000	.030	.014	.003	.001
	MSE	.000	.030	.014	.003	.001
HMIX	Bias	−.001	.006	−.004	−.004	.030
	Variance	.000	.028	.012	.003	.001
	MSE	.000	.028	.012	.003	.002
MHDEP	Bias	− .002	.003	−.006	−.005	.031
	Variance	.000	.028	.013	.003	.001
	MSE	.000	.028	.013	.003	.002
SEMIMHDE	Bias	−.002	.146	−.066	−.002	−.039
	Variance	.001	.108	.049	.011	.007
	MSE	.001	.130	.054	.011	.009

From Table 3.28 to Table 3.30, the parametric estimates MLE, HMIX, MHDEP have similar efficiency performance. Each of these parametric methods provides comparable performance metrics, reflecting their effective utilization of the underlying data structure. In contrast, the semiparametric estimation approach, specifically the SEMIMHDE, shows slightly reduced efficiency relative to the parametric methods. This observation is consistent with expectations, given that parametric estimates incorporate additional information about the distribution of each component in the mixture model. The parametric methods leverage this detailed information to enhance their efficiency and accuracy. The slightly lower efficiency of the SEMIMHDE can be attributed to its semi-parametric nature, which involves a less comprehensive modeling of the component distributions compared to the fully parametric approaches. Despite this, the SEMIMHDE still provides valuable insights and remains a viable option depending on the specific requirements of the analysis.

Case 4 use Truncated Initials $0.25 * N(0, \sqrt{2}) + 0.75 * N(4.203, 1)$, initials estimates are based on Truncated initials, overlap=0.03, sample size=100/400/800, reps=500, this is to compare the efficiency performance for Case 4 based on different initials.

Table 3.31: Case 4 Efficiency Study Estimations Comparison, Using Truncated Initials Overlap=0.03, sample size =100, reps=500

Estimations		π_1	μ_1	σ_1	μ_2	σ_2
MLE	Bias	.004	−.008	−.030	.004	−.004
	Variance	.003	.251	.095	.022	.013
	MSE	.003	.251	.095	.022	.013
HMIX	Bias	−.004	−.013	−.076	−.011	.053
	Variance	.003	.243	.077	.022	.014
	MSE	.003	.243	.083	.022	.017
MHDEP	Bias	−.003	−.006	−.072	−.010	.051
	Variance	.003	.247	.081	.022	.014
	MSE	.003	.247	.086	.022	.017
SEMIMHDE	Bias	−.035	.114	−.424	−.122	−.103
	Variance	.004	.464	.088	.042	.040
	MSE	.005	.477	.267	.057	.050

Table 3.32: Case 4 Efficiency Study Estimations Comparison, Using Truncated Initials, Overlap=0.03, sample size =400, reps=500

Estimations		π_1	μ_1	σ_1	μ_2	σ_2
MLE	Bias	.001	.020	−.004	.002	−.001
	Variance	.001	.069	.029	.006	.003
	MSE	.001	.070	.029	.006	.003
HMIX	Bias	−.004	−.005	−.025	−.010	.044
	Variance	.001	.061	.023	.006	.003
	MSE	.001	.061	.023	.006	.005
MHDEP	Bias	−.002	.011	−.016	−.006	.042
	Variance	.001	.065	.024	.006	.003
	MSE	.001	.065	.025	.006	.005
SEMIMHDE	Bias	−.036	−.108	−.271	−.097	.004
	Variance	.001	.113	.056	.012	.014
	MSE	.003	.125	.130	.022	.014

Table 3.33: Case 4 Efficiency Study Estimations Comparison, Using Truncated Initials, Overlap=0.03, sample size =800, reps=500

Estimations		π_1	μ_1	σ_1	μ_2	σ_2
MLE	Bias	.001	.007	−.001	.000	−.001
	Variance	.000	.030	.014	.003	.001
	MSE	.000	.030	.014	.003	.001
HMIX	Bias	−.004	−.017	−.017	−.009	.035
	Variance	.000	.026	.012	.003	.001
	MSE	.000	.026	.012	.003	.003
MHDEP	Bias	−.002	.002	−.005	−.005	.032
	Variance	.000	.028	.013	.003	.001
	MSE	.000	.028	.013	.003	.002
SEMIMHDE	Bias	−.034	−.111	−.233	−.089	−.003
	Variance	.001	.066	.038	.007	.008
	MSE	.002	.078	.092	.015	.008

The results presented from Table 3.31 to Table 3.33 are similar as those observed from Table 3.28 to Table 3.30. This consistency indicates that the truncated initials proposed in our study demonstrate robust performance. They function comparably to Woodward's initials in terms of efficiency for cases involving well-separated mixtures. The similarity in results across these tables reinforces the effectiveness of the truncated initials in efficiently handling well-separated mixture models.

3.4.5 Robustness Studies

Case 5 Robustness Studies using Woodward's initials

In light of the initial comparison presented in the previous section, which demonstrates that Woodward's initials outperform the truncated initials for Case 5, we have chosen to employ Woodward's initials for the robustness studies pertaining to this case. The data, inclusive of contamination, is defined as follows:

- Case 5: $0.95 * (0.25 * N(0, \sqrt{2}) + 0.75 * N(4.202, 1)) + 0.05 * N(8, 1)$.

The initial estimates utilized in this study are derived from the initial estimates proposed by Woodward et al. (1984). It is noteworthy that the probability parameter p is constrained within the interval [0.1, 0.9]. These initial estimates are computed based on data that is not contaminated. The specific initial estimates vector is:

$$\begin{pmatrix} 0.26900, & -0.0320, & 1.20593, & 4.25656, & 0.93833 \end{pmatrix}.$$

The bias of initials vector is:

$$\begin{pmatrix} 0.01900, & -0.03209, & -0.20828, & 0.05356, & -0.06167 \end{pmatrix}.$$

From the results summarized in Table 3.34, we observe three parametric estimations: Maximum Likelihood Estimation (MLE), the HMIX algorithm as reproduced from the Cutler and Cordero-Brana (1996) paper, and Parametric MHDE (MHDEP) approach. Additionally, we examine the semiparametric MHDE based on the HMIX algorithm. The robustness performance, as evidenced by these methods, reveals that the MLE exhibits the poorest robustness performance among the compared methodologies. Conversely, all three estimations based on the Hellinger distance display superior robustness performance. Notably, our semiparametric MHDE demonstrates even better robustness performance compared to the MLE.

Table 3.34: Case 5 Robustness Studies, Estimated Bias, Variance and MSE based on 500 samples, n=100,Overlap=0.03, 5% contamination, non-contaminated initials

Estimations		π_1	μ_1	σ_1	μ_2	σ_2
MLE	Bias	−.064	−.478	−.250	.062	.510
	Variance	.003	.190	.077	.026	.055
	MSE	.007	.419	.140	.030	.315
HMIX	Bias	−.044	−.280	−.208	.028	.344
	Variance	.004	.324	.105	.023	.058
	MSE	.006	.402	.148	.024	.176
MHDEP	Bias	−.051	−.358	− .181	.013	.364
	Variance	.004	.255	.082	.025	.063
	MSE	.007	.383	.115	.025	.196
SEMIMHDE	Bias	−.056	.169	− .234	−.016	−.049
	Variance	.003	.310	.068	.022	.043
	MSE	.006	.339	.123	.022	.045

Case 6 Robustness Studies Using Truncated Initials with Contaminated Data

In this subsection, we explore the robustness of parameter estimates using truncated initials in the presence of data contamination. Specifically, we consider Case 6, which involves a dataset with 5% contamination. The data distribution for Case 6 is described as follows:

$$95\% \left(0.25 \cdot N(0, \sqrt{2}) + 0.75 \cdot N(4.203, 1)\right) + 5\% \cdot N(20, 1).$$

For this study, the initial estimates are derived from the contaminated data. The overlap between the components is 0.03, and the sample size is set to 100 with 500 repetitions.

It is important to note that we are using the truncated initials based on contaminated data for this analysis. The choice of truncated initials is motivated by their superior robustness compared to Woodward's initials, particularly in scenarios where the outliers are significantly distant from the main data distribution. This robustness ensures more accurate parameter estimates even in the presence of extreme values, thereby enhancing the reliability of the resulting statistical inferences.

Through this robustness study, we aim to demonstrate the effectiveness of truncated initials in handling data contamination and to provide empirical evidence supporting their use in practical applications where data contamination is a concern.

Table 3.35: Case 6 Robustness Study Estimations Comparison, Overlap=0.03, Using Truncated Initials, sample size =100, reps=500

Estimations		π_1	μ_1	σ_1	μ_2	σ_2
MLE	Bias	.358	3.073	2.076	5.443	1.729
	Variance	.083	1.277	5.502	53.872	8.786
	MSE	.211	10.720	9.812	83.493	11.775
HMIX	Bias	.003	2.071	−.098	.168	.065
	Variance	.010	.430	.085	2.579	.444
	MSE	.010	.435	.094	2.607	.448
MHDEP	Bias	.006	.058	−.089	.196	.220
	Variance	.010	.434	.110	3.199	6.358
	MSE	.010	.438	.118	3.237	6.406
SEMIMHDE	Bias	− .053	.423	−.541	.017	.012
	Variance	.005	.536	.116	.042	1.200
	MSE	.008	.715	.409	.043	1.200

Table 3.36: Case 6 Robustness Study Estimations Comparison, Overlap=0.03, Using Truncated Initials,sample size =400, reps=500

Estimations		π_1	μ_1	σ_1	μ_2	σ_2
MLE	Bias	.267	3.110	2.537	3.630	1.528
	Variance	.084	1.741	7.037	44.374	6.783
	MSE	.155	11.415	13.474	57.550	9.117
HMIX	Bias	−.005	.002	−.057	−.011	.018
	Variance	.001	.057	.022	.005	.003
	MSE	.001	.057	.025	.006	.004
MHDEP	Bias	− .003	.013	−.023	−.006	.036
	Variance	.001	.067	.025	.006	.004
	MSE	.001	.067	.026	.006	.005
SEMIMHDE	Bias	− .054	.302	−.362	−.012	−.006
	Variance	.001	.086	.057	.010	.014
	MSE	.004	.177	.188	.010	.014

Table 3.37: Case 6 Robustness Study Estimations Comparison, Overlap=0.03, Using Truncated Initials, sample size =800, reps=500

Estimations		π_1	μ_1	σ_1	μ_2	σ_2
MLE	Bias	.216	3.303	3.150	2.103	1.457
	Variance	.062	1.483	7.470	29.525	6.554
	MSE	.109	12.388	17.394	33.949	8.676
HMIX	Bias	$-.006$	$-.025$	$-.051$	$-.013$	.017
	Variance	.000	.024	.011	.003	.001
	MSE	.000	.024	.011	.003	.001
MHDEP	Bias	$-\ .002$	.003	$-.011$	$-.005$	.029
	Variance	.000	.030	.014	.003	.002
	MSE	.000	.030	.014	.003	.002
SEMIMHDE	Bias	$-\ .052$	.295	$-.372$	$-.012$	$-.018$
	Variance	.001	.056	.033	.005	.008
	MSE	.003	.143	.171	.005	.009

From the results presented in Table 3.35 through Table 3.37, it is evident that the Maximum Likelihood Estimation (MLE) method exhibits the poorest performance in the presence of outliers, particularly when these outliers are significantly distant from the original data distribution. This lack of robustness is a critical limitation of the MLE approach, as it indicates a high sensitivity to anomalous data points which can severely skew the estimation results. In contrast, the other three estimation Hellienger Distance based- estimations demonstrate a remarkable robustness to outliers. Regardless of the outliers' distance from the original data distribution, these estimators maintain a high level of accuracy and reliability. The robustness of these alternative methods suggests their superior suitability for practical applications where data contamination is a concern. The HMIX algorithm, MHDEP, and semiparametric MHDE are able to effectively mitigate the impact of outliers, ensuring that the parameter estimates remain stable and accurate. This robustness is particularly advantageous in real-world scenarios where data quality can be compromised by the presence of extreme values or anomalies.

In summary, the comparative analysis clearly underscores the limitations of MLE in handling outliers and highlights the robustness and reliability of the alternative estimation methods. These findings provide a strong empirical basis for preferring the HMIX algorithm, MHDEP, and semiparametric MHDE in situations where data contamination is a significant concern.

3.4.6 Model Misspecification Study

In this section, we aim to examine the performance of various estimation methods when the data are generated from a Uniform distribution, yet the parametric model assumes a Normal density. Specifically, we compare the results of the semiparametric SEMIMHDE method with three other parametric estimation methods: HMIX, MLE and MHDEP. We utilize the parameter settings from Case 1, generating each component from a Uniform distribution as follows:

Case 7: $0.25 \cdot U(\mathbf{mean} = 0, \mathbf{SD} = \sqrt{2}) + 0.75 \cdot U(\mathbf{mean} = 2.573, \mathbf{SD} = 1)$.

Here, $U(\mathbf{mean} = 0, \mathbf{SD} = \sqrt{2})$ represents a Uniform distribution with a mean of 0 and a variance of 2 over the interval $[-\sqrt{6}, \sqrt{6}]$. Similarly, $U(\mathbf{mean} = 2.573, \mathbf{SD} = 1)$ represents a Uniform distribution with a mean of 2.573 and a variance of 1 over the interval $[0.841, 4.305]$. For the parametric methods HMIX, MHDEP, and MLE we assume that the data follow Normal distributions with means and variances equivalent to those of the Uniform distributions. In contrast, the SEMIMHDE method makes no such distributional assumption, providing a semiparametric approach.

The results of the four estimation methods are presented in the following Table 3.38.

Table 3.38: Case 7 Model Misspecification Studies, Estimated Bias and Variance based on 500 samples, p=0.25,n=100,Overlap=0.10, Uniform data

Estimations		π_1	μ_1	σ_1	μ_2	σ_2
MLE	Bias	−.065	−1.020	−.661	−.101	.043
	Variance	.039	.912	.074	.173	.050
	MSE	.043	1.952	.511	.183	.052
HMIX	Bias	−.043	−.819	−.524	−.083	.080
	Variance	.038	1.084	.075	.179	.054
	MSE	.040	1.755	.350	.186	.060
MHDEP	Bias	−.034	−.882	− .575	−.072	.082
	Variance	.052	1.306	.080	.238	.105
	MSE	.053	2.084	.411	.243	.112
SEMIMHDE	Bias	−.066	−.436	− .589	−.229	.070
	Variance	.018	.803	.112	.043	.119
	MSE	.022	.993	.459	.095	.124

From Table 3.38, it is evident that the SEMIMHDE method outperforms the three parametric methods, particularly in the estimation of the parameter μ_1. Furthermore, SEMIMHDE has smallest MSE for parameter π_1, μ_1 and μ_2 compared to three other parametric estimations.

This scenario presents a model misspecification problem. Our findings indicate that the semiparametric SEMIMHDE method demonstrates the best performance among the estimation methods considered. This superior performance underscores the advantage of using semiparametric methods over parametric methods. In practical data analysis, the true underlying distribution of the data is often unknown. Parametric estimation methods rely heavily on the assumption that the chosen model accurately represents the true distribution. When this assumption is violated, as in the case of model misspecification, the performance of parametric methods can degrade significantly. In contrast, semiparametric estimation methods, such as SEMIMHDE, do not require strict assumptions about the underlying distribution. As a result, these methods can maintain robust performance even in the absence of correct distributional assumptions. This robustness makes semiparametric methods particularly valuable in practical applications where the true distribution is uncertain or complex.

In addition to the Uniform distribution from Case 7, we are also interested in examining data generated from a Gamma distribution while assuming a parametric Normal distribution. The Gamma distribution is known to approximate an exponential distribution when the shape parameter approaches 1, which is quite different from a Normal distribution. Therefore, we selected the following parameter settings for Case 8:

Case 8: $0.25G(\text{mean} = 2, \text{SD} = 2) + 0.75G(\text{mean} = 6, \text{SD} = 6)$.

Here, $G(\text{mean} = 2, \text{SD} = 2)$ has a shape parameter $k = 1$ and scale parameter $\theta = 2$, while $G(\text{mean} = 6, \text{SD} = 6)$ has a shape parameter $k = 1$ and scale parameter $\theta = 6$.

In this case, each component approximates an exponential distribution, and there is approximately 63% overlap between the components. Compared to previous cases, Case 8 is more complex due to the lack of separation between the two components.

Table 3.39: Case 8 Model Misspecification Studies, Estimated Bias and Variance based on 500 samples, p=0.25, n=100, Overlap=0.63, Gamma distribution

Estimations		π_1	μ_1	σ_1	μ_2	σ_2
MLE	Bias	0.408	0.266	-0.281	4.628	0.370
	Variance	0.013	0.327	0.238	4.489	2.523
	MSE	0.179	0.398	0.317	25.91	2.660
HMIX	Bias	0.517	0.699	0.056	5.862	-1.515
	Variance	0.008	0.394	0.249	5.581	1.170
	MSE	0.275	0.883	0.252	39.944	3.465
MHDEP	Bias	0.379	0.061	-0.404	3.189	-1.34
	Variance	0.010	0.298	0.172	3.563	0.888
	MSE	0.153	0.302	0.335	13.730	2.682
SEMIMHDE	Bias	0.087	0.289	-0.681	3.597	-1.003
	Variance	0.027	0.328	0.121	1.297	2.317
	MSE	0.035	0.412	0.589	14.24	3.323

Table 3.39 shows that the bias and variance results for the four estimators are generally

poor for both the location and scale parameters. This is due to the large overlap (63%) between the two components in Case 8 which makes it complicate to estimate. Among the four estimators, SEMIMHDE demonstrates smaller bias compared to the other three parametric estimators on the first mixing proportion π_1. The biases of π_1 from MLE, HMIX, and MHDEP are very large, which are far away from the true value for π_1, this is due to the incorrect normality assumption for these three parametric estimations. Note that SEMIMHDE assumes symmetry in each component, which affects its performance in this case compared to Case 7, thus the estimates for two locations and two scales are not as well as Case 7, however, it still performs well in estimating the mixing proportion. This result underscores the limitations of parametric estimators under model misspecification, while our semiparametric MHDE remains more reliable when the true distribution is unknown.

3.5 Old Faithful Geyser Data Analysis

The Old Faithful Geyser data set is a well-known and extensively studied data set in statistical literature. This dataset contains eruption durations and waiting times between eruptions of the Old Faithful Geyser in Yellowstone National Park, notable for its bimodal waiting times, making it suitable for mixture model analysis.

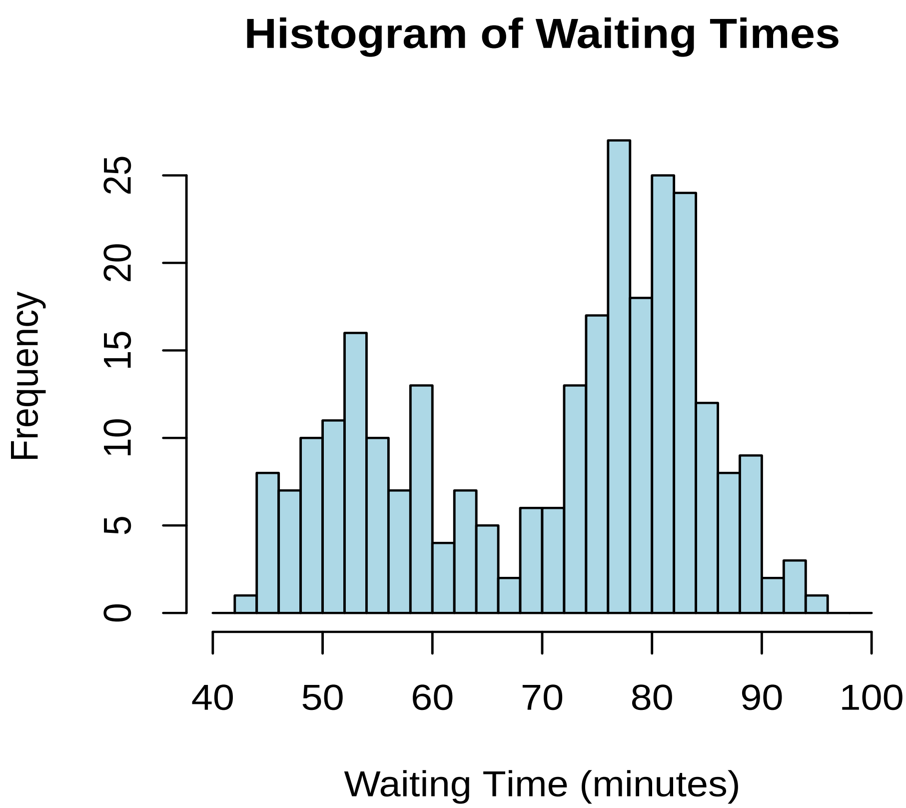


Figure 3.1: Histogram of Waiting Time

In my research, we focus on the waiting times between eruptions of the Old Faithful Geyser. The primary objective is to investigate the efficiency and robustness of various estimation techniques through comparative evaluations. This involves applying and assessing different methods to estimate the parameters of the underlying distribution of the waiting times. The estimation techniques examined in my study include Minimum Hellinger Distance Estimation for Finite Mixture (HMIX), parametric Minimum Hellinger Distance Estimation (MHDEP), Maximum Likelihood Estimation (MLE), and Semi-Parametric Minimum Hellinger Distance Estimation (SEMIMHDE). Each of these methods offers unique advantages and challenges, and my research aims to provide a comprehensive comparison of their performance in terms of both efficiency and robustness. Specifically, the HMIX algorithm is widely used for estimating parameters in finite mixture models due to its iterative nature and convergence properties. The MHDEP method, a parametric approach, focuses on minimizing the Hellinger distance between the empirical distribution and the model distribution, providing a robust alternative to MLE. MLE is a classical and extensively used estimation method known for its asymptotic efficiency under correct model specifications. However, its performance can deteriorate under model misspecification. The SEMIMHDE method combines the strengths of both parametric and non-parametric approaches, offering robustness without relying heavily on strict distributional assumptions. By applying these methods to the waiting times data from the Old Faithful Geyser, my research aims to determine which estimation technique provides the most reliable and accurate results. This analysis is crucial for understanding the underlying distribution of the waiting times and for making informed decisions based on the data.

3.5.1 Estimations Comparisons on Old Faithful Geyser Data

Since initial values play a crucial role in parametric estimations such as the HMIX algorithm, we have tested Woodward's initials based on the Old Faithful Geyser data. The Woodward's initial values used are as follows:

$$\text{ini.vector} = (0.4\,,\,55\,,\,5.93\,,\,81\,,\,4.45)^T.$$

Based on the Woodward's initials, the four estimations are presented in Table 3.41. When using Woodward's initials, it becomes apparent that the HMIX does not give good result of the parameter σ_1. Ideally, the parameter σ_1 should yield values similar to those obtained from the Minimum Hellinger Distance Estimation Parametric (MHDEP) method, as both HMIX and MHDEP are based on the Minimum Hellinger Distance criterion. Additionally, the SEMIMHDE method, when initialized with Woodward's values, shows a decrease in the estimates of σ_1 and σ_2 with each iteration. This results in unrealistically small values for σ_1 and σ_2, as evidenced by the fitted density plots from previous analyses. These unsatisfactory results necessitate the identification of more suitable initial values.

Given the promising results from the SEMIMHDE method for location-shifted models as demonstrated in Wu and Zhou (2018), which assumes equal scale parameters, we have decided to use the SEMIMHDE results as our initial values. The new initial values from SEMIMHDE are:

$$\text{ini.vector} = (0.358, 54.7, 5.87, 80, 5.87)^T.$$

From now on, we will use these SEMIMHDE-derived initial values for both efficiency and robustness studies. We have tested various estimation methods—including HMIX, MHDEP, SEMIMHDE, Maximum Likelihood Estimation (MLE) for location-scale models, MHDE for location-shift models, and MLE for location-shift models—on the waiting time data from the Old Faithful Geyser using SEMIMHDE-derived initials. The estimations are shown in Table 3.40. The NA values in Table 3.40 and Table 3.41 for the MLE of the L_2 distance arise because MLE is based on the log-likelihood objective function, not the Hellinger distance. As a result, the MLE estimates for the L_2 distance are not available. Additionally, the estimates for σ_1 and σ_2 in both the SEMIMHDE (location-shift) and MLE (location-shift) methods are unavailable due to the location-shifted mixture model, where the scale parameters are known and are the same for both components.

By comparing Table 3.40 with Table 3.41, we can tell all estimations have achieved better efficiency performance since the objective function L_2 distance in the last column are smaller when using the SEMIMHDE for location-shifted as initials, as well as the estimates for scale parameters for each estimation can better fit the data histogram. This initial choice particularly change the performance of SEMIMHDE for location-scale mixture, which can be found in SEMIMHDE parameters σ_1 and σ_2 in two tables. At the same time, it is evident that all six estimation methods yield very similar results for π_1, μ_1, and μ_2. However, the two scale parameters σ_1 and σ_2 exhibit slight differences across the six estimation methods, particularly in the SEMIMHDE method. Notably, the first three rows, representing the HMIX, MHDEP, and SEMIMHDE methods, show differing estimates for the two scale parameters, which contradicts the assumptions of the location-shifted model as outlined in Wu and Zhou (2018). Thus, it is crucial to employ location-scale mixture models to investigate inference on estimations for real data, as the majority of real-world data exhibits varying scale parameters. By considering models that account for both location and scale differences, we can achieve more accurate and reliable estimates that better reflect the inherent variability present in the data. This approach ensures that the models are robust and applicable to a wide range of practical scenarios where scale parameters differ across components.

Table 3.40: Estimations Comparisons on Old Faithful Data using SEMIMHDE for location-shift mixture as initials

Estimations	π_1	μ_1	σ_1	μ_2	σ_2	L_2 Distance
HMIX(location-scale)	0.351	55.008	5.672	79.916	5.992	0.387
MHDEP(location-scale)	0.355	54.600	5.817	80.020	6.033	0.290
SEMIMHDE(location-scale)	0.330	54.880	5.830	80.040	5.880	0.398
MLE(location-scale)	0.361	54.619	5.896	80.092	5.878	NA
SEMIMHDE(location-shift)	0.358	54.700	NA	80.050	NA	NA
MLE(location-shift)	0.361	54.61	NA	80.09	NA	NA

Table 3.41: Estimations Comparisons on Old Faithful Data using Woodward's initials

Estimations	π_1	μ_1	σ_1	μ_2	σ_2	L_2 Distance
HMIX(location-scale)	0.351	54.879	5.587	79.906	5.940	0.392
MHDEP(location-scale)	0.355	54.600	5.817	80.020	6.033	0.290
SEMIMHDE(location-scale)	0.365	55.977	4.719	80.309	4.286	0.466
MLE(location-scale)	0.361	54.619	5.896	80.092	5.878	NA
SEMIMHDE(location-shift)	0.358	54.700	NA	80.050	NA	NA
MLE(location-shift)	0.361	54.61	NA	80.09	NA	NA

3.5.2 Robustness Study on Old Faithful Geyser Data

We are also interested in evaluating the robustness performance based on the Old Faithful real data. To this end, we introduced five identical outliers with the value "200" into the dataset and recalculated the estimates, as reported in the (3.42). It is evident that the first three rows, corresponding to the HMIX, MHDEP, and SEMIMHDE methods, are not significantly influenced by the presence of outliers. The estimates produced by these methods remain almost identical to those obtained from the efficiency studies. However, for the Maximum Likelihood Estimation (MLE) method, from the estimations on mixing proportion and location parameters, we can tell MLE was obviously influenced by outliers. MLE has wrongly put the outliers of five identical 200 as the second component and the non-outlier part as the first-component. This indicates that the MLE method is affected by the presence of outliers.

The robustness of the HMIX, MHDEP, and SEMIMHDE methods can be attributed to their basis in Hellinger distance estimation. These methods demonstrate superior robustness compared to MLE when the data contains outliers. Maximum Likelihood Estimation (MLE) lacks robustness compared to Hellinger distance-based estimations. This discrepancy arises primarily from the fundamental differences in how these methods handle deviations from the assumed model, particularly in the presence of outliers. MLE operates by maximizing the likelihood function, which directly depends on the probability density function of the assumed model. As a result, MLE is highly sensitive to deviations from this model, including the presence of outliers. Outliers can disproportionately influence the likelihood function, leading to biased estimates of the parameters. This sensitivity is particularly problematic in real-world data, which often contain anomalies or noise that do not conform to the assumed model. In contrast, Hellinger distance-based estimations, such as HMIX, MHDEP, and SEMIMHDE, offer enhanced robustness. The Hellinger distance is a measure of the difference between two probability distributions. Unlike likelihood-based methods, which can be heavily influenced by outliers, the Hellinger distance provides a more resilient measure of divergence. It effectively reduces the impact of outliers by focusing on the overall shape and distribution of the data rather than individual data points. Overall, all Hellinger-distance based estimations show very strong robustness properties when data includes outliers, however MLE is not robust at all.

Table 3.42: Old Faithful Data Robustness Studies by adding 5 identical value of 200 based on Different Estimations Using Truncated Initials

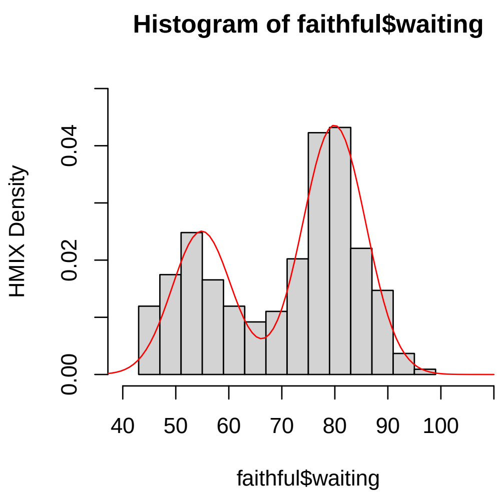
Estimations	π_1	μ_1	σ_1	μ_2	σ_2
HMIX	0.353	55.041	5.694	80.100	6.004
MHDEP	0.355	54.587	5.837	80.009	6.065
SEMIMHDE	0.333	55.434	5.681	80.058	5.798
MLE	0.982	80.897	13.578	200.00	3.387

3.5.3 Comparison among Different Estimations Fitted Density Plots

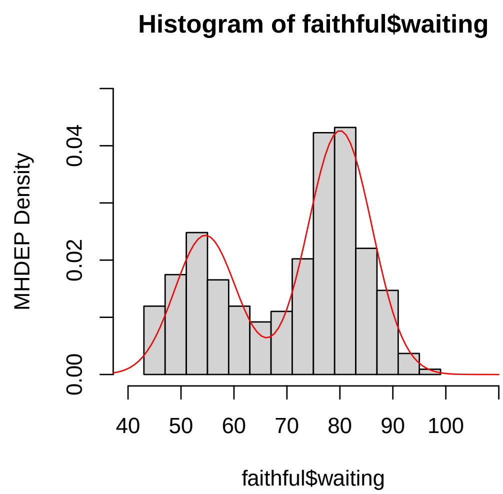
In this study, we generated fitted density plots based on the estimations and overlaid them on the histograms of the data, following the approach from Wu and Zhou (2018). The fitted densities were created using four estimation methods: HMIX, MHDEP, SEMIMHDE, and MLE. These plots offer a visual comparison of how well each method approximates the underlying data distribution. We used the results from Table 3.40 which use the SEMIMHDE location-shifted model in Wu and Zhou (2018) as the initial values for the estimations. From the density plots, we observed that the parametric methods (HMIX, MHDEP, and MLE) fit the histograms well, while the semiparametric method (SEMIMHDE) did not fit as closely. This is likely due to the smaller scale parameter estimates in the SEMIMHDE, resulting in a steeper bell curve compared to the other methods.

From the density plots in Figure 3.2, it is evident that the parametric estimation methods—HMIX, MHDEP, and MLE—fit the histogram data quite accurately. Each of these methods assumes a normal distribution for the data distribution, which appears to align closely with the observed data. This close fit can be attributed to the parametric nature of these estimations, which leverage strong assumptions about the data distribution to achieve precise parameter estimates. In contrast, the semiparametric MHDE method (SEMIMHDE) does not fit the histogram as well as the parametric methods. The density curve obtained from SEMIMHDE is notably steeper, resulting in a sharper bell curve. This discrepancy arises primarily from the differences in scale parameter estimation. For SEMIMHDE, the estimated scale parameters are smaller than those obtained through the parametric methods. Consequently, the density function generated by SEMIMHDE exhibits a steeper gradient. The semiparametric approach offers flexibility by not strictly adhering to parametric assumptions, but this comes at the cost of increased variability in the scale parameter estimates. While SEMIMHDE may provide robustness in the presence of outliers and model misspecification, its performance in terms of fitting the data distribution may vary. This trade-off between flexibility and fit highlights the nuanced differences between parametric and semiparametric estimation methods.

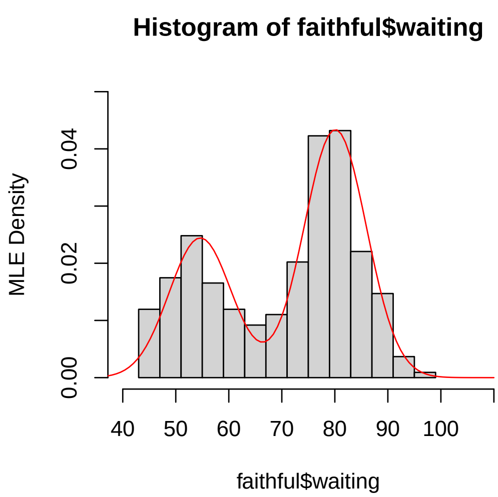
Overall, these density plots underscore the strengths and limitations of each estimation technique. The parametric methods, with their strong distributional assumptions, tend to produce more accurate fits to the histogram data. However, the semiparametric approach offers valuable robustness and flexibility, albeit with some compromise in the precision of the fit. This analysis emphasizes the importance of selecting an appropriate estimation method based on the specific characteristics and requirements of the data.



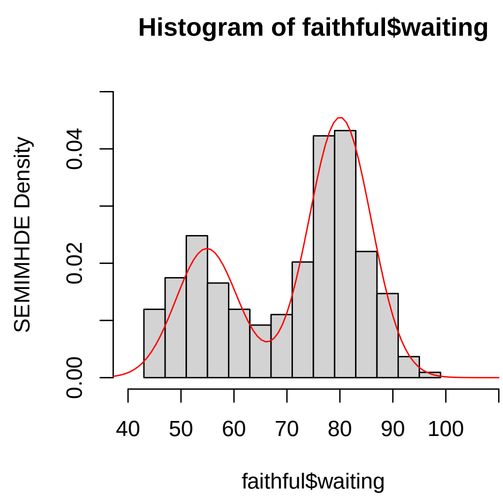
(a) HMIX Fitted Density Plot on Histogram



(b) MHDEP Fitted Density Plot on Histogram



(c) MLE Fitted Density Plot on Histogram



(d) SEMIMHDE Fitted Density Plot on Histogram

Figure 3.2: Comparison among 4 Estimations Fitted Density Plots on Histogram of Old Faithful Data base on location-shift SEMIMHDE from Wu and Zhou (2018) as initials

Chapter 4

MHDE for Semiparametric Location-Scale Mixture Model with Right-Censored Data

4.1 Introduction

Location-scale mixture models are powerful tools in statistical analysis, often used to model data that arise from multiple underlying distributions. These models assume that the observed data are generated from a mixture of distributions, each characterized by its own location (mean) and scale (variance) parameters. Formally, a location-scale mixture model can be represented as:

$$f(x;\theta) = \sum_{k=1}^{K} \pi_k \phi\left(\frac{x-\mu_k}{\sigma_k}\right),$$

where ϕ denotes the probability density function of the component distributions, μ_k and σ_k are the location and scale parameters for the k-th component, and π_k are the mixing proportions with $\sum_{k=1}^{K} \pi_k = 1$, and θ is the parameter vector of ϕ, μ and σ. These models are widely used due to their flexibility in capturing a variety of distributional shapes and their applicability in diverse fields such as finance, biology, and engineering. The foundation of location-scale mixture models dates back to early works by Pearson (1894), who introduced the method of moments for mixture distributions. Over the decades, various methods have been developed for estimating the parameters of these models, with MLE being the most prominent due to its optimality properties under regular conditions Dempster et al. (1977). Parametric approaches, such as the Weibull, exponential, and log-normal models, assume specific distributions for survival times. These models can provide more efficient estimations when the assumed distribution aligns well with the data. Nonparametric location-scale mixture models were studied by Ghosal et al. (1999). For semiparametric mixture, continuous scale mixture models were considered by Xiang et al. (2016), this paper explores semiparametric mixtures using continuous scale mixtures, offering a detailed methodological framework and applications in statistical modeling.

Given that Chapter 3 focused on MHDE for a two-component location-scale mixture model with complete data, we now aim to investigate the performance of MHDE when the mixture data are right-censored. Right-censoring is a common occurrence in survival analysis and reliability studies, where the exact value of an observation is unknown but is only known to exceed a certain threshold. Incorporating right-censored data into mixture models complicates the estimation process because the likelihood function must account for the censored observations. The probability density function for data with right-censoring is given by

$$L(\theta; x) = \prod_{i=1}^{n} \left[f(x_i; \theta)^{\delta_i} \, S(x_i; \theta)^{1-\delta_i} \right],$$

where $f(x_i; \theta)$ is the probability density function (PDF) for the i-th observation x_i, given the parameter set θ, and $S(x_i; \theta)$ is the survival function, which represents the probability that the random variable X exceeds x_i. If the observation is uncensored (i.e., we know the exact value x_i), the indicator $\delta_i = 1$, so the likelihood term for that observation is $f(x_i;\theta)$. If the observation is right-censored (i.e., we only know that the true value is greater than x_i), the indicator $\delta_i = 0$, so the likelihood term for that observation is $S(x_i;\theta)$. The inclusion of right-censored data in statistical models has significantly extended the scope of the conventional statistic methods by the non-parametric Kaplan-Meier estimator was introduced by Kaplan and Meier (1958), which has been a fundamental tool in survival analysis and is widely used for analyzing censored data. To address the challenges posed by right-censoring, various methodologies have been developed. These methods aim to handle right-censored data effectively, ensuring that estimations remain accurate and unbiased. One of the foundational methods for dealing with right-censored data is the Kaplan-Meier estimator which can be found in Kaplan and Meier (1958). It is a nonparametric estimator to estimate the survival function from lifetime data, accommodating right-censoring by adjusting the survival probability at each observed event time. Several researchers have extended kernel density estimation techniques to handle censored data. The key challenge is to modify the kernel density estimator to account for the incomplete information due to censoring. Early work by Prentice (1973) and Tanner and Wong (1983) provided methods for incorporating censored observations into the kernel density estimation framework. These approaches generally involve weighting the observations to reflect the censoring mechanism. The ability to handle incomplete data while providing accurate survival probabilities has cemented the Kaplan-Meier estimator's status as a cornerstone in the analysis of right-censored data. The paper by Lachos et al. (2017) introduces a novel approach to finite mixture modeling of censored data using the multivariate Student-t distribution. This method extends the traditional finite mixture models by incorporating the robustness of the Student-t distribution, making it particularly well-suited for data with heavy tails and outliers. The authors propose an EM algorithm for parameter estimation in the presence of right-censored data, leveraging the properties of the Student-t distribution to handle the challenges posed by censoring. This approach not only improves the robustness of the model but also enhances its flexibility in capturing the underlying structure of the data. Their parametric models are advantageous due to their ability to provide comprehensive insights, including hazard and survival functions, but their accuracy is heavily dependent on the correctness of the distributional assumptions as discussed in Lawless (2003). However, these parametric approaches can be limited by their reliance on distributional assumptions. As a result, semi-parametric and non-parametric methods have gained prominence for their flexibility and robustness. The Cox proportional hazards proposed by Cox (1972), for instance, is a semi-parametric method that allows for the inclusion of covariates without making strong assumptions about the baseline hazard function. More advances have seen the development of semi-parametric and non-parametric methods that offer greater flexibility. Dempster et al. (1977) applied the EM algorithm and its extensions have been particularly effective in handling right-censored data without relying heavily on parametric assumptions. Additionally, techniques such as the Nelson-Aalen estimator from Aalen (1978) and Nelson (1972) provide non-parametric estimations of the cumulative hazard function, serving as valuable tools in exploratory data analysis.

Multiple imputation techniques have been employed to handle the uncertainty introduced by right-censoring. These techniques involve generating multiple complete datasets by imputing the censored observations and then combining the results from each dataset to produce final estimates. This approach allows for the incorporation of additional variability due to censoring and enhances the robustness of statistical inferences as discussed in Rubin (1987).

New research has also focused on improving the robustness and efficiency of estimations in the presence of right-censored data. For instance, machine learning algorithms and Bayesian methods have been increasingly applied to survival analysis, offering new perspectives and methodologies for dealing with censored data such as van der Laan and Robins (2003) and

Ibrahim et al. (2001). These approaches leverage computational advancements to provide more accurate and flexible estimations, accommodating complex data structures and relationships. The C-mix model, introduced by Bussy et al. (2019), represents a significant advancement in the analysis of high-dimensional censored data. This model extends traditional mixture models by incorporating censored durations, providing a robust framework for analyzing complex datasets, particularly genetic data. The C-mix model leverages modern computational techniques to handle high dimensionality and censoring simultaneously, offering new insights and applications in genetic research.

The paper by Giné and Guillou (2001) makes a substantial contribution to the literature on non-parametric estimation methods for censored data. Giné and Guillou (2001) studied the consistency of kernel density estimators for randomly censored data and derived rates of convergence that hold uniformly over adaptive intervals. This work builds on the foundational results for kernel density estimation and extends them to the context of censored data, providing a rigorous theoretical framework for the consistency and convergence rates of these estimators. Giné and Guillou (2001)'s results are particularly important because they address the uniformity of convergence rates over adaptive intervals, which is crucial for practical applications where the density may vary significantly over different regions of the data. Their work ensures that the kernel density estimators are not only consistent but also provide reliable estimates across the entire range of the data, even in the presence of censoring. By establishing the consistency and convergence rates of kernel density estimators in the presence of random censoring, their work provides a robust theoretical foundation for further developments in this field. As statistical methods continue to evolve, these results will play a crucial role in enhancing the accuracy and reliability of density estimation techniques applied to censored data.

The work of Satten and Datta (2001) provides a novel interpretation of the Kaplan-Meier estimator, viewing it as an inverse-probability-of-censoring weighted (IPCW) average. This interpretation offers a deeper understanding of the estimator's properties and its connection to modern IPCW techniques. Specifically, they demonstrate that the Kaplan-Meier estimator can be seen as a weighted average of observed outcomes, where the weights are inversely related to the probability of being censored up to a given time point. This perspective links traditional non-parametric methods with contemporary IPCW approaches, highlighting the estimator's robustness in handling censored data. The insights from this work have significant implications for censored data analysis, as framing the Kaplan-Meier estimator within the IPCW framework provides a unified view that can guide both the development of new methodologies and the refinement of existing techniques. The integration of IPCW methods has proven valuable in generating unbiased and efficient estimates, and this approach continues to evolve alongside advancements in survival analysis, such as machine learning and Bayesian techniques. In this chapter, we will apply the IPCW-based Kaplan-Meier estimator to kernel density estimator, then incorporating Kaplan-Meier weighted kernel density estimator into the parametric and semiparametric MHDE. Both simulation studies and real data analyses involving right-censored data are studied in the next few sections.

Zeller et al. (2019) developed an efficient expectation-maximization (EM) algorithm for parameter estimation in the finite mixture of regression models based on scale mixtures of normal distributions model. The authors propose a finite mixture of regression models based on scale mixtures of normal distributions, allowing for the incorporation of covariates and the accommodation of heterogeneous subpopulations. Zeller et al. (2019) also explored the application of the proposed methodology to real-world data, demonstrating its effectiveness in capturing the underlying data structure and providing insightful inferences. The flexibility and robustness of the proposed model make it a valuable addition to the toolkit for analyzing censored data. In this chapter, we will use the same real dataset used by Zeller et al. (2019) as my data analysis example.

4.2 Motivation

In the literature, several authors have explored finite mixture modeling of censored data using various distributions or regression models. These studies predominantly employ parametric models and maximum likelihood estimation (MLE). However, non-parametric and semi-parametric models, as well as minimum Hellinger distance estimation (MHDE), have not been extensively utilized when data are incomplete (right-censored). This new project seeks to extend the finite location-scale mixture model and MHDE method to handle right-censored data, presenting a novel approach to estimation. These focal points aim to advance the current methodologies and provide a more comprehensive tool for analyzing finite location-scale mixture model with right-censored data.

We propose to use the KM weighted kernel density estimator based on the methods of Satten and Datta (2001), which is formed by an inverse-probability weighted average to construct the new kernel density estimation for the nonparametric estimation part when the data are right-censored in the location-scale model; After we obtain the KM weights, we can form a kernel density estimation proposed by Giné and Guillou (2001). When data are complete or not censored, the original kernel density estimator is already presented in Chapter 2 and 3 when constructing MHDEP and SEMIMHDE. A nonparametric estimator $\hat{h}_n$ of a probability density function h can be constructed using a kernel density estimator (KDE), which is a common technique in nonparametric statistics. The KDE is defined as:

$$\hat{h}_n(t) = \frac{1}{h_n} \int K\left(\frac{t-x}{h_n}\right) d\hat{F}_n(x), \quad n \in \mathbb{N}, \tag{4.1}$$

where K is the kernel function, a probability density function that is symmetric around zero and integrates to one. Common choices for K include the Gaussian (normal), Epanechnikov, and uniform kernels. The choice of kernel affects the smoothness and bias of the estimator, but the overall performance is typically more influenced by the bandwidth parameter h_n;

h_n is the bandwidth, it is a positive constant that determines the width of the kernel and, consequently, the smoothness of the estimated density. As n increases, the bandwidth h_n typically decreases (tending to zero), allowing the estimator to adapt to the local structure of the data. The bandwidth is a critical tuning parameter that controls the trade-off between bias and variance in the estimation. A smaller bandwidth leads to a more sensitive estimator (lower bias, higher variance), while a larger bandwidth provides a smoother estimator (higher bias, lower variance); $\hat{F}_n$ is the empirical distribution function based on the observed data points. It provides a step function that jumps by $1/n$ at each observed data point x_i. The integration with respect to $\hat{F}_n(x)$ in the KDE sums the contributions of the kernel function centered at each data point, scaled by the bandwidth h_n.

When data are censored, we can modify the estimator (4.1) to incorporate censoring. The idea is to calculate the Kaplan-Meier weights using the method by Satten and Datta (2001), then use the estimator given by Giné and Guillou (2001) to construct a new kernel density estimator with censored data. To construct an estimator for the density with right censored data, we can just replace $(1/n)$ in above KDE function (4.1) by $(\delta/(n * K_{hat}(X_{j-}))$, here $K_{hat}(X_{j-})$ is the left-continous version of the estimator of the survival function of the censoring data, see Satten and Datta (2001), $(\delta/(K_{hat}(X_{j-}))$ is the Inverse-Probability-of-Censoring Weight, we simply call it KM weight.

This study presents a novel approach to handling right-censored data, aiming to achieve accurate and unbiased estimations. Departing from traditional parametric and non-parametric methods, we apply the Kaplan-Meier estimator as an inverse-probability-of-censoring weighted average to Hellinger Distance estimations. This method is specifically tailored for analyzing right-censored data within two-component mixture models, applicable to both parametric and semi-parametric cases. The approach focuses on two key aspects:

1. Integration of the KM-Weighted kernel density estimator. The paper Satten and Datta (2001) introduce the concept of using Kaplan-Meier (KM) estimator as an inverse-

probability-of-censoring weighted average for dealing with right-censored data. These KM weights are integral to our methodology, particularly in the kernel density estimation within a semiparametric location-scale model. By employing the KM weights, we can effectively handle the censoring present in the data. Upon obtaining the KM weights, we will apply them to form a kernel density estimation, as proposed by Satten and Datta (2001). This approach leverages the strengths of the KM estimator in addressing right-censored data, thereby enhancing the robustness and accuracy of the kernel density estimation.

2. The KM-Weighted kernel density estimator will be applied to both the Minimum Hellinger Distance Estimator for Parametric (MHDEP) and the Semi-Parametric Minimum Hellinger Distance Estimator (SEMIMHDE) methods. The steps involved are as follows:

(2.1) **Calculation of KM Weights**: Implement the Kaplan-Meier estimator to derive the inverse-probability-of-censoring weights for the right-censored data.

(2.2) **Kernel Density Estimation for Censored Data**: Apply the obtained KM weights to the kernel density estimation framework proposed by Giné and Guillou (2001).

(2.3) **Integration with MHDEP**: Incorporate the KM-weighted kernel density estimator into the MHDEP framework to handle parametric models of censored data.

(2.4) **Integration with SEMIMHDE**: Extend the use of the KM-weighted kernel density estimator to the SEMIMHDE framework, enhancing the estimation procedure for semiparametric models of censored data.

4.3 Model and Estimation

Our proposed methodologies, the MHDEP and SEMIMHDE are applied to handle right-censored data, assuming a two-component location-scale mixture model. Although these estimators were not originally developed specifically for right-censored data, they are adapted here to address this challenge. Both approaches incorporate a nonparametric estimation component, typically utilizing KDE, which constitutes the second part of the objective function in the Minimum Hellinger Distance measure. This section details the integration of the Kaplan-Meier (KM) estimator into the kernel density estimation process, enhancing the robustness and accuracy of our models. In survival analysis and related fields, KDE is a pivotal tool for estimating the underlying probability density function of the data.

The Kaplan-Meier (KM) estimator, also known as the product-limit estimator, is a non-parametric estimator used extensively in survival analysis to estimate the survival function from lifetime data. It is particularly adept at handling right-censored data, where the event of interest (for example, failure or death) has not been observed for all subjects within the study period. This estimator was introduced by Kaplan and Meier (1958) in their seminal 1958 paper.

The survival function $S(t)$ represents the probability that the event of interest has not occurred by time t. The KM estimator is defined as follows:

$$\hat{S}(t) = \prod_{t_i \leq t} \left(1 - \frac{d_i}{n_i}\right), \tag{4.2}$$

where t_i are the distinct times at which events occur, d_i is the number of events (e.g., deaths or failures) at time t_i, n_i is the number of individuals at risk just prior to time t_i. The KM estimator works by adjusting the survival probability at each observed event time. For each distinct time point t_i, the estimator considers the proportion of subjects experiencing the event out of those at risk immediately before t_i. The survival probability is then updated by multiplying the current survival probability by the complement of this proportion. This process continues across all observed event times, yielding a step function that decreases at each event time and remains constant between events.

The KM estimator is widely appreciated for its simplicity and robustness. It provides an intuitive visual representation of the survival function through the KM curve. Additionally, it serves as the basis for various extensions and adaptations. For instance, Satten and Datta (2001) proposed an inverse-probability-of-censoring weighted version of the KM estimator, enhancing its utility in complex statistical models, such as kernel density estimation for right-censored data. Traditional KDE methods, however, are not well-suited for censored data due to their inability to account for incomplete observations. To address this limitation, we introduce a modified version of the KDE that incorporates KM weights, as proposed by Satten and Datta (2001). This approach adjusts each observation by the inverse of its probability of being uncensored, effectively mitigating the biases introduced by censored data. In Satten and Datta (2001), the Kaplan-Meier (KM) weights are introduced as a means to adjust for right-censoring in statistical analyses. These weights are essential for transforming the Kaplan-Meier estimator into an inverse-probability-of-censoring weighted average. This modification ensures that estimates remain unbiased in the presence of censored data. The KM weights are derived from the survival probabilities estimated by the Kaplan-Meier estimator. They are defined for each observation and used to account for the probability of being uncensored up to a particular time point. The weight for an individual i at time T_i is calculated as:

$$W_i = \frac{\delta_i}{\hat{S}(T_i)}, \tag{4.3}$$

where δ_i is an indicator variable that equals 1 if the event is observed (uncensored) and 0 if the data is censored. $\hat{S}(T_i)$ is the Kaplan-Meier estimate of the survival function at the observed time T_i. This weight adjusts each observation by the inverse of its probability of being uncensored, as estimated by the KM estimator. For censored observations ($\delta_i = 0$), the weight W_i is set to zero, reflecting the fact that these observations do not contribute directly to the estimation of the survival probability at later times.

In our context, the KM weights are applied to KDE process to account for right-censored data. Specifically, the weights W_i are used to adjust the contributions of each observation in the KDE, ensuring that the density estimates remain unbiased in the presence of censored data. This methodology aligns with the approach proposed by Giné and Guillou (2001), which involves using KM weights to enhance the accuracy of KDE for censored data.

For right-censored data, we incorporate the Kaplan-Meier weight into the KDE estimation process to adjust for censoring. Let $x_1, x_2, \ldots, x_n$ be the observed data points, where some observations are right-censored. The KDE estimate at a point x by adding KM weights for censored data is given by:

$$\tilde{h}(x) = \frac{1}{nb_n} \sum_{i=1}^{n} K((x - x_i)/b_n) \cdot W_i, \tag{4.4}$$

where $K(\cdot)$ is the kernel function with bandwidth b_n. W_i is the Kaplan-Meier weight for observation x_i. When data is not censored, by the definition of KM weight W_i is 1, thus $\tilde{h}(x)$ is exactly the same as the original KDE as defined in (4.1).

The incorporation of KM weights into our statistical models allows for more robust and accurate handling of right-censored data. By leveraging the inverse-probability-of-censoring weighted average approach, as detailed by Satten and Datta (2001), we can improve the precision of our nonparametric density estimations within the MHDEP and SEMIMHDE frameworks. By incorporating the KM-weighted KDE into the MHDEP and SEMIMHDE frameworks, we aim to improve the accuracy and robustness of our estimations. The weighted KDE accounts for the censoring present in the data, ensuring that our objective distance measure, based on the Hellinger Distance, remains unbiased and reflective of the true underlying distribution. This integration allows our models to effectively handle the complexities of right-censored data, providing a more reliable tool for statistical analysis in survival analysis, reliability engineering, and beyond.

Next, we explore the application of the weighted KDE function (4.4) to the Hellinger Distance estimations such as MHDEP and SEMIMHDE. These two estimations involve the Hellinger Distance, which can be divided into two parts for computational purposes. The Hellinger Distance H between two probability density functions $f(x)$ and $g(x)$ is defined as:

$$H^2(f,g) = \frac{1}{2}\int \left(\sqrt{f(x)} - \sqrt{g(x)}\right)^2 dx = 1 - \int \sqrt{f(x)g(x)}dx \in [0,1]. \tag{4.5}$$

Let $X_1, \ldots, X_n$ be independent observations with sample size of n

$$X_1, \ldots, X_n \overset{\text{i.i.d}}{\sim} h_\theta(x), \tag{4.6}$$

where $h_\theta(x)$ is the finite parametric location-scale mixture model. Then the constructed MHDEP estimator is defined as

$$\hat{\theta}_{MHDEP} = \underset{\theta\in\Theta}{\arg\min} \left\| h_\theta^{\frac{1}{2}} - \tilde{h}^{\frac{1}{2}} \right\|. \tag{4.7}$$

An SEMIMHDE of θ is proposed as

$$\hat{\theta}_{SEMIMHDE} = \underset{\theta\in\Theta}{\arg\min} \left\| h_{\theta,\hat{f}}^{\frac{1}{2}} - \tilde{h}^{\frac{1}{2}} \right\|. \tag{4.8}$$

For SEMIMHDE (4.8) and MHDEP (4.7), the second part of the Hellinger Distance $\tilde{h}$ is particularly important, which is usually a nonparametric original kernel density estimator. By substituting the original KDE function (4.1) with the weighted KDE function (4.4), we adjust for the presence of censored data in the estimation process. This substitution ensures that the density estimates used in the Hellinger Distance calculation accurately reflect the underlying data, accounting for any censoring that may have occurred.

In conclusion, the application of the weighted KDE function, as defined in equation (4.4), allows us to substitute the original KDE with the KM weights for both SEMIMHDE and MHDEP methods when data are censored. This substitution is crucial for accurately computing the second part of the Hellinger Distance, as defined in equations (4.8) and (4.7), in the presence of right-censored data. By incorporating the Kaplan-Meier weights, we can ensure that our density estimates and subsequent analyses are robust and reliable. Incorporating the KM-weighted KDE proposed in Satten and Datta (2001) into our estimations like MHDEP and SEMIMHDE helped us dealing with mixture when data is right-censored. This methodological enhancement ensures that our nonparametric estimations are both robust and accurate, maintaining the integrity of the objective distance measure in the presence of censored observations.

4.4 Simulation Experiments

In this section, we aim to test the performance of the proposed SEMIMHDE and MHDEP with weighted KM on location-scale mixture estimator with right censored data. Our simulation studies will mainly focus on 3 scenarios based on different parameter settings. The 3 scenarios are exactly the same as that in Cutler and Cordero-Brana (1996) in their evaluation of the HMIX algorithm based on the Hellinger distance. Since the 3 scenarios include small or larger overlap and larger second component probability, we can easily compare the finite sample efficiency performance from these 3 different type of scenarios. By replicating these scenarios, we ensure a consistent basis for comparison and validation of our methods. We will also compare our Hellinger distance-based estimations (SEMIMHDE and MHDEP) with likelihood-based estimation (MLE) across a range of right censoring rates, starting from no censoring (0%) and increasing to very high censoring rates, where the performance of the estimators become worse when censoring rates increase.

The three scenarios employed in our simulation studies are as follows:

Scenario 1: Small Overlap 0.03 - $0.25 \cdot N(0, \sqrt{2}) + 0.75 \cdot N(4.203, 1)$.

This scenario involves parameter settings that result in minimal overlap between the mixture components. It allows us to test the ability of the estimation methods to distinguish between components when the data is relatively well-separated.

Scenario 2: Larger Overlap 0.10 - $0.25 * N(0, \sqrt{2}) + 0.75 * N(2.573, 1)$.

Here, the parameter settings create a significant overlap between the mixture components. This scenario tests the robustness of the methods in more challenging conditions where the components are not easily distinguishable.

Scenario 3: Larger First Component Weight - $0.75 * N(0, \sqrt{2}) + 0.25 * N(2.964, 1)$.

In this scenario, the second component of the mixture model has a smaller probability. This tests the methods' performance when the mixture components have unequal contributions to the overall distribution. Here we can test our base function f in (3.6) for the case that $\pi_1 > \pi_2$.

These scenarios encompass a range of conditions that are likely to be encountered in practical applications, thereby providing a thorough evaluation of the proposed methods.

Data will be generated for each scenario using specified mixture model parameters with sample sizes of 100, 400, and 800. Right-censoring rates will range from 0% and increase until estimates show extreme bias or MSE. SEMIMHDE and MHDEP performance will be compared to MLE. Efficiency will be assessed by comparing bias, variance, and MSE across different scenarios and censoring levels.

The results from our simulation studies will provide insights into the relative performance of SEMIMHDE, MHDEP, and MLE under different conditions. By comparing the Hellinger distance-based estimations with likelihood-based estimations, we aim to demonstrate the advantages and potential limitations of our proposed methods.

4.4.1 Scenario 1: Efficiency Study Small Overlap Case on Different Censoring Rates

In this section, we focus on the efficiency study of Scenario 1: $0.25 \cdot N(0, \sqrt{2}) + 0.75 \cdot N(4.203, 1)$, with an overlap of 0.03, where N denotes each component is assumed to follow a normal distribution. This is the same parameter settings as Case 4 in the paper Cutler and Cordero-Brana (1996). The sample sizes considered are 100, 400, and 800, with 500 replications for each sample size. The censoring rates range from 0% to 80%. The parameter μ in each table on the first column is the value to determine the right censoring rate.

Simulation Results

The results of the simulation studies based on Scenario 1 are presented in Tables 4.1 through 4.9. Specifically: Tables 4.1 to 4.3 present the simulation results for SEMIMHDE based on sample sizes of 100, 400, and 800, respectively. Tables 4.4 to 4.6 present the simulation results for parametric MHDE (MHDEP) based on sample sizes of 100, 400, and 800, respectively. Tables 4.7 to 4.9 present the simulation results for MLE based on sample sizes of 100, 400, and 800, respectively.

Each table contains the estimations of 5 parameters with regarding bias, variance and MSE. Each μ value after censoring rate in the table is the value to determine the regarding censoring rate based on the true parameters values. The μ value will be used in the simulation data generation process. Bias measures the difference between the expected value of an estimator and the true value of the parameter being estimated. Mathematically, the bias of an estimator $\hat{\theta}$ for a parameter θ is defined as:

$$\text{Bias}(\hat{\theta}) = \mathbb{E}[\hat{\theta}] - \theta, \tag{4.9}$$

where $\mathbb{E}[\hat{\theta}]$ denotes the expected value of the estimator $\hat{\theta}$.

Variance measures the dispersion of the estimator around its expected value. It quantifies the variability of the estimator due to the randomness of the sample. The variance of an estimator $\hat{\theta}$ is given by:

$$\mathbb{V}ar(\hat{\theta}) = \mathbb{E}[(\hat{\theta} - \mathbb{E}[\hat{\theta}])^2]. \tag{4.10}$$

Mean Squared Error (MSE) combines both bias and variance to provide a comprehensive measure of an estimator's accuracy. It is defined as the expected value of the squared difference between the estimator and the true parameter value:

$$\text{MSE}(\hat{\theta}) = \mathbb{E}[(\hat{\theta} - \theta)^2]. \tag{4.11}$$

The MSE can also be decomposed into the sum of the variance and the square of the bias:

$$\text{MSE}(\hat{\theta}) = \mathbb{V}ar(\hat{\theta}) + \text{Bias}(\hat{\theta})^2. \tag{4.12}$$

Performance Analysis of SEMIMHDE

From Tables 4.1 to 4.3, we observe the following trends for SEMIMHDE in Scenario 1: The parameters π_1, σ_1, and σ_2 remain stable across censoring rates from 0% to 40%. The bias for the parameters μ_1 and μ_2 increases as the censoring rate increases. This increase in bias is particularly noticeable for μ_2, where the bias becomes extremely large when the censoring rate reaches 50%. The distance between the estimates of the two location parameters decreases as the censoring rate increases.

Performance Analysis of MHDEP

From Tables 4.4 to 4.6, the following trends are observed for parametric MHDE (MHDEP) in Scenario 1: The parameter σ_1 remains stable when the censoring rate is below 50%. The parameters π_1, μ_1, μ_2, and σ_2 exhibit larger biases as the censoring rate increases. Notably, when the censoring rate reaches 60%, the estimate for π_1 becomes greater than π_2, and the bias for μ_2 approaches 1. Similar to SEMIMHDE, the distance between the estimates of the two location parameters decreases as the censoring rate increases.

Performance Analysis of MLE

From Tables 4.7 to 4.9, the following trends are observed for MLE in Scenario 1: The parameter σ_1 remains stable when the censoring rate is below 50%. The parameters π_1, μ_1, μ_2, and σ_2 exhibit larger biases as the censoring rate increases. When the censoring rate reaches 60%, the estimate for π_1 becomes greater than π_2, and the bias for μ_2 approaches 1. The distance between the estimates of the two location parameters decreases as the censoring rate increases.

Comparative Performance Analysis

By comparing the results of SEMIMHDE, MHDEP, and MLE, we draw the following conclusions: Both MHDEP and MLE exhibit similar performance, with their estimation accuracy becoming unreasonable when the censoring rate exceeds 60%. SEMIMHDE shows competitive performance compared to parametric estimations (MHDEP and MLE) but begins to show increased bias starting from a censoring rate of 50%. Overall, SEMIMHDE demonstrates robustness and stability for right-censored data at lower censoring rates, providing a viable alternative to parametric methods.

In this conclusion, we have conducted a comprehensive simulation study to evaluate the performance of SEMIMHDE and MHDEP methods for small overlap location-scale mixture models with right-censored data. Our results indicate that SEMIMHDE performs competitively with MHDEP and MLE, particularly at lower censoring rates before achieving 50%. The findings suggest that SEMIMHDE is a robust method for handling right-censored data, maintaining stability and accuracy across various scenarios, even SEMIMHDE includes less information about each component distribution, it still achieves similar results as parametric estimations.

Table 4.1: Scenario 1 SEMIMHDE Efficiency Studies at Different Censoring Rate, Overlap=0.03, sample size =100, reps=100

Censor Rate/μ		π_1	μ_1	σ_1	μ_2	σ_2
0%/999	Bias	$-.035$	.114	$-.424$	$-.122$	$-.103$
	Variance	.004	.464	.088	.042	.040
	MSE	.005	.477	.267	.057	.050
20%/5.57	Bias	$-.018$	.204	$-.451$	$-.456$	$-.126$
	Variance	.005	.317	.090	.049	.040
	MSE	.005	.359	.293	.257	.056
40%/4.05	Bias	.027	.350	$-.475$	$-.979$	$-.092$
	Variance	.007	.583	.093	.137	.065
	MSE	.007	.706	.318	1.100	.074
60%/2.6	Bias	.100	.231	$-.553$	-1.680	$-.138$
	Variance	.018	.369	.106	.847	.104
	MSE	.028	.572	.411	3.680	.123
80%/0.7	Bias	.113	$-.584$	$-.580$	-2.770	$-.184$
	Variance	.030	.507	.104	.586	.093
	MSE	.043	.848	.441	8.270	.126

Table 4.2: Scenario 1 SEMIMHDE Efficiency Studies at Different Censoring Rate, Overlap=0.03, sample size =400, reps=100

Censor Rate/μ		π_1	μ_1	σ_1	μ_2	σ_2
0%/999	Bias	−.036	−.108	−.271	−.097	.004
	Variance	.001	.113	.056	.012	.014
	MSE	.003	.125	.130	.022	.014
20%/5.6	Bias	−.013	.063	−.237	−.395	−.034
	Variance	.002	.138	.045	.016	.013
	MSE	.002	.142	.101	.172	.014
40%/4.1	Bias	.029	.163	−.307	−.838	.030
	Variance	.003	.148	.043	.023	.020
	MSE	.003	.175	.138	.726	.021
60%/2.65	Bias	.160	.445	−.338	−1.510	.020
	Variance	.008	.737	.063	.153	.054
	MSE	.034	.936	.177	2.430	.054
80%/0.75	Bias	.216	−.438	−.416	−2.550	.028
	Variance	.025	.213	.092	1.480	.252
	MSE	.072	.406	.265	8.000	.252

Table 4.3: Scenario 1 SEMIMHDE Efficiency Studies at Different Censoring Rate, Overlap=0.03, sample size =800, reps=100

Censor Rate/μ		π_1	μ_1	σ_1	μ_2	σ_2
0%/999	Bias	−.034	−.111	−.233	−.089	.003
	Variance	.009	.055	.034	.006	.005
	MSE	.002	.065	.083	.014	.005
20%/5.6	Bias	−.028	−.047	−.253	−.408	−.019
	Variance	.001	.074	.027	.009	.006
	MSE	.002	.076	.091	.176	.007
40%/4.1	Bias	.039	.205	−.214	−.794	.024
	Variance	.001	.097	.026	.014	.010
	MSE	.003	.139	.072	.643	.010
60%/2.65	Bias	.162	.300	−.279	−1.430	.032
	Variance	.005	.176	.035	.074	.034
	MSE	.031	.266	.113	2.120	.037
80%/0.75	Bias	.227	−.450	−.336	−2.510	.008
	Variance	.017	.158	.086	.249	.070
	MSE	.068	.361	.198	6.530	.071

Table 4.4: Scenario 1 MHDEP Efficiency Studies at Different Censoring Rate, Overlap=0.03, sample size =100, reps=100

Censor Rate/μ		π_1	μ_1	σ_1	μ_2	σ_2
0%/999	Bias	−.003	−.006	−.072	−.010	.051
	Variance	.003	.247	.081	.022	.014
	MSE	.003	.247	.086	.022	.017
20%/5.57	Bias	.040	.241	−.026	−.271	.019
	Variance	.009	.385	.117	.040	.021
	MSE	.011	.443	.117	.113	.021
40%/4.05	Bias	.119	.330	−.112	−.633	.029
	Variance	.026	.817	.153	.165	.055
	MSE	.044	.926	.165	.565	.056
60%/2.6	Bias	.224	−.004	−.245	−1.094	.006
	Variance	.058	1.198	.184	.585	.126
	MSE	.108	1.198	.244	1.782	.126
80%/0.7	Bias	.400	−.733	.021	−1.863	.102
	Variance	.087	1.176	.133	1.212	.159
	MSE	.247	1.713	.134	4.681	.169

Table 4.5: Scenario 1 MHDEP Efficiency Studies at Different Censoring Rate, Overlap=0.03, sample size =400, reps=100

Censor Rate/μ		π_1	μ_1	σ_1	μ_2	σ_2
0%/999	Bias	−.002	.011	−.016	−.006	.042
	Variance	.001	.065	.024	.006	.003
	MSE	.001	.065	.024	.006	.005
20%/5.6	Bias	.034	.210	.051	−.278	.010
	Variance	.002	.116	.036	.010	.005
	MSE	.003	.161	.039	.088	.005
40%/4.1	Bias	.134	.450	.068	−.611	.037
	Variance	.010	.288	.058	.045	.016
	MSE	.028	.491	.063	.418	.017
60%/2.65	Bias	.268	.198	−.056	−1.050	.031
	Variance	.047	.712	.103	.381	.080
	MSE	.118	.751	.107	1.483	.081
80%/0.75	Bias	.335	−.832	−.254	−2.146	−.029
	Variance	.073	1.131	.211	1.403	.227
	MSE	.185	1.823	.275	6.008	.228

Table 4.6: Scenario 1 MHDEP Efficiency Studies at Different Censoring Rate, Overlap=0.03, sample size =800, reps=100

Censor Rate/μ		π_1	μ_1	σ_1	μ_2	σ_2
0%/999	Bias	−.002	.002	−.005	−.005	.032
	Variance	.000	.028	.013	.003	.001
	MSE	.000	.028	.013	.003	.002
20%/5.6	Bias	.033	.217	.065	−.278	.005
	Variance	.001	.074	.025	.005	.003
	MSE	.002	.121	.030	.083	.003
40%/4.1	Bias	.149	.565	.131	−.586	.020
	Variance	.006	.176	.033	.023	.008
	MSE	.028	.494	.051	.367	.008
60%/2.65	Bias	.289	.337	.018	−1.023	.036
	Variance	.032	.454	.070	.237	.050
	MSE	.115	.568	.070	1.283	.052
80%/0.75	Bias	.364	−.840	−.158	−2.017	−.037
	Variance	.080	1.220	.195	1.388	.197
	MSE	.213	1.925	.220	5.455	.198

Table 4.7: Scenario 1 MLE Efficiency Studies at Different Censoring Rate, Overlap=0.10, sample size =100, reps=100

Censor Rate/μ		π_1	μ_1	σ_1	μ_2	σ_2
0%/999	Bias	.009	.042	−.015	−.003	−.010
	Variance	.004	.300	.111	.023	.013
	MSE	.004	.302	.111	.023	.013
20%/5.57	Bias	.040	.194	.029	−.288	−.030
	Variance	.006	.371	.121	.034	.016
	MSE	.008	.409	.122	.117	.017
40%/4.05	Bias	.144	.441	.057	−.061	−.021
	Variance	.018	.518	.116	.085	.035
	MSE	.038	.712	.119	.457	.036
60%/2.6	Bias	.284	.299	.006	−1.050	.016
	Variance	.034	.499	.096	.247	.065
	MSE	.115	.589	.096	1.340	.065
80%/0.7	Bias	.357	−.825	.024	−2.170	.181
	Variance	.071	1.370	.108	1.080	.156
	MSE	.198	2.050	.109	5.77	.189

Table 4.8: Scenario 1 MLE Efficiency Studies at Different Censoring Rate, Overlap=0.10, sample size =400, reps=100

Censor Rate/μ		π_1	μ_1	σ_1	μ_2	σ_2
0%/999	Bias	.002	.003	$-.006$	.001	$-.005$
	Variance	.001	.062	.032	.006	.003
	MSE	.001	.062	.032	.006	.003
20%/5.6	Bias	.037	.212	.075	$-.263$	$-.034$
	Variance	.002	.094	.039	.008	.004
	MSE	.003	.139	.045	.077	.005
40%/4.1	Bias	.144	.533	.146	$-.574$	$-.016$
	Variance	.006	.167	.043	.022	.008
	MSE	.027	.451	.065	.351	.009
60%/2.65	Bias	.340	.581	.147	$-.930$	$-.013$
	Variance	.015	.165	.034	.084	.026
	MSE	.131	.502	.055	.948	.026
80%/0.75	Bias	.501	$-.321$	.208	-1.730	.055
	Variance	.052	.872	.054	.746	.127
	MSE	.303	.974	.097	3.740	.130

Table 4.9: Scenario 1 MLE Efficiency Studies at Different Censoring Rate, Overlap=0.10, sample size =800, reps=100

Censor Rate/μ		π_1	μ_1	σ_1	μ_2	σ_2
0%/999	Bias	.002	.016	.008	.000	$-.002$
	Variance	.000	.032	.015	.002	.002
	MSE	.000	.032	.015	.002	.002
20%/5.6	Bias	.040	.246	.098	$-.264$	$-.034$
	Variance	.001	.065	.021	.004	.002
	MSE	.003	.125	.031	.073	.003
40%/4.1	Bias	.153	.600	.181	$-.559$	$-.021$
	Variance	.004	.101	.023	.010	.005
	MSE	.027	.461	.056	.323	.005
60%/2.65	Bias	.358	.650	.180	$-.904$	$-.017$
	Variance	.008	.089	.016	.036	.012
	MSE	.136	.511	.049	.854	.012
80%/0.75	Bias	.586	$-.046$	.286	-1.490	$-.041$
	Variance	.020	.058	.014	.415	.091
	MSE	.364	.060	.096	2.640	.093

4.4.2 Scenario 2: Efficiency Study Large Overlap Case on Different Censoring Rates

In this section, we focus on Scenario 2: $0.25 \cdot N(0, \sqrt{2}) + 0.75 \cdot N(2.573, 1)$. This case has an overlap of 0.10. We examine the performance of the estimators for different sample sizes (100, 400, 800) and a range of censoring rates from 0% to 80%. Each simulation is repeated 500 times.

The results are presented in Tables 4.10 to 4.12: Simulation results for SEMIMHDE with sample sizes 100, 400, and 800, respectively, Tables 4.13 to 4.15: Simulation results for parametric MHDE (MHDEP) with sample sizes 100, 400, and 800, respectively, Tables 4.16 to 4.18: Simulation results for Maximum Likelihood Estimation (MLE) with sample sizes 100, 400, and 800, respectively.

Performance Analysis of SEMIMHDE

Tables 4.10 to 4.12 display the simulation results for SEMIMHDE for Scenario 2: Stability of Parameters: The estimates for the standard deviations σ_1 and σ_2 remain stable across censoring rates from 0% to 50%. Bias in Estimates: The bias in the estimates of π_1, μ_1, and μ_2 increases with higher censoring rates. Specifically, the bias in μ_1 becomes extremely large when the censoring rate reaches 60%. Location Parameters:The estimates for the location parameters μ_1 and μ_2 tend to decrease as the censoring rate increases.

Performance Analysis of MHDEP

Tables 4.13 to 4.15 present the results for MHDEP for Scenario 2: Stability of Parameters: The parameter σ_1 remains stable for censoring rates below 70%. Bias in Estimates: The parameters π_1, μ_1, μ_2, and σ_2 exhibit increased bias with higher censoring rates. Notably, when the censoring rate reaches 70%, π_1 is overestimated relative to π_2, and the bias in μ_2

approaches 1. Trend in Bias: The bias in μ_2 consistently increases as the censoring rate rises.

Performance Analysis of MLE

Tables 4.16 to 4.18 detail the MLE results for Scenario 2: Stability of Parameters: The parameter σ_1 is stable for censoring rates below 60%. Bias in Estimates: The parameters π_1, μ_1, μ_2, and σ_2 show significant bias at higher censoring rates. Similar to MHDEP, when the censoring rate reaches 60%, π_1 is overestimated relative to π_2, and the bias in μ_2 nears 1. Distance in Location Parameters: The distance between the estimated location parameters decreases as the censoring rate increases.

Comparative Analysis

Comparing the performance of SEMIMHDE, MHDEP, and MLE: Performance under High Censoring Rates: Both MHDEP and MLE show unreasonable performance starting from a censoring rate of 70%, while SEMIMHDE's performance degrades starting from 60%. Robustness of SEMIMHDE: The proposed SEMIMHDE method demonstrates competitive performance compared to the parametric estimations (MHDEP and MLE) for right-censored data, particularly as the censoring rate increases.

In conclusion, scenario 2 characterized by a higher overlap of 0.10, SEMIMHDE, MHDEP, and MLE all show varying degrees of robustness and bias across different censoring rates. The results indicate that while SEMIMHDE starts showing increased bias at lower censoring rates (60%), it still performs competitively in comparison to MHDEP and MLE. This suggests that SEMIMHDE is a viable method for handling right-censored data, especially in scenarios with higher censoring rates.

Table 4.10: Scenario 2 SEMIMHDE Efficiency Studies at Different Censoring Rate, Overlap=0.10, sample size =100, reps=100

Censor Rate/μ		π_1	μ_1	σ_1	μ_2	σ_2
0%/999	Bias	.012	.166	$-.570$	$-.036$	.188
	Variance	.023	.724	.110	.057	.068
	MSE	.023	.764	.431	.059	.101
20%/4	Bias	.034	.065	$-.598$	$-.343$	$-.224$
	Variance	.013	.480	.071	.070	.050
	MSE	.014	.485	.429	.188	.10
40%/2.6	Bias	.072	$-.017$	$-.624$	$-.734$	$-.259$
	Variance	.012	.366	.073	.114	.059
	MSE	.017	.366	.463	.653	.126
60%/1.36	Bias	.126	$-.350$	$-.719$	-1.210	$-.300$
	Variance	.018	.372	.077	.437	.096
	MSE	.034	.495	.594	1.900	.185
70%/0.7	Bias	.117	$-.803$	$-.671$	-1.570	$-.262$
	Variance	.022	.401	.089	.310	.100
	MSE	.035	1.040	.539	2.770	.169

Table 4.11: Scenario 2 SEMIMHDE Efficiency Studies at Different Censoring Rate, Overlap=0.10, sample size =400, reps=100

Censor Rate/μ		π_1	μ_1	σ_1	μ_2	σ_2
0%/999	Bias	.007	.090	$-.261$	$-.015$	$-.077$
	Variance	.008	.305	.040	.012	.024
	MSE	.008	.313	.109	.012	.030
20%/4.05	Bias	.028	.045	$-.305$	$-.287$	$-.127$
	Variance	.006	.184	.034	.019	.017
	MSE	.006	.186	.127	.101	.033
40%/2.6	Bias	.077	$-.233$	$-.374$	$-.612$	$-.076$
	Variance	.007	.166	.031	.030	.018
	MSE	.013	.221	.171	.404	.024
60%/1.36	Bias	.164	$-.526$	$-.409$	-1.130	$-.114$
	Variance	.008	.196	.039	.076	.037
	MSE	.035	.473	.206	1.350	.050
70%/0.7	Bias	.180	$-.848$	$-.423$	-1.470	$-.117$
	Variance	.011	.152	.056	.096	.055
	MSE	.043	.872	.235	2.240	.069

Table 4.12: Scenario 2 SEMIMHDE Efficiency Studies at Different Censoring Rate, Overlap=0.10, sample size =800, reps=100

Censor Rate/μ		π_1	μ_1	σ_1	μ_2	σ_2
0%/999	Bias	$-.014$	$-.026$	$-.204$	$-.033$	$-.052$
	Variance	.003	.172	.035	.009	.010
	MSE	.003	.172	.076	.010	.013
20%/4.07	Bias	.009	$-.056$	$-.278$	$-.282$	$-.090$
	Variance	.005	.171	.026	.011	.010
	MSE	.005	.174	.103	.091	.018
40%/2.65	Bias	.079	$-.209$	$-.325$	$-.568$	$-.078$
	Variance	.005	.108	.017	.011	.013
	MSE	.012	.151	.112	.333	.019
60%/1.36	Bias	.159	$-.620$	$-.336$	-1.040	$-.042$
	Variance	.006	.119	.033	.033	.022
	MSE	.031	.504	.146	1.120	.023
70%/0.7	Bias	.203	$-.913$	$-.328$	-1.390	$-.034$
	Variance	.008	.127	.024	.075	.029
	MSE	.049	.961	.131	2.020	.031

Table 4.13: Scenario 2 MHDEP Efficiency Studies at Different Censoring Rate, Overlap=0.10, sample size =100, reps=100

Censor Rate/μ		π_1	μ_1	σ_1	μ_2	σ_2
0%/999	Bias	.003	$-.275$	$-.250$	$-.008$	.025
	Variance	.025	.819	.124	.068	.036
	MSE	.025	.894	.187	.068	.036
20%/4	Bias	.062	$-.454$	$-.374$	$-.387$	$-.020$
	Variance	.045	1.081	.134	.157	.049
	MSE	.048	1.287	.274	.307	.050
40%/2.6	Bias	.115	$-.514$	$-.404$	$-.495$	$-.050$
	Variance	.058	1.113	.135	.290	.067
	MSE	.071	1.377	.298	.536	.069
60%/1.36	Bias	.202	$-.926$	$-.415$	$-.804$	$-.093$
	Variance	.072	1.313	.148	.518	.104
	MSE	.113	2.170	.320	1.163	.112
70%/0.7	Bias	.225	-1.289	$-.422$	-1.120	$-.109$
	Variance	.071	1.155	.173	.688	.132
	MSE	.122	2.816	.351	1.942	.144

Table 4.14: Scenario 2 MHDEP Efficiency Studies at Different Censoring Rate, Overlap=0.10, sample size =400, reps=100

Censor Rate/μ		π_1	μ_1	σ_1	μ_2	σ_2
0%/999	Bias	−.011	−.151	−.108	−.026	.041
	Variance	.007	.309	.055	.016	.006
	MSE	.007	.332	.066	.017	.008
20%/4.05	Bias	.046	−.064	−.121	−.246	−.013
	Variance	.014	.341	.050	.028	.010
	MSE	.016	.345	.065	.089	.010
40%/2.6	Bias	.133	−.245	−.172	−.538	−.009
	Variance	.030	.496	.064	.089	.025
	MSE	.047	.556	.093	.379	.025
60%/1.36	Bias	.231	−.801	−.213	−.878	.021
	Variance	.069	1.075	.113	.361	.081
	MSE	.122	1.717	.158	1.131	.082
70%/0.7	Bias	.274	−1.156	−.203	−1.118	.031
	Variance	.082	1.250	.149	.640	.120
	MSE	.157	2.586	.191	1.890	.121

Table 4.15: Scenario 2 MHDEP Efficiency Studies at Different Censoring Rate, Overlap=0.10, sample size =800, reps=100

Censor Rate/μ		π_1	μ_1	σ_1	μ_2	σ_2
0%/999	Bias	−.011	−.109	−.063	−.022	.034
	Variance	.004	.165	.032	.008	.003
	MSE	.004	.177	.036	.009	.005
20%/4.07	Bias	.045	−.001	−.069	−.247	−.012
	Variance	.009	.228	.034	.015	.005
	MSE	.011	.228	.039	.076	.005
40%/2.65	Bias	.157	−.071	−.090	−.494	−.012
	Variance	.021	.297	.032	.048	.014
	MSE	.045	.302	.040	.293	.014
60%/1.36	Bias	.265	−.576	−.094	−.843	−.007
	Variance	.046	.648	.059	.221	.051
	MSE	.116	.980	.068	.932	.051
70%/0.7	Bias	.298	−1.054	−.093	−1.093	−.010
	Variance	.071	1.178	.103	.501	.096
	MSE	.160	2.290	.112	1.696	.096

Table 4.16: Scenario 2 MLE Efficiency Studies at Different Censoring Rate, Overlap=0.10, sample size =100, reps=100

Censor Rate/μ		π_1	μ_1	σ_1	μ_2	σ_2
0%/999	Bias	.022	$-.112$	$-.137$	$-.006$	$-.037$
	Variance	.022	.727	.143	.054	.025
	MSE	.022	.740	.162	.054	.027
20%/4	Bias	.052	$-.174$	$-.192$	$-.264$	$-.068$
	Variance	.027	.706	.135	.073	.028
	MSE	.030	.736	.172	.143	.033
40%/2.6	Bias	.139	$-.287$	$-.192$	$-.531$	$-.069$
	Variance	.038	.687	.105	.128	.044
	MSE	.057	.769	.142	.410	.049
60%/1.36	Bias	.226	$-.755$	$-.140$	$-.915$	$-.019$
	Variance	.053	.979	.110	.283	.065
	MSE	.104	1.550	.130	1.120	.065
70%/0.7	Bias	.258	-1.180	$-.082$	-1.210	.028
	Variance	.066	1.310	.104	.478	.089
	MSE	.132	2.710	.111	1.930	.090

Table 4.17: Scenario 2 MLE Efficiency Studies at Different Censoring Rate, Overlap=0.10, sample size =400, reps=100

Censor Rate/μ		π_1	μ_1	σ_1	μ_2	σ_2
0%/999	Bias	.010	$-.047$	$-.055$	$-.012$	$-.013$
	Variance	.010	.340	.074	.016	.009
	MSE	.010	.341	.074	.016	.009
20%/4.05	Bias	.060	.032	$-.059$	$-.240$	$-.055$
	Variance	.012	.280	.054	.019	.009
	MSE	.015	.281	.057	.077	.012
40%/2.6	Bias	.185	.015	$-.042$	$-.497$	$-.063$
	Variance	.022	.253	.040	.043	.018
	MSE	.056	.254	.042	.289	.022
60%/1.36	Bias	.333	$-.285$	.052	$-.809$	$-.049$
	Variance	.031	.299	.033	.122	.038
	MSE	.142	.380	.035	.778	.041
70%/0.7	Bias	.361	$-.706$	.114	-1.120	.007
	Variance	.048	.638	.040	.258	.067
	MSE	.179	1.140	.053	1.520	.067

Table 4.18: Scenario 2 MLE Efficiency Studies at Different Censoring Rate, Overlap=0.10, sample size =800, reps=100

Censor Rate/μ		π_1	μ_1	σ_1	μ_2	σ_2
0%/999	Bias	.008	$-.008$	$-.018$	$-.010$	$-.005$
	Variance	.006	.222	.042	.008	.004
	MSE	.006	.222	.042	.008	.004
20%/4.07	Bias	.075	.151	$-.001$	$-.228$	$-.058$
	Variance	.008	.191	.030	.010	.005
	MSE	.014	.214	.030	.062	.008
40%/2.65	Bias	.193	.097	$-.008$	$-.475$	$-.060$
	Variance	.014	.157	.021	.021	.010
	MSE	.052	.166	.021	.246	.013
60%/1.36	Bias	.373	$-.147$	.092	$-.775$	$-.060$
	Variance	.015	.084	.011	.054	.019
	MSE	.154	.105	.019	.656	.023
70%/0.7	Bias	.430	$-.475$	.166	-1.040	$-.031$
	Variance	.024	.123	.010	.140	.041
	MSE	.209	.348	.037	1.210	.042

4.4.3 Scenario 3: Efficiency Study Larger First Component Case on Different Censoring Rates

In this section, we focus on Scenario 3, defined by the mixture model $0.75 \cdot N(0, \sqrt{2}) + 0.25 \cdot N(2.964, 1)$. This case has an overlap of 0.10. Comparing with previous two scenarios, since the second mixing proportion is 0.25, thus it is much more difficult to estimation when data are right-censored. We examine the performance of the estimators for different sample sizes (100, 400, 800) and a range of censoring rates from 0% to 60%. Each simulation is repeated 500 times.

The results are presented in the following tables: Tables 4.19 to 4.21: Simulation results for SEMIMHDE with sample sizes 100, 400, and 800, respectively. Tables 4.22 to 4.24: Simulation results for parametric MHDE (MHDEP) with sample sizes 100, 400, and 800, respectively. Tables 4.25 to 4.27: Simulation results for Maximum Likelihood Estimation (MLE) with sample sizes 100, 400, and 800, respectively.

Performance Analysis of SEMIMHDE

Tables 4.19 to 4.21 display the simulation results for SEMIMHDE for Scenario 3: Stability of Parameters: The estimates for π_1, σ_1, and σ_2 remain stable across all censoring rates. Bias in Location Parameters: The estimates for the location parameters μ_1 and μ_2 deteriorate as the censoring rate increases. The bias for μ_2 becomes particularly large when the censoring rate exceeds 40%. Effect of Censoring: The bias for μ_2 is extremely large when the censoring rate is over 20%. The estimates for μ_1 and μ_2 tend to shift leftwards as the censoring rate increases. This degradation is more pronounced in SEMIMHDE for μ_2, especially when the second component has a smaller proportion.

Performance Analysis of MHDEP

Tables 4.22 to 4.24 present the results for MHDEP for Scenario 3: Stability of Parameters: The estimates for π_1, σ_1, and σ_2 remain stable across all censoring rates. Bias in Location Parameters: The estimates for μ_1 and μ_2 show a similar trend to SEMIMHDE, with increasing bias as the censoring rate increases. The bias for μ_2 is extremely large when the censoring rate is over 40%. Efficiency Comparison: MHDEP shows better efficiency performance compared to SEMIMHDE when sample size is larger and censoring rate <= 40%. When censoring rate is 60%, SEMIMHDE is better.

Performance Analysis of MLE

Tables 4.25 to 4.27 detail the MLE results for Scenario 3: Stability of Parameters: The estimates for π_1, σ_1, and σ_2 remain stable across all censoring rates. Bias in Location Parameters: Similar to MHDEP, the estimates for μ_1 and μ_2 deteriorate as the censoring rate increases. When the censoring rate approaches 60%, both MHDEP and MLE show inconsistency, with increasing bias for μ_2 as the sample size increases, making the estimates unreasonable at higher censoring rates. Comparison with MHDEP: MLE performs similarly to MHDEP, showing better efficiency than SEMIMHDE for right-censored data.

Comparative Analysis

Comparing the performance of SEMIMHDE, MHDEP, and MLE: Performance under High Censoring Rates: Both MHDEP and MLE become inconsistent and show unreasonable performance starting from a censoring rate of 60%, while SEMIMHDE shows increased bias starting from 40%. Robustness of MHDEP and MLE: MHDEP and MLE demonstrate better efficiency than SEMIMHDE, especially when the second component has a smaller proportion (25%). Challenges with SEMIMHDE: SEMIMHDE has worse performance in estimating μ_2 when the data is right-censored and the second component has a smaller proportion, indicating a limitation of the method under these conditions.

In conclusion, scenario 3 characterized by a higher proportion (75%) for the first component and a smaller proportion (25%) for the second component, SEMIMHDE, MHDEP, and MLE show varying degrees of robustness and bias across different censoring rates. The results indicate that while SEMIMHDE starts showing increased bias at lower censoring rates (40%), it performs worse in estimating μ_2 compared to MHDEP and MLE. MHDEP and MLE, though more robust, become inconsistent at a censoring rate of 60%. Overall, MHDEP and MLE demonstrate slightly better efficiency performance than SEMIMHDE in this scenario.

Table 4.19: Scenario 3 SEMIMHDE Efficiency Studies at Different Censoring Rate, Overlap=0.10, sample size =100, reps=500

Censor Rate/μ		π_1	μ_1	σ_1	μ_2	σ_2
0%/999	Bias	−.110	−.194	−.331	−.587	−.163
	Variance	.014	.117	.161	.464	.078
	MSE	.026	.155	.271	.809	.104
20%/3.05	Bias	−.113	−.342	−.461	−1.180	−.232
	Variance	.020	.120	.121	.431	.077
	MSE	.032	.237	.334	1.810	.131
40%/1.4	Bias	−.107	−.774	−.437	−1.680	−.249
	Variance	.022	.145	.106	.375	.062
	MSE	.034	.744	.297	3.210	.124
60%/0.01	Bias	−.188	−1.710	−.452	−2.610	−.249
	Variance	.041	.766	.158	.330	.065
	MSE	.077	3.680	.362	7.120	.127

Table 4.20: Scenario 3 SEMIMHDE Efficiency Studies at Different Censoring Rate, Overlap=0.10, sample size =400, reps=500

Censor Rate/μ		π_1	μ_1	σ_1	μ_2	σ_2
0%/999	Bias	−.108	−.263	−.194	−.448	.039
	Variance	.007	.038	.039	.157	.027
	MSE	.018	.108	.077	.357	.028
20%/3.01	Bias	−.127	−.448	−.289	−1.070	−.008
	Variance	.012	.039	.032	.172	.021
	MSE	.028	.240	.116	1.310	.021
40%/1.40	Bias	−.104	−.832	−.319	−1.580	−.043
	Variance	.013	.053	.023	.132	.014
	MSE	.023	.744	.124	2.640	.016
60%/0.01	Bias	−.160	−1.600	−.267	−2.420	−.046
	Variance	.021	.117	.031	.167	.018
	MSE	.046	2.690	.102	6.040	.020

Table 4.21: Scenario 3 SEMIMHDE Efficiency Studies at Different Censoring Rate, Overlap=0.10, sample size =800, reps=500

Censor Rate/μ		π_1	μ_1	σ_1	μ_2	σ_2
0%/999	Bias	−.115	−.294	−.173	−.392	.073
	Variance	.004	.022	.014	.061	.014
	MSE	.017	.109	.044	.215	.019
20%/3.06	Bias	−.110	−.433	−.243	−.936	.032
	Variance	.005	.027	.013	.062	.014
	MSE	.018	.214	.072	.938	.015
40%/1.41	Bias	−.098	−.803	−.275	−1.540	.015
	Variance	.008	.033	.015	.073	.012
	MSE	.017	.678	.091	2.460	.012
60%/0.01	Bias	−.156	−1.570	−.237	−2.420	.019
	Variance	.015	.078	.022	.123	.012
	MSE	.039	2.540	.078	5.990	.012

Table 4.22: Scenario 3 MHDEP Efficiency Studies at Different Censoring Rate, Overlap=0.10, sample size =100, reps=500

Censor Rate/μ		π_1	μ_1	σ_1	μ_2	σ_2
0%/999	Bias	−.074	−.194	−.073	−.129	.038
	Variance	.036	.296	.069	.510	.095
	MSE	.042	.334	.074	.527	.096
20%/3.05	Bias	−.080	−.421	−.166	−.658	−.024
	Variance	.066	.595	.086	.756	.103
	MSE	.072	.772	.113	1.190	.103
40%/1.4	Bias	−.079	−.895	−.199	−1.152	−.056
	Variance	.104	1.102	.103	1.345	.152
	MSE	.110	1.902	.143	2.673	.155
60%/0.01	Bias	−.180	−1.939	−.194	−2.182	.022
	Variance	.139	1.973	.168	1.818	.183
	MSE	.172	5.734	.206	6.578	.184

Table 4.23: Scenario 3 MHDEP Efficiency Studies at Different Censoring Rate, Overlap=0.10, sample size =400, reps=500

Censor Rate/μ		π_1	μ_1	σ_1	μ_2	σ_2
0%/999	Bias	−.031	−.080	.003	−.076	.061
	Variance	.013	.084	.023	.175	.033
	MSE	.014	.091	.023	.181	.036
20%/3.06	Bias	.009	−.161	−.044	−.506	−.021
	Variance	.031	.159	.032	.377	.053
	MSE	.031	.185	.033	.634	.054
40%/1.42	Bias	.011	−.637	−.078	−.967	−.052
	Variance	.083	.780	.068	1.084	.116
	MSE	.083	1.187	.075	2.019	.119
60%/0.01	Bias	−.240	−2.052	−.097	−2.564	.181
	Variance	.163	2.663	.207	1.891	.174
	MSE	.220	6.872	.216	8.464	.207

Table 4.24: Scenario 3 MHDEP Efficiency Studies at Different Censoring Rate, Overlap=0.10, sample size =800, reps=500

Censor Rate/μ		π_1	μ_1	σ_1	μ_2	σ_2
0%/999	Bias	$-.014$	$-.029$	.018	$-.033$	.045
	Variance	.005	.036	.010	.080	.016
	MSE	.005	.037	.010	.081	.018
20%/3.06	Bias	.045	$-.070$	$-.005$	$-.459$	$-.039$
	Variance	.015	.097	.013	.202	.035
	MSE	.017	.102	.013	.413	.037
40%/1.41	Bias	.100	$-.403$	$-.003$	$-.714$	$-.113$
	Variance	.047	.419	.037	.694	.085
	MSE	.057	.582	.037	1.204	.098
60%/0.01	Bias	$-.243$	-2.033	$-.018$	-2.799	.214
	Variance	.165	2.854	.181	1.514	.164
	MSE	.225	6.987	.182	9.350	.210

Table 4.25: Scenario 3 MLE Efficiency Studies at Different Censoring Rate, Overlap=0.10, sample size =100, reps=500

Censor Rate/μ		π_1	μ_1	σ_1	μ_2	σ_2
0%/999	Bias	$-.075$	$-.202$	$-.097$	$-.205$	.024
	Variance	.034	.328	.068	.492	.097
	MSE	.039	.369	.077	.534	.097
20%/3.05	Bias	$-.119$	$-.459$	$-.190$	$-.922$	.042
	Variance	.060	.598	.081	.751	.118
	MSE	.074	.809	.117	1.600	.120
40%/1.4	Bias	$-.128$	$-.846$	$-.203$	-1.620	.114
	Variance	.074	.795	.087	1.030	.147
	MSE	.090	1.510	.128	3.660	.160
60%/0.01	Bias	$-.233$	-1.840	$-.090$	-2.660	.185
	Variance	.094	1.820	.173	1.120	.156
	MSE	.148	5.210	.181	8.210	.190

Table 4.26: Scenario 3 MLE Efficiency Studies at Different Censoring Rate, Overlap=0.10, sample size =400, reps=500

Censor Rate/μ		π_1	μ_1	σ_1	μ_2	σ_2
0%/999	Bias	−.024	−.055	−.028	−.061	−.001
	Variance	.010	.072	.019	.159	.038
	MSE	.011	.075	.020	.163	.038
20%/3.06	Bias	.010	−.133	−.062	−.538	−.069
	Variance	.030	.135	.025	.396	.074
	MSE	.030	.153	.029	.685	.079
40%/1.42	Bias	−.043	−.569	−.106	−1.350	.054
	Variance	.079	.785	.062	1.140	.161
	MSE	.081	1.110	.073	2.950	.164
60%/0.01	Bias	−.288	−1.700	.005	−3.070	.029
	Variance	.101	2.240	.211	.836	.145
	MSE	.184	5.150	.211	10.3	.229

Table 4.27: Scenario 3 MLE Efficiency Studies at Different Censoring Rate, Overlap=0.10, sample size =800, reps=500

Censor Rate/μ		π_1	μ_1	σ_1	μ_2	σ_2
0%/999	Bias	−.012	−.024	−.011	−.036	.000
	Variance	.005	.034	.010	.067	.017
	MSE	.005	.035	.010	.068	.017
20%/3.06	Bias	.052	−.053	−.026	−.450	−.084
	Variance	.013	.058	.012	.173	.035
	MSE	.016	.061	.013	.375	.042
40%/1.41	Bias	.031	−.420	−.031	−1.120	−.016
	Variance	.061	.511	.043	.996	.135
	MSE	.062	.687	.044	2.260	.136
60%/0.01	Bias	−.308	−1.450	.122	−3.320	.327
	Variance	.085	1.480	.178	.479	.121
	MSE	.180	3.580	.193	11.5	.228

4.5 Real Data Analysis

In this section, we illustrate the proposed MHDE with weighted KM estimator by applying it to a real data set which is Panel Study of Income Dynamics (PSID) Data. After the simulation studies, since all three estimations SEMIMHDE, MHDEP and MLE show good results when right censoring rate is around 40% in scenario 1 and 2, we apply our proposed SEMIMHDE and MHDEP methods to a real-world dataset. The dataset is obtained from Mroz (1987) and has been previously used by Zeller et al. (2019) to illustrate their methods. This dataset is available in the R package "CensMixReg". The dataset was originally used to study the determinants of labor force participation and hours worked by married women. It includes observations for 753 married women. Among these women, 325 reported zero hours worked outside the home. This implies a substantial number of observations where the dependent variable (hours worked) is zero, which is why the dataset is often used in censored regression models. From the figure presented in the paper Zeller et al. (2019), it shows the data is a mixture of two distributions, but the data are left-censored with a censoring rate of 43%. Left-censoring occurs when values below a certain threshold are not observed or recorded. To make this data suitable for our methods, we will convert it from left-censored to right-censored data.

There are several methods to convert left-censored data into right-censored data. These methods include: Reflection Method: Reflecting the data around a certain point, such that values below the threshold become values above a corresponding upper threshold. Transformation Method: Applying a monotonic transformation that reverses the censoring direction. Interval Censoring: Treating left-censored observations as being within a certain interval, which can then be interpreted in a right-censored framework. For our analysis, we will use the reflection method, which is straightforward and preserves the essential characteristics of the data. After converting the data to right-censored form, we will apply our proposed

SEMIMHDE and MHDEP methods.

We will apply SEMIMHDE and MHDEP to the converted right-censored dataset. These estimations will be compared against the Maximum Likelihood Estimation (MLE) method to evaluate their performance. The focus will be on examining the robustness and accuracy of SEMIMHDE and MHDEP in handling real-world data with significant censoring.

The dataset has a left-censoring rate of 43%, which, after conversion, translates to a right-censoring rate of 43%. This relatively high level of censoring presents a rigorous test for our proposed methods. From the previous three scenarios simulation studies, all three estimations can still have good performance when censoring rates achieves 40% in scenario 1 and 2, which gives us motivation to apply MHDE to this real data set with censoring rate of 43%.

Figures and tables will be presented to illustrate the results of the SEMIMHDE, MHDEP, and MLE methods. These will include: Parameter Estimates: Comparison of the parameter estimates obtained from SEMIMHDE, MHDEP, and MLE. We will present this part in a table. Model Fit: Graphical representations of the estimated distributions to visually compare the fit of each method to the data. We will combine each fitted density plot based on regarding estimation results with the histogram of the PSID data, which is is the wife's annual work hours outside home divided by 1000. We will present this part in several figures.

The results will be discussed in the context of the performance of SEMIMHDE and MHDEP compared to MLE. Particular attention will be paid to the following: Stability of Estimates: How the estimates from each method behave under the high censoring rate. Robustness: The ability of SEMIMHDE and MHDEP to provide reliable estimates despite the significant level of censoring. Comparative Performance: How the proposed methods stack up against the traditional MLE approach in terms of accuracy and computational efficiency.

In summary, this section demonstrates the application of SEMIMHDE and MHDEP to a real-world dataset with a high censoring rate. By comparing these methods with MLE, we aim to highlight the advantages and potential limitations of our proposed approaches in practical settings.

Based on the truncated initial estimates, we obtain the initial vector from the PSID data as follows:

$$(\pi_1, \mu_1, \sigma_1, \mu_2, \sigma_2) = (0.3, -1.882, 0.1894, -0.2835, 0.1984)$$

Using these truncated initial estimates, we present three estimation methods in the following table, and in order to see the influence of the bandwidth selection and initial selection, SEMIMHDE and MHDEP are present in the last two rows in the table, the last column is the L_2-distance which is also the objective function for SEMIMHDE and MHDEP, the estimations results are present in the following table:

Table 4.28: Comparison of three estimations based on truncated initial estimates

Estimations	π_1	μ_1	σ_1	μ_2	σ_2	L_2 Distance
θ_{MLE}	0.457	-1.549	0.657	-0.243	0.155	NA
θ_{MHDEP}	0.374	-1.536	0.548	-0.280	0.150	0.648
θ_{SEMIMHDE}	0.272	-1.873	0.189	-0.303	0.194	0.982

Table 4.29: Comparison of two MHD estimations based on MLE as initial estimates and appropriate bandwidth

Estimations	π_1	μ_1	σ_1	μ_2	σ_2	L_2 Distance
θ_{MHDEP}	0.372	-1.540	0.577	-0.227	0.189	0.428
θ_{SEMIMHDE}	0.355	-1.420	0.456	-0.316	0.140	0.262

From Table 4.28, the θ_{SEMIMHDE} estimate using truncated initials and bandwidth following the setting in the HMIX algorithm differs from the other two estimations. To investigate whether SEMIMHDE using MLE estimates as initials and select a different bandwidth would acheive better results, we recalculated SEMIMHDE and MHDEP using above MLE as the initial estimates and change the bandwidth to be larger. The SEMIMHDE and MHDEP results are presented in the Table 4.29. By comparing two pairs of MHDEP and SEMIMHDE, we can see both MHDEP and SEMIMHDE have better results since the L2 distance is smaller when using MLE as initials and larger bandwidth, also each parameter estimation has less difference by comparing with MLE.

The three estimations show slightly different results when using MLE as initials and choosing larger bandwidth. The two parametric estimations, MLE and MHDEP, yield similar outcomes, while SEMIMHDE indicates a smaller mixing proportion and scale parameter for the first component when using truncated initials and smaller bandwidth.

Based on the estimations presented in Table 4.28 and Table 4.29, we overlay the fitted density plots on the histogram of the original PSID data to assess which estimation method best fits the data. Given that the PSID data contains 43% censored observations, we are constrained to constructing histograms based solely on the available censored data. It is important to highlight that the fitted density plots are derived from parameter estimates obtained through various estimation methods. Specifically, these plots are based on five distinct sets of parameter estimates. It should be noted that each of Figure 4.1 and 4.3 combines a fitted density plot with the histogram of the censored data. Consequently, the density plots and the histograms represent different aspects of the data. The density plots are generated using the parameter estimates, aiming to provide a visual approximation of the underlying data distribution. In contrast, the histograms reflect the observed distribution of the censored data points. This distinction is crucial because the density plots attempt to infer the complete data distribution, which remains partially unobserved due to the censored nature of the data. Therefore, while the histograms of the censored data are included to provide a reference point, they do not fully represent the same subject as the density plots. The histograms serve as a rough comparison tool, offering a visual context for the censored observations and allowing for a comparative analysis against the fitted density estimates. Therefore, the histograms illustrate the distribution of the censored data, the fitted density plots represent an estimation of the complete data distribution.

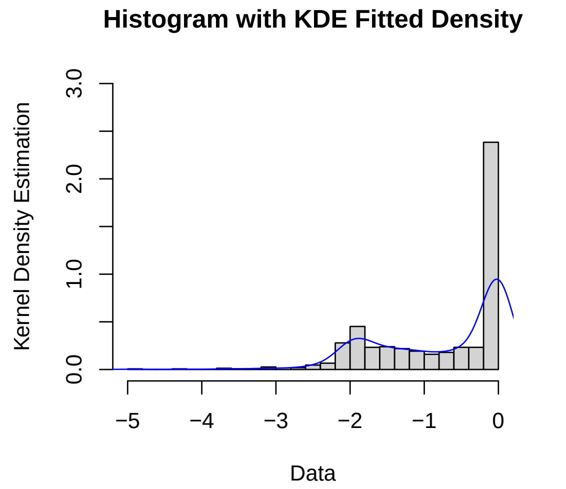
Figure 4.1 (a) illustrates the nonparametric kernel density estimation (KDE) fitted density plot superimposed on the data histogram. This figure serves as the symmetric counterpart to the one presented in Zeller et al. (2019). It demonstrates that the KDE estimation for the second location parameter, μ_2, is zero. Figures 4.1 (b) and 4.1 (c) depict the fitted density plots obtained from the maximum likelihood estimation (MLE) and the minimum Hellinger distance estimation for parametric models (MHDEP), respectively. Both parametric estimation methods produce similar density plots. Notably, for both MLE and MHDEP, the first location parameter, μ_1, is shifted to the right compared to the KDE. Additionally, the first scale parameter, σ_1, is larger, which results in a poorer fit for the first component. Moreover, the second scale parameter, σ_2, is smaller, causing the density plot of the second component to be higher and narrower. Figure 4.1 (d) illustrates the fitted density plot for the semiparametric minimum Hellinger distance estimation (SEMIMHDE). This method results in smaller scale parameters, σ_1, producing higher and narrower curve for the first component. Given that the histogram is derived from real data, which includes 43% zeros, the actual shape of the data histogram, when not censored, remains uncertain. Consequently, it is challenging to conclusively determine which estimation method provides a superior fit to the data.

Since the four figures in Figure 4.1 only present the four estimations using truncated initial values and a small bandwidth, and given that the SEMIMHDE estimations from Table 4.1 and the plots in Figure 4.1 show a obvious difference compared to the other three estimations, we are interested in evaluating how these four estimations fit the PSID data when the censored data are removed. Figure 4.2 displays the four estimations' fitted density plots using a same bandwidth with Figure 4.1 and truncated initial values overlaid on the histogram of the data with censored observations removed. All four density plots in Figure 4.2 are represented by dashed lines. From Figure 4.2, it is evident that the density plots for all four estimations are significantly higher than the histogram for the second component. This discrepancy arises because the 43% of censored data predominantly belong to the second component, which is near zero. Regarding the first component, it is apparent that the SEMIMHDE provides a superior fit compared to the other three estimations.

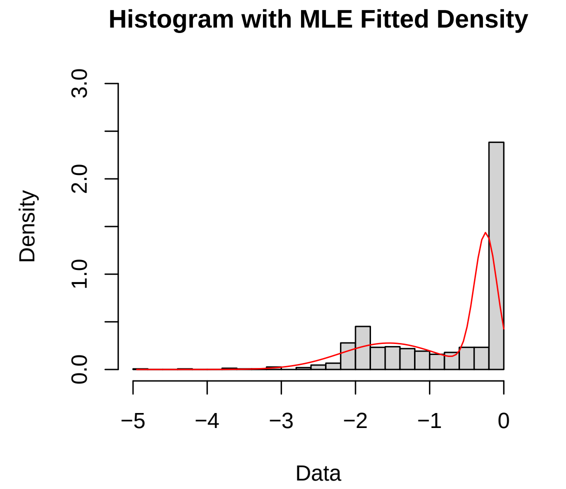
Based on the last two rows of SEMIMHDE and MHDEP from Table 4.1, it is evident that both SEMIMHDE and MHDEP exhibit better performance when utilizing a larger bandwidth and MLE as initial estimates. Therefore, we are interested in evaluating how the fitted densities of these two estimations perform on the histograms of the PSID data. The corresponding figures are presented in Figure 4.3. We can compare MHDEP and SEMIMHDE with different bandwidths and initial values vertically in Figure 4.3. For MHDEP, Figure 4.3 (c) demonstrates a better fit for the second component compared to Figure 4.3 (a), as it includes a larger estimate for the second scale parameter. For SEMIMHDE, Figure 4.3 (d) provides a better fit for the first component compared to Figure 4.3 (b), due to a larger estimate for the first scale parameter. Thus we can conclude that both MHDEP and SEMIMHDE yield better-fitted density plots on the histogram of PSID data when using MLE as initial estimates and a larger bandwidth.

4.6 Conclusion and Discussion

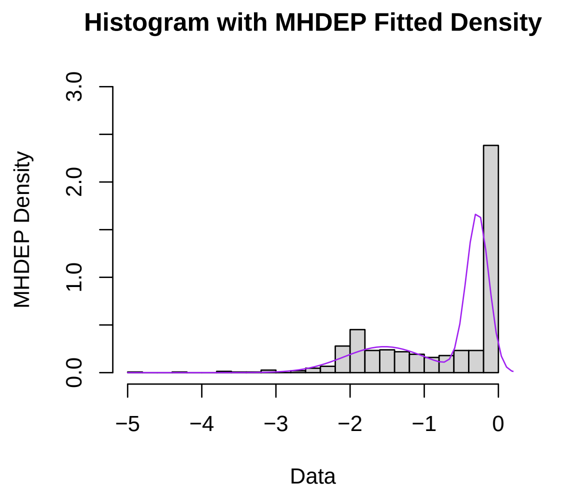
In conclusion, this chapter has demonstrated the performance of SEMIMHDE and MHDEP by simulation and their application to a real world data set characterized by a high censoring rate. By comparing these semiparametric methods with the traditional MLE, we aimed to elucidate the strengths and potential limitations of our proposed approaches in practical scenarios. Our analysis reveals that the selection of appropriate initial values and bandwidths is crucial when applying SEMIMHDE and MHDEP to real data. The importance of these initial values and tuning parameters cannot be overstated, as they significantly influence the performance and accuracy of the density estimations. When suitable initial values and bandwidths are chosen, both MHDEP and SEMIMHDE can perform comparably to the MLE, demonstrating their robustness and adaptability. The results indicate that MHDEP and SEMIMHDE not only are viable alternatives to MLE but also offer specific advantages, particularly in handling datasets with substantial censoring. These methods provide more flexibility in modeling complex data structures where the underlying distribution may not be fully known or easily specified. However, the necessity of carefully selecting tuning parameters highlights a potential limitation, suggesting that further research into automated or data-driven selection techniques could enhance the usability and efficiency of these methods. With judicious tuning parameter selection, these methods can achieve performance levels akin to those of MLE, thereby offering valuable tools for practitioners dealing with intricate data scenarios. In the literature, several authors have explored finite mixture modeling of censored data using various distributions or regression models. These studies predominantly employ parametric models and maximum likelihood estimation (MLE). However, it is important to note that this thesis does not address regression models or multivariate distributions, such as the multivariate Student-t distribution. These aspects remain beyond the current scope and are slated for investigation in our future research endeavors.



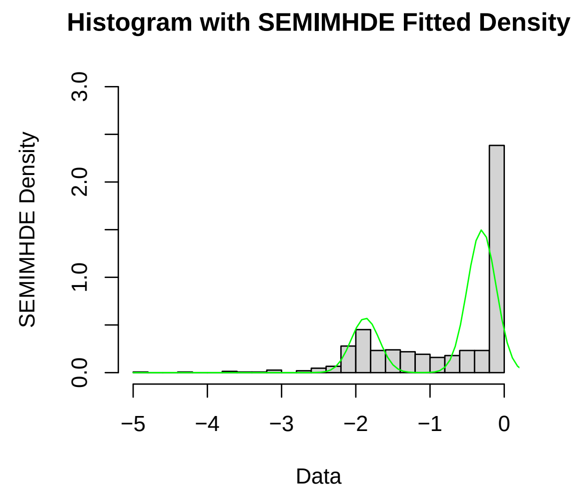
(a) KDE Fitted Density Plot on Histogram



(b) MLE Fitted Density Plot on Histogram

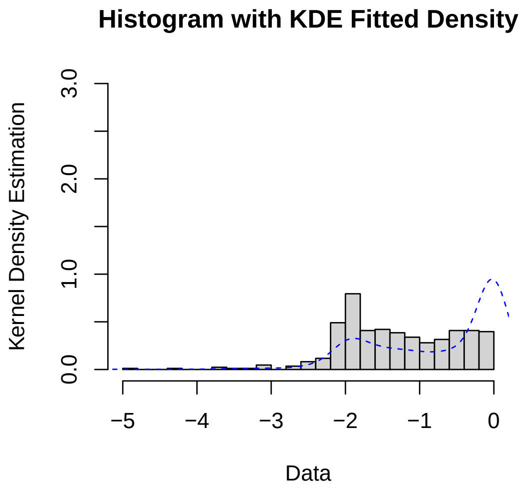


(c) MHDEP Fitted Density Plot on Histogram

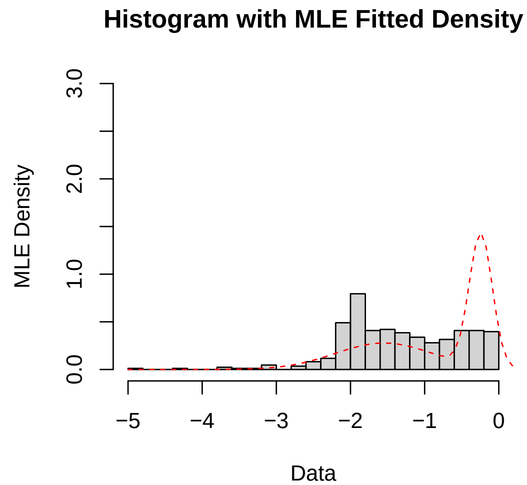


(d) SEMIMHDE Fitted Density Plot on Histogram

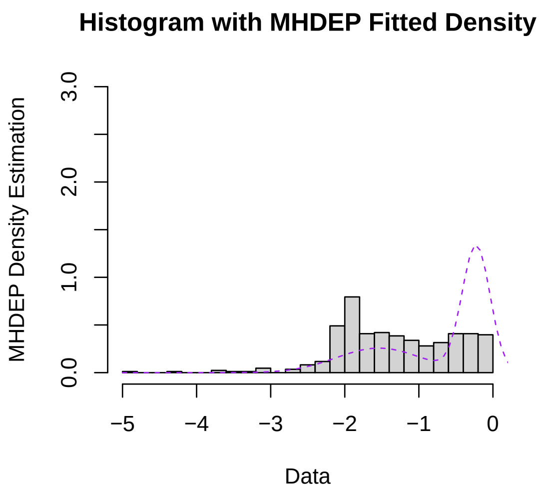
Figure 4.1: Comparison among 4 Estimations Fitted Density Plots on Histogram of PSID data



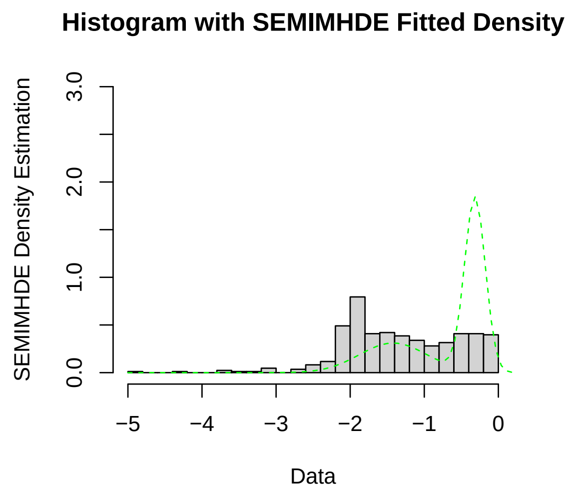
(a) KDE Fitted Density Plot on Histogram



(b) MLE Fitted Density Plot on Histogram

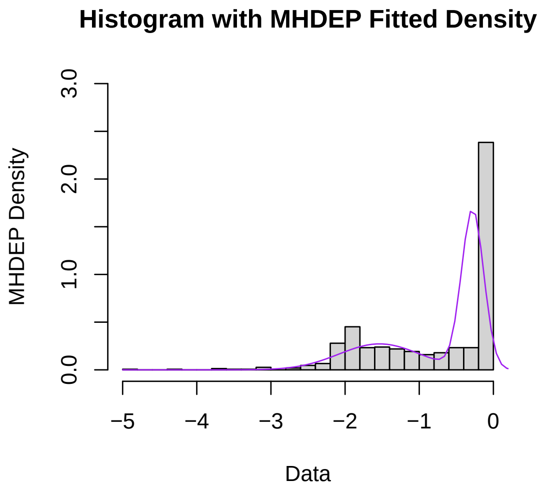


(c) MHDEP Fitted Density Plot on Histogram

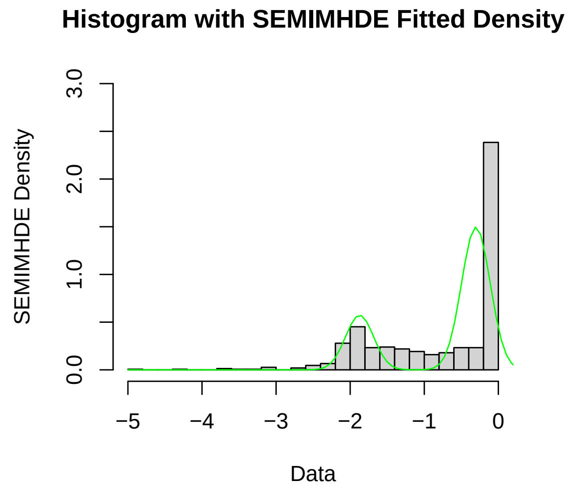


(d) SEMIMHDE Fitted Density Plot on Histogram

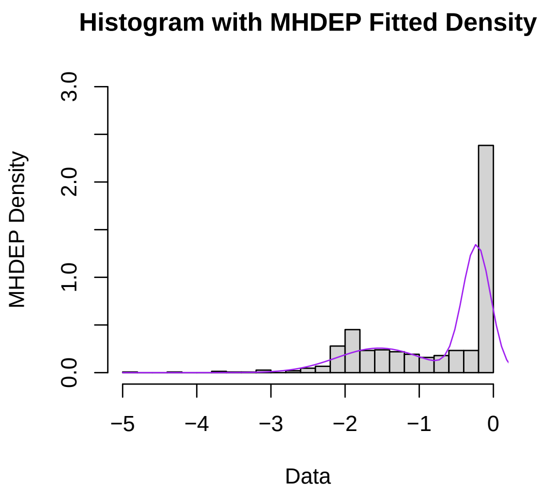
Figure 4.2: Comparison among 4 Estimations Fitted Density Plots on Histogram of PSID Data with zero hours removed



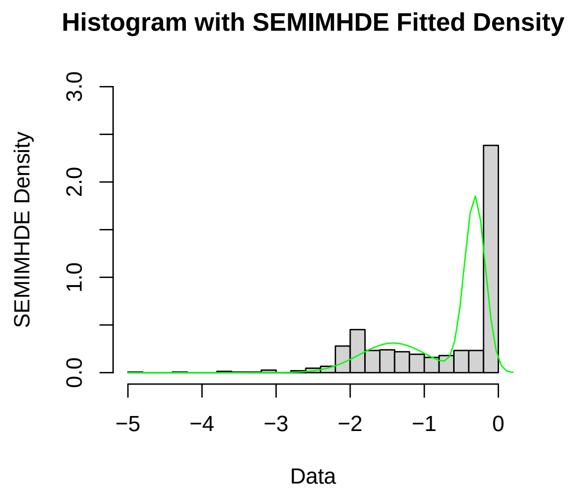
(a) MHDEP Fitted Density Plot on Histogram, with smaller bandwidth and truncated initials from Table 4.28



(b) SEMIMHDE Fitted Density Plot on Histogram, with smaller bandwidth and truncated initials from Table 4.28



(c) MHDEP Fitted Density Plot on Histogram, with larger bandwidth and MLE as initials from Table 4.29



(d) SEMIMHDE Fitted Density Plot on Histogram, with larger bandwidth and MLE as initials from Table 4.29

Figure 4.3: Comparison among 2 MHD Estimations Fitted Density Plots on Histogram of PSID Data with different bandwidth and initials

Chapter 5

Summary and Discussion

This dissertation mainly focuses on the problems of the minimum Hellinger distance estimation of the unknown parameters in two-component location-scale mixture models, with particular emphasis on parametric mixture models and semiparametric mixture models when data are complete and right-censored, respectively. In Chapter 2, we started with parametric location-scale mixture where the distribution of each component is known. In Chapter 3, we focus on the semiparametric mixture models when the distribution is unknown and data are complete. After we investigate both parametric and semiparametric mixtures with Hellinger distance estimation for complete data, we also tested the estimation for mixture when data is right-censored in Chapter 4. Mixture models are a vital tool in statistics for analyzing heterogeneous populations, capturing the underlying sub-populations by modeling the data as a mixture of distributions. While Maximum Likelihood Estimation (MLE) is a widely used method for parameter estimation in these models, its sensitivity to outliers and model misspecification often leads to biased or inefficient estimates. This thesis explores efficient and robust alternatives, specifically focusing on the Minimum Hellinger Distance Estimation (MHDE) method, which provides resilience to deviations from model assumptions and offers reliable estimates even in the presence of outliers.

The primary contributions of this dissertation focus on three key objectives: developing efficient and robust parameter estimation (Minimum Hellinger Distance Estimations) for both parametric and semiparametric location-scale mixture models, and extending these techniques to handle right-censored data, where complete observations may be unavailable. A summary of each chapter is provided below.

In Chapter 1, we present the foundational concepts of mixture models, including location-shifted and location-scale mixture models. This chapter also provides a comprehensive literature review on mixture models and estimation techniques, with a particular focus on methods based on Hellinger Distance. We reviewed parametric, nonparametric, and semiparametric estimation approaches, highlighting their respective advantages and limitations. Special attention is given to the shortcomings of traditional methods like MLE, especially in the presence of outliers or model misspecification. Additionally, we reviewed robust alternatives, emphasizing Hellinger Distance Estimation as a more reliable approach in such scenarios. Furthermore, we review the literature on identifiability for both parametric location-scale mixture models and semiparametric location-shifted mixture models.

In Chapter 2, we focus on parametric MHDEP for location-scale mixtures. Firstly, the theoretical properties of MHDEP are explored, such as consistency and asymptotic normality. Simulation studies showed that MHDEP offers enhanced robustness and efficiency compared to classical parametric MLE, particularly in scenarios where MLE performs poorly due to model misspecification or data with outliers. We also applied MHDEP to the real dataset hydropower data, and MHDEP can fit the data very well after we preprocessed the missing data and treat the new dataset as complete data.

Chapter 3 extends SEMIMHDE method to semiparametric location-scale mixture models, where the distribution of the mixture components is unspecified. We first derive the base function for each component distribution, then develop a modified HMIX algorithm for semiparametric mixture that builds on the parametric HMIX algorithm, substituting the component distributions with the derived base functions. The simulation studies focus on evaluating efficiency, robustness, and the impact of model misspecification. Results from these simulations demonstrate that SEMIMHDE outperforms traditional MLE-based methods, particularly in scenarios involving complex or unknown distributions, or when the data contains outliers. In the real data analysis, we apply the proposed methods to the Old Faithful geyser dataset, focusing on the waiting time between eruptions. We compare the performance of MHDEP, SEMIMHDE, HMIX, and MLE in terms of efficiency and robustness on this dataset. Four estimations show similar results regarding efficiency, however, all three distance-based estimations have very strong robustness and MLE is not robust to outliers at all.

In Chapter 4, we tackle the challenge of parameter estimation in mixture models when data are right-censored, a common issue in applications like survival analysis, where some observations are incomplete. We refine and extend MHDEP and SEMIMHDE to accommodate right-censored data, enhancing their robustness and applicability in these contexts. The Kaplan-Meier (KM)-weighted kernel density estimation to deal with right-censored data is introduced and incorporated into both MHDEP and SEMIMHDE. Simulation studies, ranging from no censoring to high right-censoring rates, demonstrate that KM-weighted SEMIMHDE and MHDEP maintain good performance even with significant right-censoring in location-scale mixture models. The analysis of the Panel Study of Income Dynamics (PSID) dataset further confirms the effectiveness of these methods, showing that both MHDEP and SEMIMHDE perform well even when the right-censoring rate reaches 40%.

While our work has focused on developing Minimum Hellinger Distance Estimation (MHDE) for two-component location-scale mixtures, several intriguing challenges remain, presenting opportunities for future research. In Chapter 2, we introduced MHDEP for parametric two-component location-scale mixtures, and a natural extension would be to apply MHDEP to three or more component mixtures. In Chapter 3, we considered the identifiability of semiparametric mixtures, but due to the flexibility of location-scale mixtures, we only established some necessary conditions under a constrained parameter space for identifiability. A promising direction for future work is to establish both the sufficient and necessary conditions for the identifiability of two-component semiparametric location-scale mixtures, ensuring that the model parameters can be uniquely determined from the observed data. Theoretical properties such as consistency and asymptotic normality also need to be proved for two-component semiparametric location-scale mixture models. In Chapter 4, we focused on estimation in right-censored mixtures, employing Hellinger Distance Estimation with kernel density estimation using Kaplan-Meier weights. The real data analysis revealed that different bandwidth selections can impact the estimation results. Therefore, a future avenue of research could involve bandwidth selection for nonparametric kernel density estimation in two-component location-scale mixtures to further improve the performance of MHDEP and SEMIMHDE.

Bibliography

Aalen, O. (1978). Nonparametric estimation of partial transition probabilities in multiple decrement models. *The Annals of Statistics*, 6(3):534–545.

Aragam, B. et al. (2020). Identifiability of nonparametric mixture models and bayes optimal clustering. *The Annals of Statistics*, 48(2):1037–1064.

Basu, A. (1996). Existence of a Normal scale mixture with fixed variance and a percentile. *Journal of Statistical Planning and Inference*, 55(1):19–39.

Benaglia, T., Chauveau, D., and Hunter, D. R. (2009). An EM-like algorithm for semi- and nonparametric estimation in multivariate mixtures. *Journal of Computational and Graphical Statistics*, 18(4):505–526.

Beran, R. (1977). Minimum Hellinger distance estimates for parametric models. *Annals of Statistics*, 5(3):445–463.

Bickel, P. J., Klaassen, C. A., Bickel, P. J., Ritov, Y., Klaassen, J., Wellner, J. A., and Ritov, Y. (1993). *Efficient and adaptive estimation for semiparametric models*, volume 4. Springer.

Bordes, L., Mottelet, S., and Vandekerkhove, P. (2006). Semiparametric estimation of a two-component mixture model. *The Annals of Statistics*, 34(3):1204–1232.

Bussy, S., Guilloux, A., Gaïffas, S., and Jannot, A.-S. (2019). C-Mix: A high-dimensional mixture model for censored durations, with applications to genetic data. *Bioinformatics*, 35(14):i481–i489.

Butucea, C. and Vandekerkhove, P. (2014). Semiparametric mixtures of symmetric distributions. *Scandinavian Journal of Statistics*, 41(1):227–239.

Chen, H., Chen, J., and Kalbfleisch, J. D. (2001). A modified likelihood ratio test for homogeneity in finite mixture models. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 63(1):19–29.

Chen, J., Tan, X., and Zhang, R. (2008). Consistency of penalized MLE for Normal mixtures in mean and variance. *Statistica Sinica*, 18:443–465.

Chen, Y., Wainwright, M. J., and Jordan, M. I. (2017). A general framework for identifiability of mixture models. *Electronic Journal of Statistics*, 11:1097–1125.

Chiu, S.-T. (1991). Bandwidth selection for kernel density estimation. *The Annals of Statistics*, 19(4):1883–1905.

Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, 34(2):187–202.

Cutler, A. and Cordero-Brana, O. (1996). Minimum Hellinger distance estimation for finite mixture models. *Journal of the American Statistical Association*, 91(436):1716–1723.

Day, N. (1969). Estimating the components of a mixture of Normal distributions. *Biometrika*, 56:463–474.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1):1–38.

Dick, N. P. and Bowden, D. C. (1973). Maximum likelihood estimation for mixtures of two Normal distributions. *Biometrics*, 29(4):781.

Du Roy de Chaumaray, M. and Marbac, M. (2024). Full-model estimation for Non-parametric multivariate finite mixture models. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 86(4):896–921.

Frühwirth-Schnatter, S. and Pyne, S. (2010). Bayesian inference for finite mixtures of univariate and multivariate Skew-Normal and Skew-T distributions. *Biostatistics*, 11(2):317–336.

Ghosal, S., Ghosh, J. K., and Ramamoorthi, R. (1999). Posterior consistency of Dirichlet mixtures in density estimation. *The Annals of Statistics*, 27(1):143–158.

Giné, E. and Guillou, A. (2001). On consistency of kernel density estimators for randomly censored data: rates holding uniformly over adaptive intervals. *Annales de l'Institut Henri Poincare (B) Probability and Statistics*, 37(4):503–522.

Hall, P. and Zhou, X. (2003). Nonparametric estimation of component distributions in a multivariate mixture. *Annals of Statistics*, 31(1):201–224.

Hathaway, R. (1985). A constrained formulation of Maximum-Likelihood estimation for Normal mixture distributions. *Annals of Statistics*, 13:795–800.

Helali, S., Masmoudi, A., and Slaoui, Y. (2022). Semi-parametric estimation using Bernstein polynomial and a finite Gaussian mixture model. *Entropy*, 24(3):315.

Hennig, C. (2004). Breakdown points for Maximum Likelihood estimators of location-scale mixtures. *Annals of Statistics*, 32(4):1313–1340.

Holzmann, H., Munk, A., and Gneiting, T. (2006). Identifiability of finite mixtures of Elliptical distributions. *Scandinavian Journal of Statistics*, 33(4):753–763.

Huang, M., Li, R., and Wang, S. (2013). Nonparametric mixture of regression models. *Journal of the American Statistical Association*, 108(503):929–941.

Hunter, D. R., Wang, S., and Hettmansperger, T. P. (2007). Inference for mixtures of symmetric distributions. *The Annals of Statistics*, 35(1):224–251.

Ibrahim, J., Chen, M., and Sinha, D. (2001). *Bayesian Survival Analysis*. Springer Series in Statistics. Springer.

Jordan, M. I., Ghahramani, Z., Jaakkola, T. S., and Saul, L. K. (1999). An introduction to variational methods for graphical models. *Machine Learning*, 37(2):183–233.

Kang, J. (2020). Robust estimation for longitudinal data based upon minimum Hellinger distance. *Journal of Applied Statistics*, 47(1):150–159.

Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, 53(282):457–481.

Kasahara, H. and Shimotsu, K. (2015). Testing the number of components in Normal mixture regression models. *Journal of the American Statistical Association*, 110(509):163–174.

Kiefer, J. and Wolfowitz, J. (1952). Stochastic estimation of the maximum of a regression function. *The Annals of Mathematical Statistics*, 23(3):462–466.

Lachos, V. H., López Moreno, E. J., Chen, K., and Cabral, C. R. B. (2017). Finite mixture modeling of censored data using the multivariate Student-T distribution. *Statistics in Medicine*, 36(21):3437–3459.

Lambert, M., Chewi, S., Bach, F., Bonnabel, S., and Rigollet, P. (2022). Variational inference via Wasserstein gradient flows. *Advances in Neural Information Processing Systems*, 35:14434–14447.

Lawless, J. F. (2003). *Statistical Models and Methods for Lifetime Data*. Wiley-Interscience, Hoboken, NJ, 2nd edition.

Lindsay, B. and Basak, P. (1993). Likelihood-based inference for mixture distributions. *Journal of Statistical Planning and Inference*, 18:251–275.

Lindsay, B. G. (1995). *Mixture Models: Theory, Geometry, and Applications*, volume 5. Institute of Mathematical Statistics.

Maleki, M. and Mahmoudi, M. R. (2017). Two-piece location-scale distributions based on scale mixtures of Normal family. *Communications in Statistics-Theory and Methods*, 46(24):12356–12369.

McLachlan, G. and Basford, K. (1988). *Mixture Models: Inference and Applications to Clustering*. Marcel Dekker.

McLachlan, G. and Peel, D. (2000). *Finite Mixture Models*. Wiley.

Mroz, T. A. (1987). The sensitivity of an empirical model of married women's hours of work to economic and statistical assumptions. *Econometrica*, 55(4):765–799.

Nelson, W. (1972). Theory and applications of hazard plotting for censored failure data. *Technometrics*, 14(4):945–966.

Newton, J., Titterington, D. M., Smith, A. F. M., and Makov, U. E. (1986). Statistical analysis of finite mixture distributions. *Biometrics*, 42(3):679.

Otiniano, J. et al. (2015). Identifiability of finite mixtures of skew-Normal and skew-T distributions. *Journal of Multivariate Analysis*, 135:1–14.

Parzen, E. (1962). On estimation of a probability density function and mode. *The Annals of Mathematical Statistics*, 33(3):1065–1076.

Pearson, K. (1894). Contributions to the mathematical theory of evolution. *Philosophical Transactions of the Royal Society of London. A*, 185:71–110.

Prentice, R. L. (1973). Exponential survivals with censoring and explanatory variables. *Biometrika*, 60(2):279–288.

Pu, X. and Arias-Castro, E. (2020). An EM algorithm for fitting a mixture model with symmetric log-concave densities. *Communications in Statistics-Theory and Methods*, 49(1):78–87.

Quandt, R. and Ramsey, J. (1978). Estimation of mixture models using the EM algorithm. *Journal of the American Statistical Association*, 73:793–797.

Redner, R. (1981). Note on the consistency of the Maximum Likelihood estimate for non-identifiable distributions. *Annals of Statistics*, 9:225–228.

Redner, R. and Walker, H. (1984). Mixture densities, Maximum Likelihood and the EM algorithm. *SIAM Review*, 26:195–239.

Roeder, K. (1994). A graphical technique for determining the number of components in a mixture of Normals. *Journal of the American Statistical Association*, 89(427):487–495.

Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*. John Wiley & Sons, New York.

Satten, G. A. and Datta, S. (2001). The Kaplan-Meier estimator as an inverse-probability-of-censoring weighted average. *The American Statistician*, 55(3):207–210.

Scott, D. W. (1992). *Multivariate Density Estimation: Theory, Practice, and Visualization*. John Wiley & Sons, New York.

Tanner, M. A. and Wong, W. H. (1983). The estimation of the hazard function from randomly censored data by the kernel method. *Annals of Statistics*, 11(4):1316–1323.

Teicher, H. (1960). On the mixtures of distributions. *Annals of Mathematical Statistics*, 31:55–73.

Teicher, H. (1961). Identifiability of finite mixtures. *Annals of Mathematical Statistics*, 32:244–248.

van der Laan, M. J. and Robins, J. M. (2003). *Unified Methods for Censored Longitudinal Data and Causality*. Springer, New York.

Wang, Z. (2023). *Bayesian Finite Mixture Regression Models With Cluster-Specific Variable Selection*. PhD thesis, The University of Iowa.

Woodward, W. A., Parr, W. C., Schucany, W. R., and Lindsey, H. (1984). A comparison of minimum distance and maximum likelihood estimation of a mixture proportion. *Journal of the American Statistical Association*, 79(387):590–598.

Wu, J. and Karunamuni, R. J. (2009). On minimum Hellinger distance estimation. *Canadian Journal of Statistics*, 37(4):514–533.

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

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(11):2128–2137.

Wu, J. and Zhou, X. (2018). Minimum Hellinger distance estimation for a semiparametric location-shifted mixture model. *Journal of Statistical Computation and Simulation*, 88(13):2507–2527.

Xiang, S., Yao, W., and Seo, B. (2016). Semiparametric mixture: Continuous scale mixture approach. *Computational Statistics & Data Analysis*, 103:413–425.

Xiang, S., Yao, W., and Wu, J. (2014). Minimum profile Hellinger distance estimation for a semiparametric mixture model. *Canadian Journal of Statistics*, 42(2):246–267.

Xiang, S., Yao, W., and Yang, G. (2019). An overview of semiparametric extensions of finite mixture models. *Statistical Science*, 34(3).

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

Zeller, C. B., Cabral, C. R. B., Lachos, V. H., and Benites, L. (2019). Finite mixture of regression models for censored data based on scale mixtures of Normal distributions. *Advances in Data Analysis and Classification*, 13:89–116.

Zhang, Q. and Chen, J. (2012). Minimum Wasserstein distance estimator under finite location-scale mixtures. In *Advances and Innovations in Statistics and Data Science*, pages 69–98. Springer.

Zitouni, M., Zribi, M., and Masmoudi, A. (2018). Asymptotic properties of the estimator for a finite mixture of Exponential dispersion models. *Filomat*, 32(19):6575–6598.